

基于混合编码策略的电弧炉冶炼钢种的碳含量预报模型

姜 静¹ 孟利东¹ 李素玲¹ 姜 琳²

(1 山东理工大学电气学院, 淄博 255049; 2 安阳钢铁公司二炼厂, 安阳 455004)

摘 要 提出一种混合编码策略的遗传算法(GA)训练电炉钢碳含量神经网络预报模型。先采用二进制编码策略,再采用十进制编码策略继续优化预报模型的权阈值,这种混合编码策略综合了二进制编码搜索能力强和十进制编码变异量可任意小的优点。仿真结果表明,混合编码策略的遗传算法(GA)具有更快的收敛速度和更好的寻优性能。对100 t电弧炉冶炼0.85%~1.00% C的钢种,预报碳含量的精度为 $\pm 0.04\%$ 时混合编码GA的命中率为96%,二进制编码GA的命中率为90%。

关键词 电弧炉 遗传算法 混合编码 预报模型

Predictive Model for Carbon Content in Steel Grade Melting by Arc Furnace Based on Hybrid Coding Method

Jiang Jing¹, Meng Lidong¹, Li Suling¹ and Jiang Lin²

(1 Electric School, Shandong University of Science and Engineering, Zibo 255049;
2 No2 Steel Plant, Anyang Iron and Steel Corp., Anyang 455004)

Abstract A new genetic algorithm (GA) using hybrid coding method to train the neural network model for prediction of carbon content in arc furnace steel is proposed, that is first to use binary coding method then to use decimal coding method continuously to optimize and train the weighted threshold of predictive model. The hybrid coding method combines the advantages of binary coding method with strong search ability and decimal coding method with arbitrarily small variance. Simulation results show that the genetic algorithm (GA) using hybrid coding method has faster convergence rate and better search performance, for instance as steel grade with 0.85%~1.00% C is melted by 100 t arc furnace, for $\pm 0.04\%$ precision of predictive carbon content in steel the hit ratio by using hybrid coding method GA is 96%, but the hit ratio by using binary coding method GA is 90%.

Material Index Arc Furnace, Genetic Algorithm, Hybrid Coding, Predictive Model

编码机制对遗传算法GA(Genetic Algorithm)的性能影响很大,二进制编码和十进制编码存在很大的不同,具体表现在:所需训练时间不同,实数编码是连续参数优化问题的直接的自然描述,实数编码省去编码和解码的过程从而明显缩短了训练用的时间;不同编码方法表示的精度不同,二进制编码存在着连续函数离散化时的映射误差,个体编码串的长度较短时,可能达不到精度的要求;交叉和变异作用对不同的编码机制作用不同,表现在遍历空间,种群的多样性,搜索能力等方面;二进制编码的理论起源早,并仍在继续发展。

遗传算法起源于对生物系统所进行的计算机模拟研究,Holland首先用模式定理和隐含并行性(基于二进制编码)解释了遗传算法的高效搜索能力^[1]。模式定理的理论仍在发展中,最近,一些学者对模式定理的正确性提出了质疑,尽管大量成功的实际应用支持了模式定理所依赖的积木块假设^[2],但至今还没有一种方法用来判别“对于一个给定的问题,积木块假设是否成立”。

本文提出一种混合编码的遗传算法,先采用二进制编码训练预报模型,适应度函数增大到某一预定值时,用十进制编码继续对预报模型进行训练,这种混合编码策略综合了二进制编码搜索能力强和十进制编码变异量可任意小的优点,可以提高算法的收敛速度和优化性能。

1 碳含量预报模型的结构

电弧炉冶炼主要任务是向二次精炼装置提供合乎要求的钢水。本文以钢厂100 t交流电弧炉为研究对象,通过对100 t电弧炉冶炼工艺深入分析,确定供电量、供氧量、烧嘴氧耗、烧嘴油耗、废钢量、铁水量为预报模型的输入变量,模型只对碳含量进行预报,网络结构为6输入单输出,电弧炉冶炼终点碳含量预报模型隐层节点数由经验公式取为8。

图1中, p 表示6维的输入向量, W_1 表示输入层到隐含层的连接权系数, b_1 表示隐含层节点的阈值; W_2 表示隐含层到输出层的连接权系数, b_2 表示输出层节点的阈值。隐含层传函取为tansig函数,

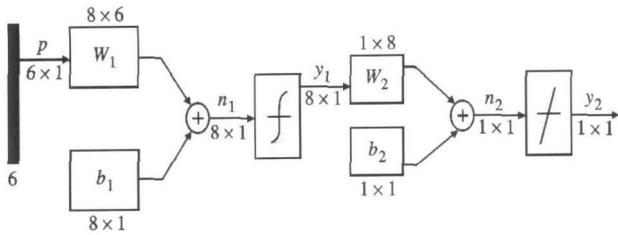


图 1 钢中碳含量预报模型结构简图

Fig. 1 Schematics of structure of predictive model for carbon content in steel

输出层传函取为 purelin 函数。

训练样本数必须多于网络模型的连接权数,一般为连接权数的 2~10 倍。鉴于这个原则,从安阳钢铁公司炼钢厂采集 680 炉高碳钢的炉况数据作为训练和测试预报模型的数据基础,其中 480 炉次数据用于训练模型,200 炉次数据用于对模型测试。每炉次钢可得到的初始条件、过程信息和终点结果的信息包括计划钢种、炉号、装入废钢量,装入的铁水量、石灰和轻烧白云石量,熔清钢水温度和成分、烧嘴氧耗和油耗、碳氧枪喷加的碳粉和氧耗、冶炼时间、送电时间和电耗、出钢量、出钢温度和成分等。为了消除不同变量因量纲或量级不同对网络训练造成的影响,必须对预报模型的输入量和输出量数据按下式进行归一化处理:

$$X_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中: X_i - 数据 i 的归一化结果; x_i - 数据 i 的实际数据; x_{\min} - 参数的最小值; x_{\max} - 参数的最大值。相应地,模型计算结果需参考归一化公式逆变换为实际生产数据^[3]。

2 用混合编码的遗传算法训练预报模型

实数编码的个体只能以单个实数作为最小单位进行交叉,而二进制编码的个体可以以单个位(0 或 1)进行交叉操作。十进制编码的个体经交换后产生的新个体只能在两个父个体的超体空间的顶点上,而二进制编码的个体经交叉后可以产生数目更多的新个体,而且产生的新个体的分布范围更广,不受父个体超体空间的限制,这也说明了二进制编码的群体经交叉后具有更强的搜索能力,群体规模越大,这种差别就越明显。

假定种群中个体数目为 n , x_t^i 表示第 t 代的第 i 个个体, $i \in \{1, 2, \dots, n\}$ 。每个个体的基因位置数为 m ,即个体由 m 个实数构成。这样第 t 代的种群 X_t 可以表示为一个 $n \times m$ 的实数矩阵。

命题:设种群中两个互异的个体 x_t^i 和 x_t^j , $i, j \in \{1, 2, \dots, n\}$,且 $i \neq j$,且 $|x_t^{i(k)} - x_t^{j(k)}| = \delta$, $k = 1, 2, \dots, m$ 。那么它们交换后产生的两个新个体(都用 x_t^i 表示),那么满足 $|x_t^{i(k)} - x_t^{p(k)}| = \delta$,要么满足 $|x_t^{i(k)} - x_t^{p(k)}| = 0$,其中 $p \in \{i, j\}$ 。

证明:因为交换后产生的新个体 x_t^i 的第 k 个基因位 $x_t^{i(k)} \in \{x_t^{i(k)}, x_t^{j(k)}\}$,显然命题成立。

命题表明了对于种群中任意两个互异的个体,交换后产生的新个体一定在这两个父个体所构成的超体 $\prod_{k=1}^m [\min(x_t^i, x_t^j), \max(x_t^i, x_t^j)]$ 的顶点上^[4]。由此可以得出,十进制编码的两个父个体产生的新个体搜索能力是十分有限的。

用二进制编码时,不失一般性,设每个个体包含 m 个实数,每个实数用 l 位二进制数表示,第 t 代种群 X_t 可以表示为一个 $n \times ml$ 的矩阵。个体 x_t^i 的第 k 个长度为 l 的二进制码串转化为实数的解码函数为:

$$\Gamma(x_t^i, k) = u_k + \frac{v_k - u_k}{2^l - 1} \left(\sum_{j=1}^l x_t^{i(kl+j)} \times 2^{j-1} \right) \quad (2)$$

式中: u_k, v_k - 分别是个体 x_t^i 的第 k 个长度为 l 的二进制码串表示的实数范围的上、下界。

二进制编码的两个父个体经交叉后产生的新个体在搜索空间分布范围更广,而十进制编码的两个父个体经交叉后产生的新个体只能在父个体的超体空间的顶点上。二进制编码产生的新个体可能在父个体的超体空间内,也可能在父个体的超体空间外。

先用二进制编码策略的遗传算法对预报模型进行训练,步骤如下:

(1) 种群规模和适应度函数的确定。群体规模可以根据实际情况在 10~150 选择。碳含量预报模型的种群规模选为 90,预报模型需要训练的权阈值的数目为 65,每一个权系数(阈值)用 5 位二进制数表示,权系数和阈值设定在 $[-4, 4]$ 之间。这样种群由 $90 \times (65 \times 5)$ 二进制码矩阵组成。

遗传算法寻优的本质是以群体中各个体的适应度为依据,通过选择、交叉等操作反复迭代,不断寻出适应度较好的个体,最终得到问题的最优解。适应度函数是评价群体中个体好坏的标准,是模拟自然选择的唯一依据^[5]。

碳含量预报模型的适应度函数定义为:

$$F = \frac{1}{1 + E} \quad (3)$$

式中: E - 网络学习的误差能量函数。

(2) 选择方法和最优保留策略。选择操作就是用来确定如何从父代群体中选取个体遗传到下一代的运算方法,采用最多的选择方式是轮盘赌选择,也称为比例选择或复制。该方法首先计算某代群体中个体位串的适应值,然后计算此适应值在群体总适应值中所占的比例,即为该个体在选择过程中被选中的概率。最优保存策略就是把每一代遗传操作后产生的新一代群体的最高适应值与上一代群体的最高适应值作比较,如果小于上一代的最高适应值,就随机淘汰新一代中的一个个体,把上一代中具有最高适应值的个体加入到新一代中^[6]。从理论上讲,在 t 趋于 ∞ 时,采用最优保留策略的遗传算法依概率 1 收敛到全局最优解。

(3) 交叉和变异操作。遗传算法中的交叉运算,是指对两个相互配对的染色体按某种方式相互交换其部分基因,从而形成优秀基因组合的新个体。变异算子使得遗传算法在接近最优解邻域时能加速向最优解收敛,并可以维持群体多样性,避免未成熟收敛。

依交叉概率($p_c = 0.6$)选择群体中要进行交叉变换的个体,对选中的个体进行随机配对,配对个体的等位基因采用两点交叉方法,依变异概率($p_m = 0.005$)对个体上的每一位进行变异操作。

种群通过复制、交叉、变异等不断演化,产生出新的更加优良的种群,经过若干代的进化,若种群中最高个体适应度值大于某一预定值时,采用浮点数编码策略继续对种群进行优化,交叉策略选为均匀交叉,采用高斯变异算子进行变异操作,对于十进制编码变异后的种群中的各个体采用模拟退火操作,模拟退火算法是模拟固体熔化和结晶过程而得到的优化算法,它的一个特点是除接收优化解外,还在一个限定范围内接受恶化解,模拟退火算法作为一种自适应变概率的变异操作,增强和补充了遗传算法的进化能力。

3 仿真结果分析

分别采用混合编码策略的遗传算法和只采用二进制编码的遗传算法对 680 炉高碳钢中的碳含量预报模型进行训练,预报钢种的碳含量范围在 0.85% ~ 1.00%,两种算法的训练过程和预报模型的命中率如图 2 和表 1 所示。

从图 2 和表 1 可以看出,用混合编码策略训练预报模型在相同的精度范围内有更高的命中率,混

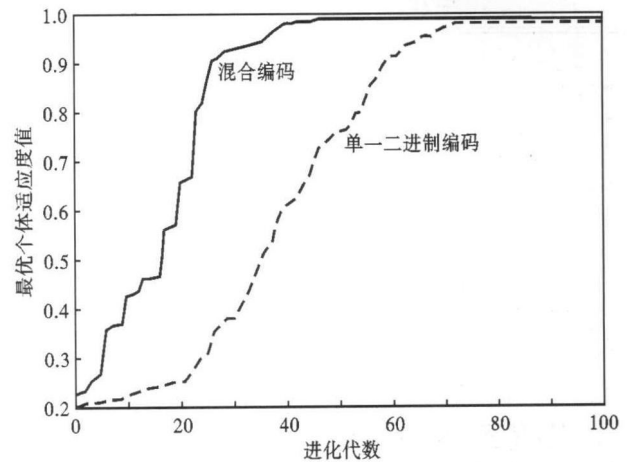


图 2 混合编码遗传算法(GA)和二进制编码遗传算法(GA)学习过程比较

Fig. 2 Comparison between training processes of hybrid code method GA and binary code method GA

表 1 二进制编码 GA 和混合编码 GA 预报钢中碳含量不同精度范围命中率比较/%

Table 1 Comparison between hit ratios for carbon content in steel at different precision range predicted by hybrid code method GA and binary code method GA /%

碳含量精度范围	二进制编码(GA)	混合编码(GA)
± 0.04	90	96
± 0.03	89	93
± 0.02	81	86
± 0.01	71	75

合编码策略综合了二进制编码可以产生更多的新个体和搜索空间大,而十进制编码变异量可以任意小的优点,加快了训练时的收敛速度并提高了寻优性能。

参考文献

- Goldberg D E. Genetic Algorithms in Search, Optimization and Machine Learning. Reading, MA: Addison-Wesley, 1989
- 戴晓晖,李敏强. 遗传算法理论研究综述. 控制与决策, 2000, 15(3): 263
- 李 亮,姜周华,王文忠. VD 炉终点钢液温度预报. 钢铁, 2003, 38(1): 17
- 张晓缙,方 浩,戴冠中. 遗传算法的编码机制研究. 信息与控制, 1997, 26(2): 135
- 张思才,张方晓. 一种遗传算法适应度函数的改进方法. 计算机应用与软件, 2006, 23(2): 108
- 王 蕾,沈庭芝,招 扬. 一种改进的自适应遗传算法. 系统工程与电子技术, 2002, 24(5): 76

姜 静(1974-),女,博士,讲师,智能控制理论在工业生产中的应用研究。

收稿日期:2010-06-30