

DOI: 10.19650/j.cnki.cjsi.J1904809

基于改进遗传算法的转炉炼钢过程数据特征选择*

刘 辉^{1,2}, 曾鹏飞^{1,2}, 巫乔顺³, 陈甫刚³

(1.昆明理工大学信息工程与自动化学院 昆明 650500; 2.昆明理工大学 云南省人工智能重点实验室 昆明 650500; 3.云南昆钢电子信息科技有限公司 昆明 650500)

摘要:转炉炼钢生产过程数据特征选择是实现终点碳温预报的关键,针对生产过程高维数据不利于快速精确预测终点碳温的问题,提出一种改进遗传算法的转炉炼钢生产过程数据特征选择方法。首先采用皮尔逊相关系数衡量不同特征的重要贡献度,进而构造反映过程数据特征与终点碳温相关性的目标函数;然后通过目标函数定义了种群的最大、最小、平均适应度和随机个体适应度值4个变量,建立了一种自适应调节交叉变异概率机制,使得迭代寻优时种群分布更加合理的同时又提高了算法后期收敛速度,防止陷入局部最优。最后进行实际钢厂生产过程数据特征选择验证和对比实验,结果表明,特征选择平均用时为0.25 s,用于终点预报中温度误差在 $\pm 5^\circ\text{C}$ 的精度为85.67%,碳含量预测误差在 $\pm 0.01\%$ 的精度为80.67%。

关键词: 转炉炼钢;特征选择;碳温预测;改进遗传算法

中图分类号: TH865 TF31 **文献标识码:** A **国家标准学科分类代码:** 510.8040

Feature selection of converter steelmaking process based on the improved genetic algorithm

Liu Hui^{1,2}, Zeng Pengfei^{1,2}, Wu Qiaoshun³, Chen Fugang³

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;

2. Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China;

3. Yunnan Kungang Electronic Information Technology Co. Ltd., Kunming 650500, China)

Abstract: Data feature selection of converter steelmaking process is the key step to realize the end point carbon content and temperature prediction. The high-dimensional data of production process are not conducive to the rapid and accurate prediction of the end point carbon temperature. To address this problem, an improved genetic algorithm is proposed to select the data feature of converter steelmaking process. Firstly, Pearson correlation coefficient is used to measure the important contribution of different features. Then, the objective function is formulated to reflect the correlation between process data feature and terminal carbon temperature. The maximum, minimum, average fitness and random individual fitness of the population are defined by the objective function. In this way, an adaptive crossover mutation probability mechanism is established. This method not only makes the population distribution more reasonable during the iteration optimization, but also improves the late convergence speed to prevent the algorithm from falling into local optimization. Through verification and comparison experiments of data feature selection in actual steel mills, results show that the average time of feature selection is 0.25 s, the accuracy of temperature error within $\pm 5^\circ\text{C}$ in terminal prediction is 85.67%, and the accuracy of carbon content prediction error within $\pm 0.01\%$ is 80.67%.

Keywords: converter steel-making; feature selection; carbon temperature prediction; improved genetic algorithm

0 引言

目前,转炉炼钢是主要的炼钢生产方式,转炉以其高生产率和低成本的优势在世界范围内得到广泛应用。转炉炼钢的关键任务之一就是实现对吹炼终点的准确控制,终点控制的主要目标是使得熔池内碳含量和温度达到出钢的工艺要求。因此,终点熔池碳、温的快速、准确、实时预报对提高炼钢生产效率、降低冶炼成本、节约原材料和节能减排有着重要的意义^[1]。

其中,碳含量和温度的检测方法主要是副枪检测,但副枪在吹炼的全过程中只能间断性探测,不能实现连续实时测量,且副枪探头长期处于高温腐蚀环境中,使用和维护成本较高。温宏愿等^[2-3]利用炉口光谱辐射法实现对终点碳温的预报,同时指出在炼钢现场受钢包等其他高温物体的影响,获取到的光谱图片存在一定的干扰性。

在转炉炼钢过程中,加入的石灰石、铁水、废钢、吹氧量等过程数据与终点碳含量和温度有密切的对应关系^[4],采用过程数据分析实现转炉炼钢终点碳、温预测成为目前的研究热点。使用过程数据预测碳、温的方法中,Kubat等^[5]使用模糊方法建立了碱性氧气转炉(basic oxygen furnace, BOF)过程的静态模型;Coxa等^[6]采用BP神经网络(back propagation, BP)模型对BOF的动态过程进行控制;谢书明等^[7-8]综合使用径向基函数(radial basis function, RBF)神经网络对转炉终点碳含量和温度进行预报。在上述研究中,转炉炼钢过程数据全部特征作为预测模型的输入变量,包含着大量的冗余数据和无效特征,进而导致模型预测时间过长、精度下降^[9-10]。为解决上述问题,把特征选择作为重要的处理步骤,以此从大量数据集中剔除冗余信息,选择出具有代表性的特征子集,从而实现对数据规模和维度的精简,提升数据分析和处理的效率。因此,王心哲等^[11]提出基于变量选择的转炉终点预报模型;代照坤^[12]采用灰色关联分析(grey relational analysis, GRA)与贡献度(important contribution degree, ICD)相结合的方法(GRA-ICD算法)对炼钢过程输入特征进行选择。然而,随着炼钢过程传感器的增设,又会产生更多的过程数据,导致特征选择过程计算复杂度增加、时间过长,成为终点碳、温预测实时性的突出问题,进而错过转炉出钢的最佳时刻。因此,研究特征选择的最优性和实时性,能解决因冗余特征造成预测模型性能降低的问题,同时可以提高终点控制的准确性和实时性^[13]。

本文提出一种基于改进遗传算法(improved genetic algorithm, IGA)的特征选择方法,并将其应用于转炉炼钢终点碳含量和温度的预测任务中。在特征选择时,首先建立描述转炉炼钢生产过程数据特征与终点碳、温关

联度的目标函数,并作为遗传算法迭代优化的目标函数。其次构建自适应调节交叉变异的概率模型,合理调节种群个体适应度值分布,有效克服了后期收敛速度过慢、容易陷入局部最优、目标函数精度低的问题,进而达到快速、准确选择最优特征子集的目的。采用实际炼钢转炉数据进行实验,结果表明,相比于传统遗传算法(traditional genetic algorithm, TGA),改进遗传算法可以快速、精确得到最优特征子集,有效缩短预测模型时间,提高预测精度和预测的实时性。

1 问题分析

本文处理转炉炼钢过程数据特征选择的流程为:首先获取炼钢过程原始数据,进行数据预处理以剔除异常数据,随后的关键步骤是利用遗传算法对特征选择目标函数进行寻优,以获取最优特征。其中,在特征选择之前需要进行特征相关性分析,进而建立特征选择目标函数,然后使用遗传算法进行特征选择,最后将获取的最优特征子集输入预测模型中进行碳、温验证预测。其思路流程如图1所示。

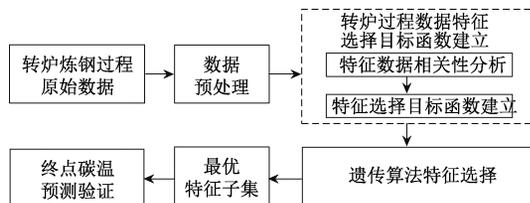


图1 本文思路流程

Fig.1 Flowchart of the study motivation

下文分别分析了目标函数的建立过程,以及传统遗传算法在转炉炼钢过程数据特征选择的思路及出现的问题。

1.1 转炉过程数据特征选择目标函数建立

选择输入参量中的关键特征从而降低输入数据维度是提高碳、温预报模型效率和精度的关键。因此,为了选择最优特征子集,对转炉炼钢过程输入数据和输出碳、温的相关性建立衡量特征重要度的评价指标,并以此作为待优化的目标函数,实现转炉炼钢过程数据中与终点碳、温具有较高关联度的关键特征选择,进而实现对终点碳、温准确实时预报的目标。

以实际钢厂近几年的转炉炼钢过程生产数据集为研究对象,其过程数据为装入铁水量 h_1 ,装入生铁量 h_2 ,装入废钢量 h_3 ,转入总量 h_4 ,铁水温度 h_5 ,铁水C含量 h_6 ,铁水SI含量 h_7 ,铁水MN含量 h_8 ,铁水S含量 h_9 ,铁水P含量 h_{10} 等120维数据,以上数据为转炉吹炼输入输出数据,其中输出为决定炼钢终点的碳含量和温度。通过建

立输入与输出之间的量化关系,提取重要的过程数据特征。原始数据集 \mathbf{H} 表示如下:

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,120} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,120} \\ \vdots & \vdots & \ddots & \vdots \\ h_{3\ 390,1} & h_{3\ 390,2} & \cdots & h_{3\ 390,120} \end{bmatrix}_{3\ 390 \times 120}$$

数据集 \mathbf{H} 由 3 390 炉次实际炼钢数据组成,每一炉次转炉炼钢数据包含输入与输出共 120 维数据,其中第 113 维和第 114 维特征分别表示终点输出碳含量和温度。

相关性分析数学模型如下:给定 \mathbf{X} 表示输入数据, \mathbf{Y} 表示输出数据,转炉炼钢过程输入数据 $\mathbf{X} \in R^{m \times d}$, 其中 $\mathbf{X} = [x_1, x_2, x_3, \dots, x_{m-1}, x_m]^T$, $x_i \in R^d (i = 1, 2, \dots, m)$, 行向量表示每一炉次样本数据,列向量表示样本特征数。 \mathbf{Y} 表示出炉时钢水碳含量和温度, $\mathbf{Y} \in P^{m \times n}$, 其中 $n = 2$, $\mathbf{Y} = [y_1, y_2, y_3, \dots, y_{m-1}, y_m]^T$, $y_i \in P^n (i = 1, 2, \dots, m)$, 行向量表示每一炉次钢水碳含量和温度。为了对各维度的特征进行评估,采用皮尔逊相关系数 (Pearson correlation coefficient, PCCs) 度量特征之间的相关性^[14], 整理后可得:

$$p_{i,j} = \frac{\text{cov}(x_i, y_j)}{\sigma_{x_i} \sigma_{y_j}} \quad i \in (1, m), j \in (1, 2) \quad (1)$$

由式(1)计算得到各输入特征与输出钢水温度和碳含量的皮尔逊相关系数 $p_{i,j}$ 。因无法建立有效的转炉炼钢过程数据特征选择目标函数,为此定义一种基于重要贡献度算法^[15-17]的目标函数为:

$$R_{ICD} = \left| \frac{p_{i,j}}{\text{Max}(p_{i,j})} \right| \times 100\% \quad (2)$$

式(1)和(2)建立了衡量转炉炼钢过程数据特征重要度的目标函数,特征越重要,目标函数 R_{ICD} 值越大。选取排序靠前的 M 个 R_{ICD} 值以达到特征选择目的。通过理论计算及文献[18]分析,求解最优特征时存在问题如下:1) 寻优空间过高,难以有效找出最优特征子集。2) 计算复杂、耗时过长。

1.2 遗传算法特征选择

为了得到 1.1 节建立的特征选择目标函数的最优解,在由特征构成的解空间中进行不断迭代从而使目标函数逼近最优,这一过程就是对目标函数的优化。具体到本文应用中,目标函数用于衡量特征与终点碳温的关联度,优化过程的进行意味着依次从过程数据特征中找到与终点碳温关联度较大的前 M 个特征。如图 1 所示,遗传算法是实现目标函数优化的关键步骤,该算法在转炉炼钢过程数据特征选择过程中的基本思路如下所述。

遗传算法是一种基于自然选择和基因遗传学原理的随机并行搜索算法,是一种寻求全局最优解而不需要任

何初始化信息的高效优化算法^[19]。算法主要由编码机制、控制参数、目标函数和选择算子 4 部分组成,将其应用于转炉炼钢过程数据特征选择中时,式(1)中的协方差与标准差之积两个变量组成一个 20 位二进制的染色体,当遗传算法寻优出适应度较高的前 M 个染色体时,通过反解码定位转炉炼钢生产过程数据与终点碳温关系最为密切的关键特征,进而实现对转炉炼钢终点碳温的准确实时预报。遗传算法特征选择流程如下:

1) 随机初始化种群 $\mathbf{X} = \{x_i^0, \dots, x_N^0\}$, 种群规模为 N 。随后将衡量特征重要度的目标函数中的两个变量编码成两个十位二进制数。染色体编码与解码对应关系为: k 表示染色体基因个数 ($k = 10$), b_i 表示染色体第 i 个基因位的二进制数。编码过程如下所示:

$$\begin{aligned} \overbrace{000 \cdots 0000}^{10 \text{位二进制数}} &= 0 & y_{\min} \\ 000 \cdots 0001 &= 1 & y_{\min} + \delta \\ 000 \cdots 0010 &= 2 & y_{\min} + 2\delta \\ &\vdots & \\ 111 \cdots 1111 &= 2^k - 1 & y_{\max} \end{aligned}$$

$$\delta = \frac{y_{\max} - y_{\min}}{2^k - 1} \quad (3)$$

解码过程:

$$x = y_{\min} + \left(\sum_{i=1}^k b_i \times 2^{k-1} \right) \times \delta \quad (4)$$

2) 计算种群中每个个体适应度值。适应度是描述遗传算法中单个染色体使得目标函数趋于最优解的程度,在本文应用中,此处的适应度描述当前染色体所代表的特征与终点碳温的关联程度。为了进行后续的算法设计,分别计算得到个体适应度值、最大个体适应度值、最小个体适应度值、种群平均适应度值,用于描述和设计后续自适应交叉、变异因子。个体适应度值越大,则在下一轮迭代寻优时被保留下来的概率就越大。种群平均适应度值的分布用于引导遗传算法进化的方向。当算法达到最优解时,选取使得遗传算法目标函数最优的 M 个染色体构成选择后的特征子集。转炉炼钢过程数据特征选择是寻求目标函数值最大的个体。因此,将式(1)和式(2)结合的目标函数作为遗传算法的适应度函数,可以得出特征越好,适应度值越大的结论。迭代第 t 次时全部个体的适应度值为 $f(x_t)$, 其中 $i = 1, \dots, N$ 。

3) 交叉和变异。交叉是选取成对父代个体按一定概率进行单点交叉、多点交叉。变异是按照某一概率确定变异个体,并实行相应变异操作,以保持种群个体的多样性,防止陷入局部最优。

4) 选择算子。通常采用精英个体保留策略和赌轮选择算子,即保留适应度值较大的个体。在轮盘赌中群体大小为 N , 个体 x_i 的适应度为 $f(x_i)$, 则个体 x_i 的选择概

率为:

$$p(x_i) = \frac{f(x_i)}{\sum_{j=1}^N f(x_j)} \quad (5)$$

其每个个体的累积概率为:

$$Q_i = \sum_{j=1}^i p(x_j) \quad (6)$$

此时,产生一个均匀分布的随机数 $r \in [0, 1]$,当 r 在累计区间 Q_i 中保留第 i 个个体,一般适应度值越大的个体保留下来的概率越大。

5) 迭代终止条件。满足设置的迭代终止条件时寻优结束,在转炉炼钢过程数据中选择适应度排名靠前的 M 个特征,组成最优特征子集 F ; 否则转至步骤 2), 计算新一代种群中每个个体的适应度值。

通过实验结果及相关文献分析^[20-21],传统遗传算法 (traditional genetic algorithm, TGA) 在种群个体适应度值分布不合理时,直接导致遗传算法收敛速度慢,易陷入局部最优和目标函数精度低的问题。而进化机制中交叉、变异是影响种群分布的关键。因此, TGA 在转炉炼钢过程数据特征选择时存在以下问题: (1) 局部搜索能力差; (2) 全局搜索过程耗时较高; (3) 后期收敛速度慢; (4) 目标函数优化精度不足。

上述问题会造成转炉炼钢过程数据特征选择时的快速性、准确性的问题^[22],从而降低终点碳温实时预测的性能。因此,针对上述不足,提出改进遗传算法,目的在于解决传统遗传算法在转炉炼钢过程中的数据特征选择问题。

2 改进遗传算法

文献[23-24]研究分析表明, TGA 中交叉变异的合理选取能有效避免收敛速度慢,易陷入局部最优和优化的目标函数精度低的问题。因此本文对 TGA 进行改进,建立了一种新的自适应交叉、变异概率模型。通过分析种群的离散度进而合理自适应调整交叉、变异概率,使得种群分布更优。

改进遗传算法调整种群最优分布过程描述如下:首先,随着种群初始化,通过已构造的目标函数计算得到种群最大、最小适应度值、种群平均适应度值和种群随机个体适应度值 4 个变量。其次,利用设置的变量建立自适应交叉、变异模型,通过比较种群平均适应度值与种群随机适应度值的大小分析出种群的分布情况,以此引导交叉、变异模型的自适应变化,使得迭代寻优时种群分布更加合理,加速算法后期收敛速度,防止陷入局部最优。从而实现种群分布达到最优,进而获取生产过程数据特征与终点碳温相关度最高的特征子集。

2.1 自适应交叉 p_c 概率模型建立

p_c 算子直接影响种群的进化速度和群体质量。种群进化中交叉概率过大,会导致较优个体被破坏;过小,会导致算法搜索速度慢。根据种群个体的适应度值离散分布情况,自适应调整 p_c 大小,能有效克服算法进化不足的问题。当适应度值分散时, p_c 值变大,使搜索空间扩大,种群更新速度加快;种群个体适应度值较为集中, p_c 需适当减小,进而保证优良个体不被破坏,同时也保证了收敛速度。

设 f_{avg} 表示种群平均适应度值, f' 表示随机选取的个体适应度值,采用 f_{avg} 与 f' 二者的差来表示种群的离散程度,并将结果对 p_c 加权从而实现交叉概率的最佳取值。同时为服从遗传算法自身随机进化的特性,保证交叉、变异概率取值的合理性^[25],令 p_{c1} 、 p_{c2} 、 p_{c3} 、 p_{c4} 代表 4 个不同的交叉概率变量,其取值范围均为 $(0, 1)$ 之间的常数。利用 sigmoid 函数非线性化处理,实现交叉算子自适应调整。设 f_{max} 表示种群中最大个体适应度值, f_{min} 表示种群中最小个体适应度值,用 $f_{max} - f_{avg}$ 、 $f_{max} - f_{min}$ 表征种群差异程度,结合 sigmoid 函数 $\left(g(x) = \frac{1}{1 + e^{-x}}\right)$,指导劣势个体的进化程度和进化方向。因此,自适应交叉模型如式(7)所示。

$$p_c = \begin{cases} \left| \frac{p_{c1}(f_{avg} - f') + p_{c2}(f' - f_{avg})}{1 + e^{-(f_{max} - f_{avg})}} \right|, & |f_{avg} - f'| = \varepsilon, \varepsilon \in (0.2, 1) \\ \left| \frac{p_{c3}(f_{avg} - f') + p_{c4}(f_{avg} - f')}{1 + e^{-(f_{max} - f_{min})}} \right|, & |f_{avg} - f'| = \varepsilon, \varepsilon \in (0, 0.2) \end{cases} \quad (7)$$

式中: $|f_{avg} - f'| = \varepsilon$ 。通过实验证明,当 $\varepsilon \in (0, 0.2)$ 时,离散程度低;当 $\varepsilon \in (0.2, 1)$ 时,离散程度高。

2.2 自适应变异 p_m 概率模型建立

p_m 算子合理设置能保持种群多样性,防止算法陷入局部最优,提高优化精度。进化过程中,种群分布离散时,交叉概率要大,变异概率要小,以防止少数优良个体被破坏;种群差异较小时,需要适当地减小交叉概率,增大变异概率,以增强种群的多样性。根据上述目的及自适应交叉模型分析思路,令代表 4 个不同的变异概率变量取值范围均为 $(0, 1)$ 之间的常数。采用 f_{avg} 与 f' 二者的差表示种群的离散程度,并将结果对 p_m 加权从而实现变异概率的最佳取值。种群分布离散度高,就利用 $f_{max} - f_{avg}$ 结合 sigmoid 函数自适应提高 p_c 交叉概率, $f_{max} - f_{min}$ 结合 sigmoid 函数自适应降低 p_m 变异概率。种群分布离散低,则利用 $f_{max} - f_{min}$ 结合 sigmoid 函数自适应降低 p_c 交叉概率, $f_{max} - f_{avg}$ 结合 sigmoid 函数自适应提高 p_m 概率,从而达到变异算子和交叉算子相互协调的目的。因此,自

适应变异模型如式(8)所示。

$$P_m = \begin{cases} \left| \frac{p_{m1}(f_{avg} - f') + p_{m2}(f' - f_{avg})}{1 + e^{-(f_{min} - f_{max})}} \right|, \\ |f_{avg} - f'| = \varepsilon, \varepsilon \in (0.2, 1) \\ \left| \frac{p_{m3}(f_{avg} - f') + p_{m4}(f_{avg} - f')}{1 + e^{-(f_{min} - f_{max})}} \right|, \\ |f_{avg} - f'| = \varepsilon, \varepsilon \in (0, 0.2) \end{cases} \quad (8)$$

通过以上自适应交叉、变异概率模型的建立,实现种群平均适应度与随机适应度判断种群的分布情况,进而实现交叉、变异模型有方向性的调整其概率大小,使得种群分布达到最优。通过有限次数的迭代进化,达到种群快速收敛,防止陷入局部最优的目的。

3 仿真实验及分析

3.1 实验平台内容介绍

系统所用计算机配置为处理器 i5-cpu,运行内存(RAM)4G,操作类型 64 位,改进遗传算法对比实验运行平台 Python 和 MATLAB。

为了能够从算法和应用两个角度说明本方法的有效性,实验分为以下 3 个模块开展研究:首先,验证传统遗传算法在转炉炼钢过程数据特征选择中的可行性及不足之处,针对其在特征选择时存在计算速度慢、陷入局部最优解的问题进行后续改进;其次,验证基于改进遗传算法的转炉炼钢生产过程数据特征选择的有效性和先进性,并与同类其他改进算法进行对比分析;最后,为了进一步验证特征选择的有效性,在实际数据中对终点碳温进行预报实验。

1) 传统遗传算法转炉炼钢过程数据特征选择基础实验。其目的在于验证遗传算法在转炉炼钢生产过程数据特征选择应用中的适应性和有效性;其次发现传统遗传算法在转炉炼钢生产过程数据特征选择应用中的不足之处,并为基于改进遗传算法的特征选择方法提供思路 and 方向。

2) 改进遗传算法转炉炼钢生产过程数据特征选择及对比实验。用于验证改进遗传算法对改善种群分布的有效性,以及避免陷入局部最优、加快收敛速度方面的性能。以种群平均适应度值作为性能衡量指标,与文献[24]中改进的遗传算法进行对比实验。

此外,为了验证本文特征选择方法的快速性和准确性,选取 TGA、文献[21]、[22]、[24]中的改进遗传算法在转炉炼钢过程数据特征选择中的对比实验。

3) 基于特征选择的转炉炼钢终点碳温预报及对比实验。为了验证本文 IGA 方法在特征选择精度和有效性方面的优势,与文献[24]中改进遗传算法,以及文

献[11]特征选择的方法进行对比实验,并以终点碳、温预测的精度作为衡量指标对结果进行评价和分析。

3.2 改进遗传算法转炉炼钢生产过程数据特征选择及对比实验

1) 改进遗传算法有效性验证及对比实验

为验证改进遗传算法在改善种群个体适应度分布方面的有效性。实验首先以种群平均适应度值为衡量指标,其次将转炉炼钢过程数据特征选择目标函数作为寻优对象,并与文献[24]改进遗传算法的种群分布情况进行对比。由于文献[24]主要用于解决路径优化中遗传算法收敛速度慢、容易陷入局部最优的问题,这与本实验的研究对象具有类似的问题特点,因此将其作为对比。实验结果如图 2~5 所示。

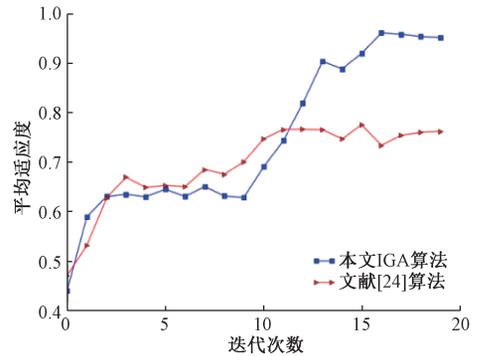


图 2 迭代 20 次种群平均适应度值
Fig.2 The average fitness value of the population obtained after 20 iterations

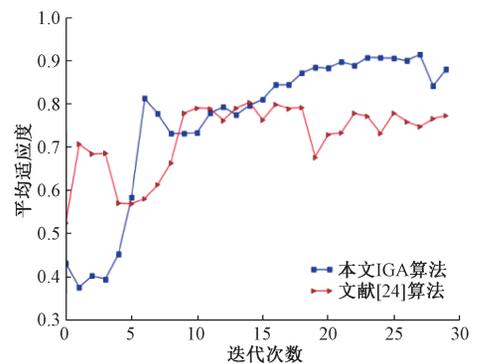


图 3 迭代 30 次种群平均适应度值
Fig.3 The average fitness value of the population obtained after 30 iterations

图 2~5 所示分别为迭代 20 次、30 次、40 次、50 次时种群平均适应度值。在种群分布对比实验初始阶段中,种群适应度平均值均在 0.4 附近。随着迭代开始,IGA 算法种群平均适应度值均优于文献[24]改进遗传算法,说明 IGA 能合理分析出种群分布的离散情况,实

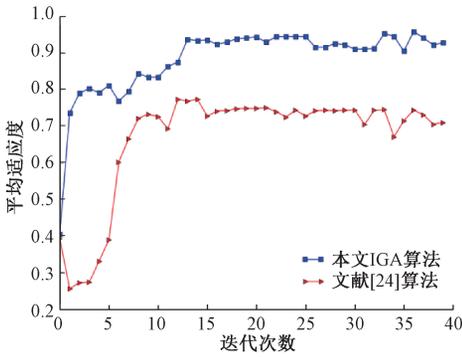


图4 迭代40次种群平均适应度值

Fig.4 The average fitness value of the population obtained after 40 iterations

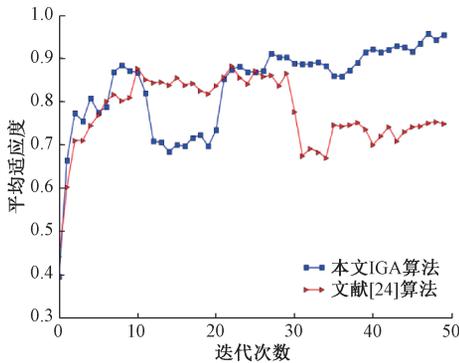


图5 迭代50次种群平均适应度值

Fig.5 The average fitness value of the population obtained after 50 iterations

时调整交叉、变异概率。当种群分布比较离散时,交叉概率变大,变异概率变小,进而使种群适应度值分布更佳且集中。算法后期,文献[24]改进遗传算法平均适应度值在0.7~0.8之间缓慢增加,有陷入局部最优的趋势。但IGA变异概率适当变大,交叉概率减少,从而能产生更优个体,使种群进化更加合理。此外,遗传算法进化的随机特性,有时会导致在寻优过程中后一次迭代的寻优效果比前一代差,从而造成算法稳定性不高的问题。本文提出的改进遗传算法在最大程度上克服了稳定性不高的问题,保证了实验的可靠性。实验结果表明,改进遗传算法有效克服后期收敛速度慢,容易陷入局部最优解的不足。

2) 改进遗传算法转炉炼钢生产过程数据特征选择有效性验证及对比实验

由于转炉炼钢终点碳温预报对实时性、准确性具有较高的要求,为了验证本文方法是否满足实际工业要求,以实际转炉炼钢生产过程数据为研究对象,采用初始值、迭代完成时是否收敛、以及迭代完成时的最优解3个参

数衡量指标对结果进行评价和分析。

(1) 初始值。初始值的大小影响寻优时间,初始值越大寻优时间则越少;

(2) 迭代完成时是否收敛。当达到终止条件时,算法是否收敛是衡量其优劣的重要指标;

(3) 迭代完成时的最优解。对比实验中算法收敛时其最优解表征是否陷入局部最优解。

同时,为了进一步说明本文方法在特征选择中快速性和准确性两方面的优势,选取了在不同应用背景下通过改进遗传算法提升速度和准确度性能的方法作为对比,因此选取了文献[21]、[22]、[24]中的改进遗传算法以及TGA算法作为参考,实现对转炉炼钢过程数据特征选择的对比实验。

对比实验结果如图6~10所示。

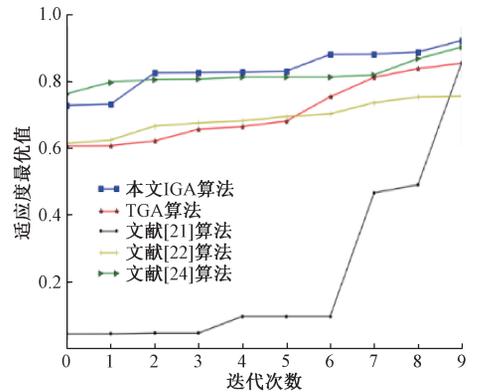


图6 转炉炼钢过程数据特征选择迭代10次

Fig.6 The data feature of BOF steelmaking process selected after 10 iterations

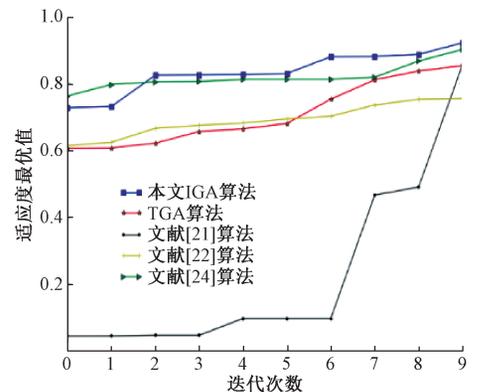


图7 转炉炼钢过程数据特征选择迭代20次

Fig.7 The data feature of BOF steelmaking process selected after 20 iterations

对图6~10所示实验结果进行统计,得到对比实验中的初始值、迭代完成时是否收敛、迭代完成时的最优解

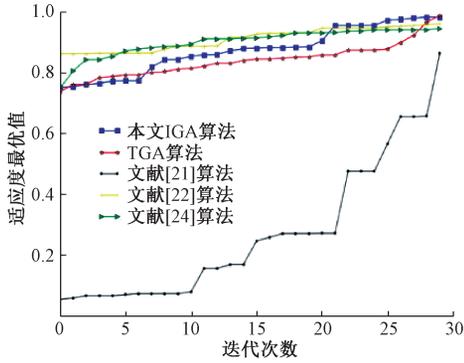


图 8 转炉炼钢过程数据特征选择迭代 30 次
Fig.8 The data feature of BOF steelmaking process selected after 30 iterations

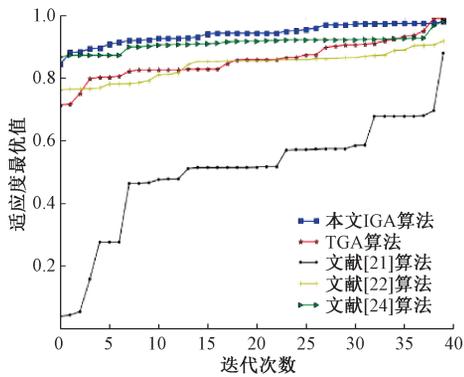


图 9 转炉炼钢过程数据特征选择迭代 40 次
Fig.9 The data feature of BOF steelmaking process selected after 40 iterations

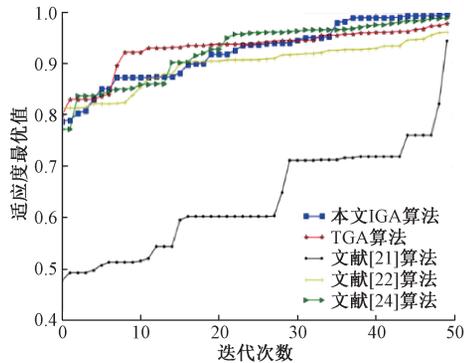


图 10 转炉炼钢过程数据特征选择迭代 50 次
Fig.10 The data feature of BOF steelmaking process selected after 50 iterations

3 个参数,结果如表 1~5 所示。本文改进遗传算法与 TGA、文献[21]、[22]、[24]方法用于转炉炼钢生产过程数据特征选择计算效率对比如表 6 所示。

表 1 迭代 10 次参数情况

Table 1 Parameter after 10 iterations

项目	IGA 算法	TGA 算法	文献[21] 算法	文献[22] 算法	文献[24] 算法
初始值	0.729	0.609	0.043	0.614	0.765
是否收敛	否	否	否	否	否
最优解	无	无	无	无	无

表 2 迭代 20 次参数情况

Table 2 Parameter after 20 iterations

项目	IGA 算法	TGA 算法	文献[21] 算法	文献[22] 算法	文献[24] 算法
初始值	0.899	0.840	0.588	0.823	0.853
是否收敛	收敛	准收敛	否	否	收敛
最优解	0.957	无	无	无	0.917

表 3 迭代 30 次参数情况

Table 3 Parameter after 30 iterations

项目	IGA 算法	TGA 算法	文献[21] 算法	文献[22] 算法	文献[24] 算法
初始值	0.755	0.734	0.051	0.895	0.749
是否收敛	收敛	准收敛	否	否	否
最优解	0.979	无	无	无	无

表 4 迭代 40 次参数情况

Table 4 Parameter after 40 iterations

项目	IGA 算法	TGA 算法	文献[21] 算法	文献[22] 算法	文献[24] 算法
初始值	0.841	0.714	0.031	0.760	0.867
是否收敛	收敛	准收敛	否	否	否
最优解	0.985	无	无	无	无

表 5 迭代 50 次参数情况

Table 5 Parameter after 50 iterations

项目	IGA 算法	TGA 算法	文献[21] 算法	文献[22] 算法	文献[24] 算法
初始值	0.791	0.802	0.479	0.815	0.867
是否收敛	收敛	准收敛	否	准收敛	收敛
最优解	0.985	无	无	无	0.980

表 6 转炉炼钢生产过程数据特征选择平均耗时

Table 6 The average time of data feature selection of converter steelmaking production process (s)

算法	IGA 算法	TGA 算法	文献[21] 算法	文献[22] 算法	文献[24] 算法
特征选择时间	0.25	1.1	0.7	0.46	0.42

在对比实验中,迭代 10 次时的结果如图 6 所示,实验结果显示算法寻优均没有达到收敛,但本文所提方法 IGA 最优个体适应度值为 0.86,且保持最佳,表明 IGA 进化速度快。迭代 20 次时的结果如图 7 所示,IGA 达到收敛,最优解为 0.957,TGA 即将收敛,文献[21-24]改进遗传算法却收敛缓慢,原因在于局部搜索能力不强,无法进化出更优个体。为了判断 IGA 最优解 0.975 是否为局部最优,实验迭代次数依次从 30 次、40 次增加到 50 次。实验结果如图 6~10 所示,结果显示 IGA 全部收敛且最优解分别为 0.979、0.985、0.985。因遗传算法自身进化的随机性的特点,造成了收敛曲线趋势都不同,但均在误差允许范围内,且趋于稳定,最大程度保证了实验的可靠性。而对比算法中后期却存在收敛性不稳定、陷入局部最优、精度不高等问题。实验证明 IGA 达到了后期收敛速度快、跳出局部最优。同时转炉炼钢生产过程数据特征选择的平均用时对比实验结果如表 6 所示,其中本文所提方法在转炉炼钢生产过程数据特征选择的平均用时为 0.25 s,相比于其他改进遗传算法在特征选择计算速度上有一定优势,为实现转炉炼钢过程数据终点碳温实时预报提供速度保障。

3.3 基于特征选择的转炉炼钢终点碳温预报及对比实验

该部分实验用于验证本文特征选择方法在转炉炼钢终点碳温实时预报中的有效性,以终点碳温预报性能为评判指标并与同类方法进行对比。由于文献[11]同样是对转炉炼钢生产过程数据进行特征选择进而实现终点碳温预报研究,与本文具有相同的研究对象和目标,因此将其作为本节对比方法之一。同时,为了测试其他改进遗传算法在转炉炼钢终点碳温预报方面的性能,考虑到文献[24]在已有改进遗传算法中具有较好的实验性能,因此也将其作为本节用于碳温预报实验的对比方法之一。本文改进遗传算法最优特征选择结果如表 7 所示。

实验中训练样本为 1 000 炉次,测试样本 60 炉次。实验结果如图 11~16 所示,同时统计出终点温度预测误差分别在 $\pm 5^\circ\text{C}$ 、 $\pm 10^\circ\text{C}$ 、 $\pm 15^\circ\text{C}$ 时的命中率和终点碳含量预测误差在 $\pm 0.01\%$ 、 $\pm 0.02\%$ 、 $\pm 0.03\%$ 时的命中率结果如表 8 所示。

对比实验结果中终点碳温预报精度统计结果如表 8 所示,在 IGA 特征选择验证实验中,钢水预测温度误差在 $\pm 10^\circ\text{C}$ 时预测精度达到 93.67%,终点碳含量预测 $\pm 0.03\%$ 时预测精度达到 93.33%。相比文献[11]和文献[24]特征选择结果,IGA 特征选择的结果在终点碳、温预测实验中有明显优势,说明 IGA 对遗传算法目标函数优化精度更高,有力地证明了 IGA 特征选择的有效性,达到了转炉炼钢终点碳温实时连续预测的目的。

表 7 终点碳温预报模型的最优特征

Table 7 Optimal feature of the end-point carbon temperature prediction model

输入变量	输出变量
氧压 31	终点钢水碳含量
兑铁时长	
兑铁结束到开氧时间	
铁水温度	
枪位 36	
兑铁开始与上炉出钢时长	
装入废钢量	
装入生铁量	
铁水 P	
枪位 22	
氧压 29	终点钢水温度
氧压 11	
枪位 16	
氧压 34	

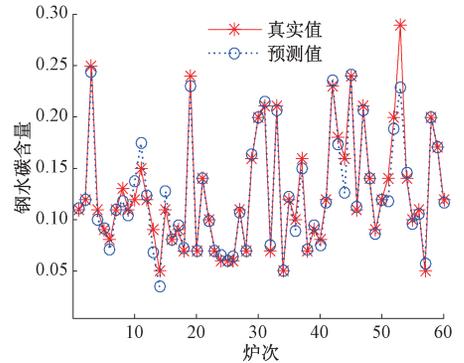


图 11 IGA 特征选择终点碳含量预测

Fig.11 Endpoint carbon content prediction based on the selected feature of IGA

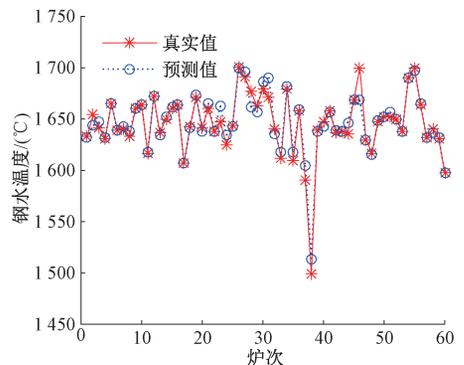


图 12 IGA 特征选择终点温度预测

Fig.12 Terminal temperature prediction based on the selected feature by IGA

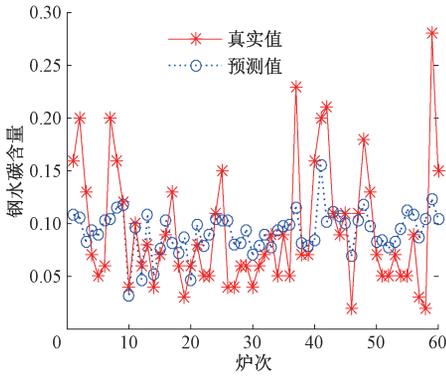


图 13 文献[11]特征选择终点碳含量预测

Fig.13 Terminal carbon content prediction based on the selected feature by reference [11]

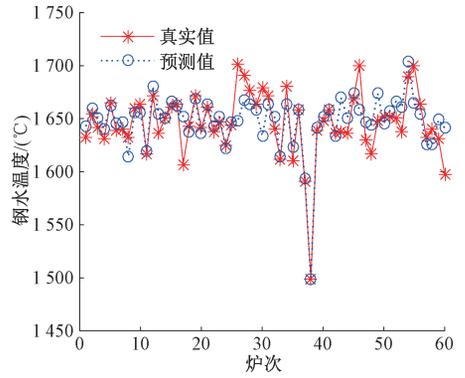


图 16 文献[24]特征选择终点温度预测

Fig.16 Terminal temperature prediction based on the feature selected by reference [24]

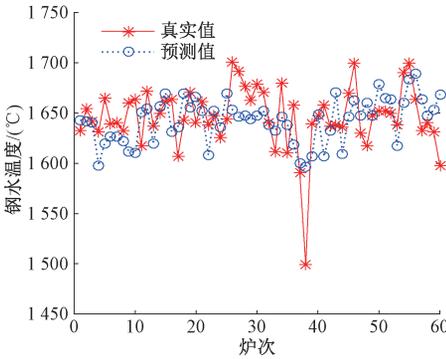


图 14 文献[11]特征选择终点温度预测

Fig.14 Terminal temperature prediction based on the selected feature by reference [11]

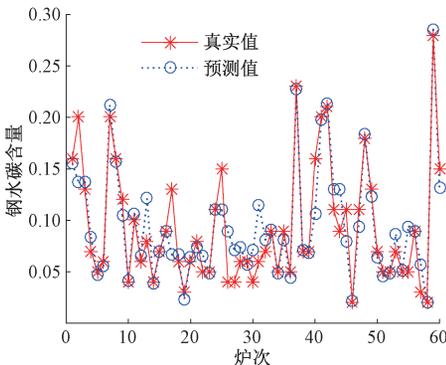


图 15 文献[24]特征选择终点碳含量预测

Fig.15 Terminal carbon content prediction based on the feature selected by reference [24]

表 8 验证对比实验终点碳温预测结果

Table 8 Prediction results verification and comparison of endpoint carbon temperature (%)

预测误差	预测精度		
	IGA	文献[11]	文献[24]
终点温度预测误差 $\pm 5^{\circ}\text{C}$	85.67	21.67	46.67
终点温度预测误差 $\pm 10^{\circ}\text{C}$	93.67	33.33	56.67
终点温度预测误差 $\pm 15^{\circ}\text{C}$	98.33	45.00	71.67
终点碳含量预测误差 ± 0.01	80.67	16.67	51.67
终点碳含量预测误差 ± 0.02	80.67	31.67	75.00
终点碳含量预测误差 ± 0.03	93.33	50.00	85.00

使算法能够很快地进入到最优值附近,并逼近最优精度,尽快跳出局部收敛。IGA 无论在收敛速度还是对目标函数优化的精度上都取得了良好效果。由此可见,本文提出的自适应交叉、变异概率模型对提高遗传算法的收敛速度、局部搜索能力、目标函数的优化精度是有效的。

4 结 论

本文提出一种改进遗传算法(IGA)的转炉炼钢过程数据特征选取方法,实现了对转炉炼钢终点碳温实时连续预测,具体内容描述如下:

1) 通过种群最大、最小适应度值、种群平均适应度值和种群随机个体适应度值 4 个变量作为自适应调节交叉变异概率模型的衡量标准,从而使种群个体适应度值快速达到最优效果。

2) 种群个体适应度值最优分布,加快了遗传算法后期收敛速度(缩短了特征选择时间),进而能跳出局部最优,提高寻优精度,有效选择出最优特征子集(因减少预测模型输入),从而降低预测模型时间,保证了预测精度。

以上实验结果表明,在进行特征选择时,IGA 在收敛性能、搜索能力、优化目标函数精度方面都有明显提高。可见,自适应交叉、变异概率让算法有效地解决了遗传算法中未成熟收敛的“早熟”问题,改进了局部搜索性能,

3) 实际钢厂转炉炼钢过程数据特征选择任务中进行验证分析。结果表明,改进后的遗传算法能够在精度和速度方面优于传统算法,实现对终点碳含量和温度的准确实时预报。

实际应用中,本文改进遗传算法能有效用于转炉炼钢生产数据关键特征的选择,从而为实现转炉炼钢计算机自动控制任务打下良好的基础。

参考文献

- [1] SHAO Y M, ZHOU M C, CHEN Y R, et al. BOF end-point prediction based on the flame radiation by hybrid SVC and SVR [J]. *Optik-International Journal for Light and Electron Optics*, 2014, 125(111): 2491-2496.
- [2] 温宏愿, 赵琦, 陈延如, 等. 基于炉口辐射和改进神经网络的转炉终点预测模型[J]. *光学学报*, 2008, 28(11): 2131-2135.
WEN H Y, ZHAO Q, CHEN Y R, et al. Basic oxygen furnace endpoint forecasting model based on radiation and modified ne-ural networ [J]. *Acta Optica Sinica*, 2008, 28(11): 2131-2135.
- [3] 温宏愿, 赵琦, 陈延如, 等. 基于辐射信息分析的转炉终点控制回归模型[J]. *仪器仪表学报*, 2008, 29(8): 1633-1637.
WEN H Y, ZHAO Q, CHEN Y R, et al. Converter end point control regression model based on radiation information analysis [J]. *Chinese Journal of Science Instrument*, 2008, 29(8): 1633-1637.
- [4] 韩敏, 张瑞全, 许美玲. 一种基于改进灰色关联分析的变量选择算法[J]. *控制与决策*, 2017, 32(9): 1647-1652.
HAN M, ZHANG R Q, XU M L. A variable selection algorithm based on improved gray relational analysis [J]. *Control and Decision*, 2017, 32(9): 1647-1652.
- [5] KUBAT C, TASKIN H, ARTIR R, et al. Bofy-fuzzy logic control for the basic oxygen furnace (BOF) [J]. *Robotics and Autonomous Systems*, 2004, 49(3-4): 193-205.
- [6] COXA I J, LEWIS R W, RANSING R S, et al. Application of neural computing in basic oxygen steelmaking [J]. *J of Materials Processing Technology*, 2002, 120(1-3): 310-315.
- [7] 谢书明, 陶钧, 柴天佑. 基于神经网络的转炉炼钢终点控制[J]. *控制理论与应用*, 2003, 20(6): 903-907.
XIE SH M, TAO J, CHAI T Y. BOF steel making endpoint control based on neural network [J]. *Control Theory and Applications*, 2003, 20(6): 903-907.
- [8] 谢书明, 柴天佑, 陶钧. 一种转炉炼钢动态终点预报的新方法[J]. *自动化学报*, 2001, 27(1): 136-139.
XIE SH M, CHAI T Y, TAO J. A new method of converter steelmaking dynamic end point prediction [J]. *Acta Automatica Sinica*, 2001, 27(1): 136-139.
- [9] 黄南天, 王达, 刘座铭, 等. 复杂噪声环境下电能质量复合扰动特征选择[J]. *仪器仪表学报*, 2018, 39(4): 82-90.
HUANG N T, WANG D, LIU Z M, et al. Power quality complex disturbance feature selection in complex noise environment [J]. *Journal of instrumentation*, 2008, 39(4): 82-90.
- [10] 盛云龙, 魏长安, 刘玉奇, 等. 时序约束条件下序列测试建模方法[J]. *仪器仪表学报*, 2019, 40(6): 213-220.
SHENG Y L, WEI CH AN, LIU Y Q, et al. Sequence test modeling method under timing constraints [J]. *Journal of Instrumentation*, 2019, 40(6): 213-220.
- [11] 王心哲, 韩敏. 基于变量选择的转炉炼钢终点预报模型[J]. *控制与决策*, 2010, 25(10): 1589-1592.
WANG X ZH, HAN M. A prediction model for converter steelmaking endpoint based on variable selection [J]. *Control and Decision*, 2010, 25(10): 1589-1592.
- [12] 代照坤. 面向转炉炼钢碳温预报模型的输入参量特征选择方法研究[D]. 昆明: 昆明理工大学, 2018.
DAI ZH K. Research on input parameter feature selection method for carbon steel temperature prediction model of converter steelmaking [D]. Kunming: Kunming University of Science and Technology, 2018.
- [13] 孙广路, 宋智超, 刘金来, 等. 基于最大信息系数和近似马尔科夫毯的特征选择方法[J]. *自动化学报*, 2017, 43(5): 795-805.
XUN G L, SONG ZH CH, LIU J L, et al. Feature selection method based on maximum information coefficient and approximate Markov blanket [J]. *Acta Automatica Sinica*, 2017, 43(5): 795-805.
- [14] 贾亚超, 李国龙, 何坤, 等. 基于灰色准则与 EEMD 的滚刀振动信号降噪方法[J]. *仪器仪表学*, 2019, 40(7): 187-194.
JIA Y CH, LI G L, HE K, et al. Noise reduction method of hob vibration signal based on grey criterion and EEMD [J]. *Chinese journal of instrumentation*, 2019, 40(7): 187-194.
- [15] 孟佳娜, 林鸿飞, 李彦鹏. 基于特征贡献度的特征选择方法在文本分类中应用[J]. *大连理工大学学报*, 2011, 51(4): 611-615.
MENG J N, LIN H F, LI Y P. Application of feature selection method to text categorization based on feature contribution degree [J]. *Journal of Dalian University of Technology*, 2011, 51(4): 611-615.
- [16] 魏东升, 周晓光. 顾及纹理特征贡献度的变化影像对象提取算法[J]. *测绘学报*, 2017, 46(5): 605-613.

- WEI D SH, ZHOU X G. Image Object extraction algorithm considering the variation of texture feature bution [J]. Acta Geochimica Sinica, 2017, 46 (5): 605-613.
- [17] 夏瑜, 吴小俊, 冯振华. 基于特征贡献度的 mean shift 视频跟踪算法 [J]. 控制与决策, 2012, 27 (7): 1021-1026.
- XIA Y, WU X J, FENG ZH H. Mean shift algorithm for visual tracking based on feature contribution [J]. Control and Decision, 2012, 27(7): 1021-1026.
- [18] 官伯林, 贾建援, 朱应敏. 基于自适应遗传算法的三轴光电跟踪策略 [J]. 仪器仪表学报, 2012, 33 (8): 1758-1764.
- GUAN B L, JIA J Y, ZHU Y M. Triaxial photo electric tracking strategy based on adaptive genetic algorithm [J]. Chinese journal of instrumentation, 2012, 33 (8): 1758-1764.
- [19] 王勇臻, 陈燕, 于莹莹. 求解多旅行商问题的改进分组遗传算法 [J]. 电子与信息学报, 2017, 39(1): 198-205.
- WAANG Y ZH, CHEN Y, YU Y Y. An improved grouping genetic algorithm for solving multi-traveling salesman problem [J]. Journal of Electronics & Information Technology, 2017, 39 (1): 198-205.
- [20] 赵世田, 赵东标, 付莹莹, 等. 基于改进遗传算法求取加工自由曲面的最大刀具尺寸 [J]. 计算机集成制造系统, 2011, 17(2): 289-293.
- ZHAO SH T, ZHAO D B, FU Y Y, et al. Maximum cutter size calculating for free-form surface machining based on improved gene-tic algorithm [J]. Computer Integrated Manufacturing Systems, 2011, 17 (2): 289-293.
- [21] 闫春, 厉美璇, 周潇. 基于改进的遗传算法在函数优化中的应用 [J]. 计算机应用研究, 2019, 36 (10): 2982-2985.
- YAN CH, LI M X, ZHOU X. Application of improved genetic algorithm in function optimization [J]. Application Research of Computers, 2019, 36(10): 2982-2985.
- [22] 王吉权, 程志文, 张攀利, 等. 求解有约束优化问题的实数遗传算法改进研究 [J]. 控制与决策, 2019, 34 (5): 937-946.
- WANG J Q, CHENG ZH W, ZHANG P L, et al. Research on improvement of real-coded genetic algorithm for solving constrained optimization problems [J]. Control and Decision, 2019, 34(5): 937-946.
- [23] 韩煜东, 董双飞, 谭柏川. 基于改进遗传算法的混装线多目标优化 [J]. 计算机集成制造系统, 2015, 21(6): 1476-1485.
- HAN Y D, DONG SH F, TAN B CH. Multiobjective optimization for mixed-model assembly line balancing problem based on improved genetic algorithm [J]. Computer Integrated Manufacturing Systems, 2015, 21(6): 1476-1485.
- [24] 刘二辉, 姚锡凡. 基于改进遗传算法的自动导引小车路径规划及其实现平台 [J]. 计算机集成制造系统, 2017, 23(3): 465-472.
- LIU E H, YAO X F. Path planning and its implementation platform of automatic guided vehicle based on improved genetic algorithm [J]. Computer integrated manufacturing system, 2017, 23(3): 465-472.
- [25] 宋建厚, 陈良琼, 刘道华. 改进的免疫算法参数自适应调整的优化设计 [J]. 计算机测量与控制, 2013, 21 (5): 1297-1300.
- SONG J H, CHEN L Q, LIU D H. Optimal design of improved immune algorithm parameters adaptive adjustment [J]. Computer measurement and control, 2013, 21(5): 1297-1300.

作者简介



刘辉, 2012 年于昆明理工大学获得博士学位, 现为昆明理工大学副教授。主要研究方向为数据分析、图像处理、模式识别。

E-mail: liuhui621@126.com

Liu Hui received his Ph. D. degree from Kunming University of Science and Technology.

He is currently an associate professor at Kunming University of Science and Technology. His main research interests include data analysis, image processing and pattern recognition.



曾鹏飞 (通信作者), 2017 年于华中科技大学武昌分校获得学士学位, 现为昆明理工大学硕士研究生, 主要研究方向为数据分析、图像处理、模式识别。

E-mail: 1431919050@qq.com

Zeng Pengfei (Corresponding author) received his B. Sc. degree from Huazhong University of Technology Wuchang Branch in 2017. He is currently a M. Sc. candidate at Kunming University of Science and Technology. His main research interests include data analysis, image processing and pattern recognition.