

• 应用与服役 •



基于一类对余类法的遗传算法优化决策树 漏钢预报模型研究

余浩辰¹, 张本国¹, 吴 恒¹, 张瑞忠²

(1 盐城工学院优集学院, 盐城 224051; 2 河钢集团钢研总院工艺研究所, 石家庄 050000)

摘 要:针对决策树模型在小样本训练数据的多分类问题上难以获得较高预报准确率的问题,建立了一种基于一类对余类法的遗传算法优化决策树漏钢预报模型,通过充分利用遗传算法的全局搜索能力以及其具有的鲁棒性,加强搜索控制和优化过程的监督,提高了模型的准确性。结合某钢厂连铸生产数据,对基于一类对余类法的遗传算法优化决策树漏钢预报模型进行了测试。测试表明,遗传算法仅在 10 次迭代后就可以使基于一类对余类法的遗传算法优化决策树漏钢预报模型达到 98.39% 的准确率和 100% 的报出率,该算法相比传统决策树算法可以在极少的迭代次数下得到更高的准确率和更好的泛化性。

关键词:遗传算法; 决策树; 连铸; 漏钢预报

DOI:10.20057/j.1003-8620.2023-00245 **中图分类号:**TF777; TP181

Research on a Steel Leakage Prediction Model Based on One vs Rest Genetic Algorithm Optimization Decision Tree

Yu Haochen¹, Zhang Benguo¹, Wu Heng¹, Zhang Ruizhong²

(1 School of Mechanical Engineering, Yancheng Institute of Technology, Yancheng 224051, China;

2 Institute of Materials Technology, HISCO Group, Shijiazhuang 050000, China)

Abstract: Aiming at the problem that the decision tree model of small sample training data is difficult to obtain a high prediction accuracy rate on multi classification problems, this paper establishes a breakout prediction model based on one vs rest genetic algorithm to optimize the decision tree. By making full use of the global search ability and robustness of genetic algorithm, strengthening the search control and supervision of the optimization process, the accuracy of the model is improved. Combined with the continuous casting production data of a steel plant, the breakout prediction model of genetic algorithm optimization decision tree based on one kind of congruence method was tested. The tests shows that genetic algorithms can achieve an accuracy of 98.39% and a reporting rate of 100% for the optimization decision tree steel leakage prediction model based on a class to class genetic algorithm after only 10 iterations. Compared with traditional decision tree algorithms, this algorithm can achieve higher accuracy and better generalization in very few iterations.

Key Words: Genetic Algorithm; Decision Tree; Continuous Casting; Steel Leakage Prediction

漏钢是连铸过程中非常严重的事故^[1],它会导致设备寿命降低、财物损失,并且降低企业的经济效益^[2-4]。为了预测漏钢,在工业领域广泛采用逻辑判断预报模型、神经网络预报模型以及热电偶测温法等方法^[5-6],如李珊^[7]在 BP 神经网络的训练过程中引入改进的粒子群算法,建立了基于粒子群优化算法的 BP 神经网络漏钢预报模型,段海洋等^[8]运用密度聚类方法聚集和分离正常工况、粘结漏钢状况下的温度样本,来达到漏钢预报的目的,可已知存在的漏钢预报模型需要大规模数据长时间训练且可解释性较差而无法广泛应用,而机器学习决策树

模型具有可在小样本中快速训练且决策规则简单易于理解的特性,因此,选用决策树模型来弥补现有模型的缺点。但是在实际生产过程中,工况复杂,简单的决策树模型并不能适应现今的生产实际,泛化性、准确性不高及容易过拟合等问题突出。针对这些问题本文引入遗传算法 (Genetic Algorithm, GA) 对决策树 (Decision Tree) 模型的一系列超参数进行寻优,使其在二分类问题上能够保证良好的性能,并用一类对余类 (One versus rest, OvR) 法将该模型从二分类模型转变为一个三分类模型。最后,用某钢厂现场连铸数据对模型进行训练和测

基金项目: 江苏省基础研究计划资助项目 (BK20150429)

作者简介: 余浩辰 (1998—), 男, 硕士; **E-mail:** yuhaochen077970@163.com; **收稿日期:** 2023-12-09

通信作者: 张本国 (1982—), 男, 副教授, 博士; **E-mail:** zhangbenguo@163.com

试,并以此来验证模型性能。

1 漏钢预报原理分析

漏钢预报方法主要有热流检测法、摩擦力测量法、热电偶测温法等^[9]。其中,热电偶测温法是漏钢预报领域中应用最广泛的方法,该方法通过检测安装在结晶器铜板上的热电偶温度数值,并分析其变化,以确定是否具有漏钢时的温度波形特性,继而实现漏钢发现与报警^[10-11]。图1中表明了,在正常浇注状况下,上排热电偶的监测温度远大于下排热电偶,并且下排热电偶的工作温度变动范围也较小。在结晶器坯壳进行粘接后,随着结晶器的上下抖动,粘接中心将向下移动,这导致上排热电偶的温度先升高后降低,下排热电偶的温度也会出现相应的波形变化,并且会出现温度“倒置”的现象^[12-13]。

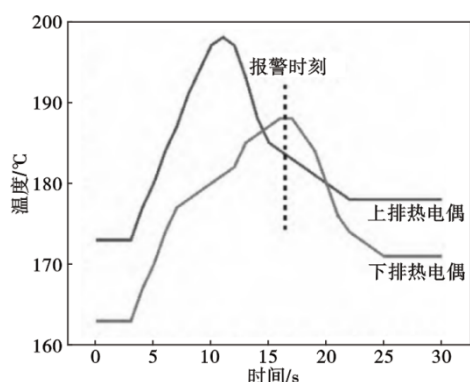


图1 粘结漏钢预报原理图

Fig. 1 Schematic diagram of bonding breakout prediction

2 基于SURE方法的小波阈值降噪

板坯连铸过程中存在许多干扰和噪声信号,限制了热电偶采集的温度数据对连铸坯表面品质的判断。因此,首先采用基于SURE方法的小波阈值降噪方法对温度数据进行降噪处理。小波阈值降噪的流程图如图2所示。

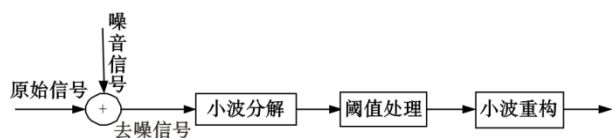


图2 小波阈值去噪流程图

Fig. 2 Flow chart of wavelet threshold denoising

SURE(Stein's unbiased risk Estimate)方法是小波去噪中一种常用的阈值选择方法,它的基本思想

是通过最小化估计软阈值方法进行小波去噪处理。在进行小波阈值分解的过程中,将其中的阈值处理设置为使用SURE方法选择最优阈值,接着使用软阈值方法进行小波去噪处理。信号降噪效果通常由信噪比(Signalnoise rate, SNR)和重构误差(Reconstruction Error, RE)两种参数表征,分别计算如下:

$$SNR = 10 \lg \left\{ \frac{\sum_{n=1}^N x^2(n)}{\sum_{n=1}^N [\hat{x}(n) - x(n)]^2} \right\} \quad (1)$$

$$RE = \frac{\|X - \text{denoised_}X\|}{\|X\|} \quad (2)$$

式中, $x(n)$ 为数字信号; X 为原始信号; $\text{denoised_}X$ 为经过小波去噪后的信号。

3 遗传算法优化决策树模型的建立

3.1 决策树模型

决策树算法是一种十分高效且易于理解的分类型算法^[14]。是一种基于概率分析的图解法,可以在已知情况发生概率的基础上预测净现值的期望值,从而评价项目风险和可行性。机器学习有着强大的数据处理和预测能力^[15],在机器学习中,决策树模型也是一个预测模型,能够帮助构建目标属性和对象值之间的映射关系。算法示意如图3所示。

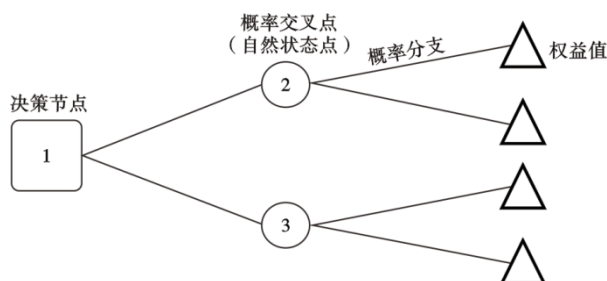


图3 决策树示意图

Fig. 3 Schematic diagram of the decision tree

在决策树的建立过程中,涉及到很多对当前结点集合的划分操作,为了提高决策树模型对漏钢样本的预测精度,通常希望分支节点所包含的样本属于同一类别,即节点纯度越高越好。

基尼指数: CART决策树在各类决策树中是唯一使用基尼指数(Gini index)来划分属性。

$$\text{Gini}(D) = 1 - \sum_{k=1}^{|Y|} P_k^2 \quad (3)$$

式中, P_k 为第 k 类子集在集合 D 中的占比。

直观来说, $\text{Gini}(D)$ 可以反映数据集的纯度, Gini指数越小代表数据集的纯度越高。属性 a 的基尼指数定义如式(4)。

$$\text{Gini_index}(D,a)=\sum_{v=1}^V \frac{|D^v|}{|D|} \text{Gini}(D^v) \quad (4)$$

式中, D^v 为第 V 个分支结点包含了 D 在所有属性 a 上取值为 a^v 的样本。

于是在候选属性集 A 中, 选择哪个使得划分后基尼指数最小的属性作为最优划分属性, 即:

$$a^* = \arg \min \text{Gini_index}(D,a) \quad (5)$$

3.2 遗传算法优化决策树

遗传算法(GA)是一种基于自然界遗传法则和适者生存规律而产生的算法^[16]。它根据了达尔文的进化论学说, 通过模仿自然选择、交叉和突变的过程, 不断优化个体的适应度, 从而找到最优解或者接近最优解的解决方案。在原理上突破了常规的优化算法框架, 适用于解决各种类型的优化问题, 包括连续优化问题、离散优化问题、多目标优化问题等, 具有很强的通用性。

遗传算法最主要是用来对决策树的各种超参数(叶节点最小样本数, 叶节点最大分裂次数, 父节点最小样本数, 每个节点的特征数)进行调优, 将寻得的最优的决策树超参数建立相关决策树的模型, 用来对漏钢情况发生的预报。遗传算法优化决策树详细流程如图4所示。

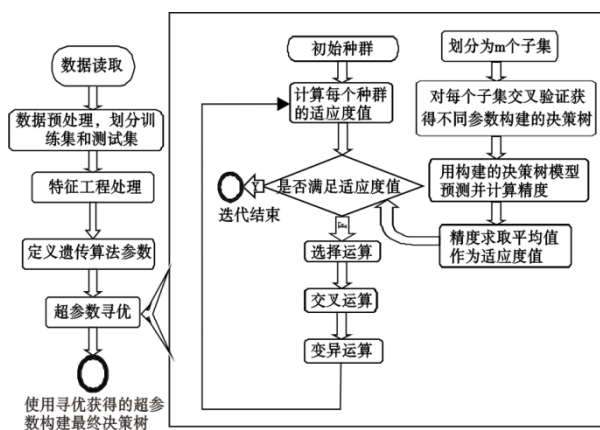


图4 遗传算法优化决策树流程图
Fig. 4 GA - Decision Tree Flowchart

图4所示遗传算法优化决策树流程, 通过选用交叉验证^[17]来对模型进行评估, 分析结果泛化到独立于训练数据的其他数据集, 通过对每个训练集计算准确率然后求取平均值, 最后, 以此结果来作为遗传算法的适应度值。

3.3 一类对余类法

一类对余类法(OvR)是最早出现也是目前在机

器学习、文本分类、图像识别领域应用最为广泛的方法之一, 在这种方法中, 对于一个有 N 个类别的问题, 通过训练 N 个二分类器, 每个二分类器都用来区分一个类别和其他所有类别的组合。当任务要对一个新的样本进行分类时, 用这 N 个二分类器分别对其进行分类, 然后选择得票最多的类别作为最终的分类结果。因此, 根据研究的三分类模型, 可以用一类对余类法设置两个决策树, 将决策树模型在二分类问题上的优良表现在三分类问题上充分发挥。一类对余类法的具体分类操作如图5所示。

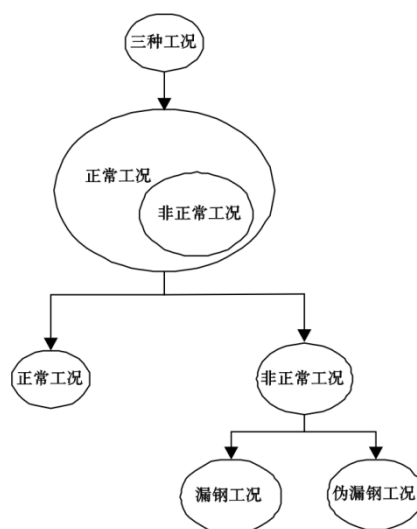


图5 OvR法流程图
Fig. 5 OvR method flowchart

如图5所示, 首先用CART决策树模型将训练集的三个工况分类为正常工况和非正常工况(漏钢工况和伪漏钢工况), 再用该模型对非正常工况进行二次划分以达到对三种不同工况进行分类学习的目的。然后用交叉验证法对模型训练结果进行评估, 选出遗传算法合适的适应度值。之后遗传算法根据该适应度值选出模型合适的参数组, 用该参数组构建新的决策树模型。最后针对测试集使用新构建的决策树同样使用二次划分的方法进行分类得出测试结果。综上, 模型整体流程如图6所示。

4 实验结果及分析

4.1 模型参数设置见表1.

4.2 小波阈值降噪实验对比

为验证SURE方法的降噪性能, 选用不同方法进行实验对比, 从某钢厂实际生产数据中, 截取了30 s(总共60个样本)的温度序列样本, 分别用

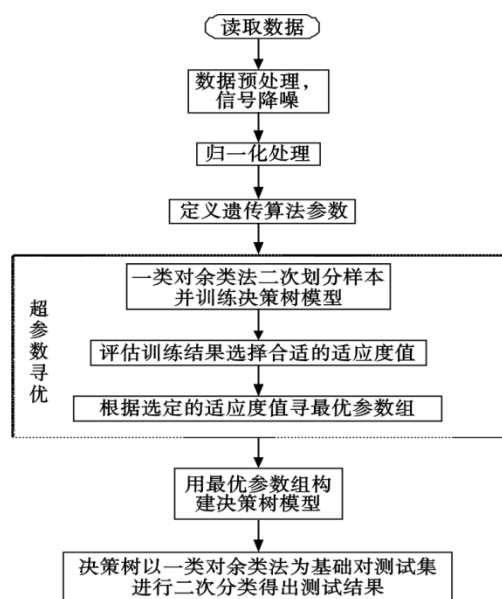


图6 模型流程图

Fig. 6 Model flowchart

SURE方法和Birge-Massart方法进行降噪,图7为两种方法降噪对比图。基于这两种方法降噪的对照详情见表2。

通过对信噪比和重构误差以及能量保留率的数据对比,基于SURE方法阈值降噪的信噪比以及能量保留率均高于基于Birge-Massart方法的阈值降

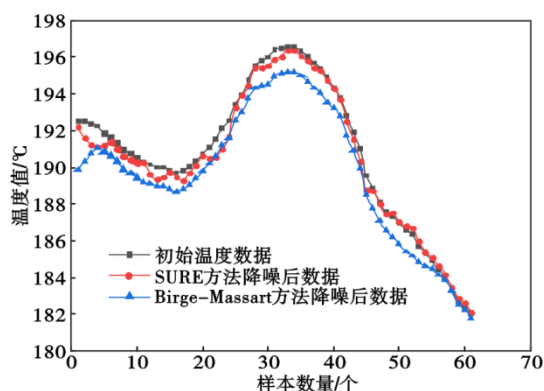


图7 两种方式降噪对比图

Fig. 7 Comparison of two methods for noise reduction

噪,降噪效果较好,且通过信号重构误差可以看出,SURE方法信号保留率更好。因此,基于SURE方法的小波阈值降噪在漏钢预报模型里有更好的降噪效果。

4.3 遗传算法优化决策树的模型实验效果

在本次课题中采用了一类对余类法对样本进行了二次分类,因此,随机选出其中一个分类决策树做实验验证可以得到10折交叉验证的平均准确率为95%,标准差为3.24%,这个结果表明,训练集与测试集的划分对正确率的影响是有限的,该模型在分类任务中表现良好,并且具有较好的稳定性和泛化能力。

在可视化方面,使用混淆矩阵图来直观地表示分类模型的性能表现,以便更好地进行分析和比较。实验结果混淆矩阵如图8所示。

通过分析该混淆矩阵图,可以看到模型在特定类别上的准确度和误分类情况。其中,工况1和工况2的发生次数均为20次,模型对两种工况的预报率和准确率都达到了100%。而最终模型的寻优结果见表3。

基于以上结果,可以进一步评估遗传算法优化决策树的漏钢预报模型。为了更全面地了解模型的性能。可以通过绘制ROC曲线来观察模型在不同阈值下的真阳性率(TPR)和假阳性率(FPR)之间的关系。与此同时,可以计算模型的AUC值。AUC是ROC曲线下的面积,它可以提供一个综合评估模型分类能力的指标。AUC值越接近1,表示模型的分类能力越强。通过计算AUC值,可以更直观地了解遗传算法优化决策树的漏钢预报模型在整体数据样本中的表现结果。ROC曲线如图9所示。从图9可以看出,AUC值为0.996 47,表示分类模型的性能非常高。

4.4 整体模型实验结果及对照

为测试模型性能,采用309组样本(每个样本28个特征)进行测试,随机选取了数据集80%为训练

表1 三类模型参数设置

Table 1 Three types of model parameter settings

遗传算法		降噪模型		决策树	
项目	参数	项目	参数	项目	参数
种群数目/个	300	小波基	coif1	叶节点最小样本数/个	120
迭代次数/次	10	分解层数/层	4	叶节点最大分裂次数/个	1 100
交叉概率/%	80			父节点最小样本数/个	120
变异概率/%	10			每个节点的特征数/个	1 100
K折交叉验证/折	10				

表 2 两种降噪方法效果对照表

Table 2 Comparison of the effects of two noise reduction methods

	信噪比/ dB	重构误差/°C	能量保留 率/%	标准差比
SURE	56.21	0.002 3	99.88	0.99
Birge-Massart	42.21	0.007 8	96.23	0.96

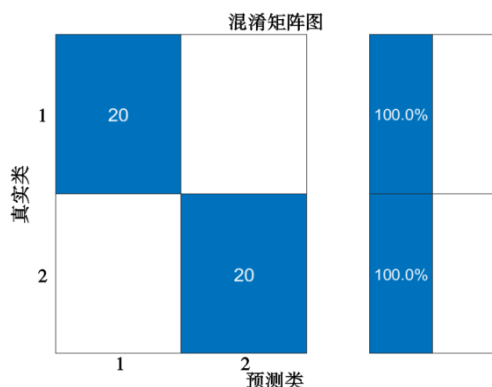


图 8 混淆矩阵图

Fig. 8 Confusion matrix diagram

表 3 决策树寻优结果

Table 3 Decision tree optimization results

叶节点最小样本数	叶节点最大 分裂次数	父节点最小 样本数	每个节点的特征数
15.279 028 525 569 6	1	1	1

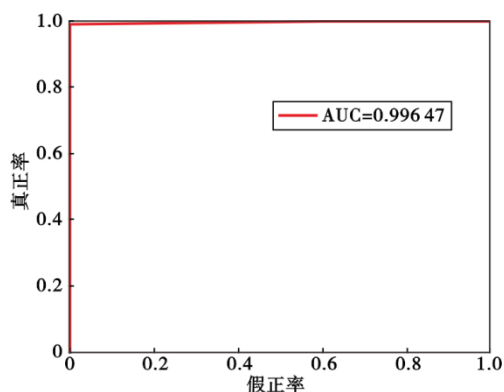


图 9 ROC 曲线图

Fig. 9 Receiver operating characteristic curve graph

集, 20% 为测试集即测试样本数量为 62 组, 测试样本中包含了 20 个非粘结样本, 20 个粘结样本以及 22 个正在发生粘结漏钢的样本。根据预测结果显示, 基于 One-vs-Rest 的遗传算法优化决策树漏钢预报模型(模型 1)的准确率高达 98.39%, 其中, 错误预测的样本是将正常工况的样本错误预测为已经发生漏钢的工况。而 BP 神经网络模型(模型 2)的预测准确率仅为 65%, 普通决策树模型(模型 3)的预

测准确率为 93.55%, 遗传算法优化决策树基础模型(模型 4)的准确率为 67.7%, 三种模型对于故障样本的预报率均不如模型 1, 具体对比结果见表 4。

表 4 不同模型分类结果对照表

Table 4 Comparison table of classification results of different models

模型	准确率 ^a /%	预报率 ^b /%
1	98.39	100
2	65.00	65
3	93.55	90
4	67.70	70

表中, a 为测试集预测正确样本数/测试集总样本数; b 为正确预测的漏钢工况样本数/测试集漏钢工况样本数。

由对照结果可知, 遗传算法优化决策树的漏钢预报模型在同类型模型中对漏钢样本的检测分类更敏感, 并在整体的数据样本训练测试结果中有更好的表现, 且相比于同类型模型更加具有应用价值。

由于该模型并没有采取限制树的深度这一方式来防止过拟合, 也为了进一步评估模型的泛化性能, 分析模型在不同数量样本下测试报错和训练报错的变化曲线如图 10 所示。

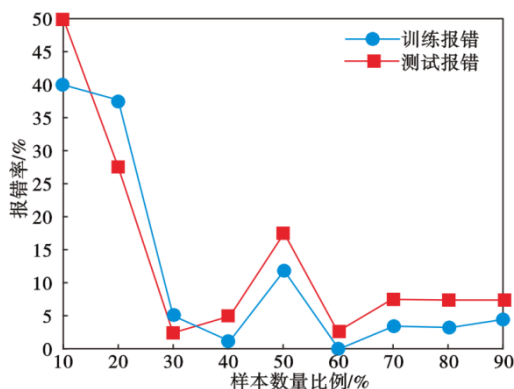


图 10 测试报错和训练报错的变化曲线图

Fig. 10 Change curve of test error and training error

从图 10 中可以观察到, 在 10 次迭代寻优的基础上, 随着样本数量的增加, 模型的训练报错和测试报错随着样本数量的增加而减小并趋于稳定。这表明模型具有较好的学习能力和泛化能力。

5 结论

(1) 利用遗传算法超空间网络寻优能力对决策树的各项超参数寻最优解, 建立了基于一类对余类

法的遗传算法优化决策树漏钢预报模型,弥补了决策树模型在数据复杂时无法对多个分类目标同时做到高预报率和高准确率缺点,同时,用交叉验证评估模型和报错曲线图,防止了决策树在某些情况下出现过拟合等问题。

参考文献

- [1] 李小明,史雷刚,崔雅茹,等. H型钢连铸质量缺陷和浸入式水口结构优化的数值模拟[J]. 特殊钢, 2013, 34(5): 9-12.
- [2] Saha Roy P D, Tiwari P K. Knowledge discovery and predictive accuracy comparison of different classification algorithms for mould level fluctuation phenomenon in thin slab caster[J]. Journal of Intelligent Manufacturing, 2019, 30(1): 241-254.
- [3] Zhang Y X, Wang W L, Zhang H H. Development of a mold cracking simulator: The study of breakout and crack formation in continuous casting mold[J]. Metallurgical and Materials Transactions B, 2016, 47(4): 2244-2252.
- [4] 王叶婷,赵洪强,国兴龙,等. 连铸粘结漏钢成因机理分析[J]. 连铸, 2011, 36(S1): 134-137.
- [5] 吴迪,张本国,生万宝,等. 基于差分进化-灰狼优化算法的支持向量机连铸漏钢预报系统研究[J]. 特种铸造及有色合金, 2023, 43(2): 174-177.
- [6] 苏春阳,陈君,姜亚清. 基于PCA-BP神经网络的LF精炼终点温度预测[J]. 特殊钢, 2023, 44(6): 39-44.
- [7] 李珊. 基于粒子群优化算法的BP神经网络漏钢预报模型[D]. 秦皇岛:燕山大学, 2014.
- [8] 段海洋,王旭东,姚曼. 基于密度聚类 and 动态时间弯曲的结晶器黏结漏钢预报方法的开发[J]. 工程科学学报, 2020, 42(3): 348-357.
- [9] 孙立根,张家泉. 基于逻辑判断的板坯漏钢预报系统研究[J]. 冶金自动化, 2009, 33(1): 16-20+25.
- [10] 秦旭,朱超甫,高光河,等. 板坯漏钢机理及基于热电偶检测的漏钢预报技术研究[J]. 钢铁研究学报, 2011, 23(2): 7-11.
- [11] 张勇,张慧,袁林涛,等. 板坯漏钢预报系统的研究及应用[J]. 钢铁研究, 2013, 41(5): 23-26.
- [12] 徐海伦,马春武,李清忠,等. 结晶器漏钢过程解析及预报原理[J]. 钢铁钒钛, 2012, 33(5): 35-40.
- [13] 生万宝,张本国,吴迪,等. 基于蚁群算法-BP神经网络的漏钢预报模型研究[J]. 特种铸造及有色合金, 2022, 42(11): 1366-1369.
- [14] 郑力嘉,宋冰. 决策树分类算法的预剪枝与优化[J]. 自动化仪表, 2023, 44(5): 56-62.
- [15] 温湖峰,尚天帅,李剑,等. 基于决策树方法的奇A核基态自旋预测[J]. 物理学报, 2023, 72(15): 27-35.
- [16] 何享,张笑非. 基于OVR-SVM及fMRI的词语识别认知状态分类研究[J]. 计算机与数字工程, 2021, 49(6): 1158-1162.
- [17] 王雨芹. 遗传算法在辐射防护中的应用[J]. 电子技术与软件工程, 2022, 232(14): 143-146.

欢迎订阅 2024 年《特殊钢》期刊

全国各地邮局均可订阅(可破订)

邮发代号:38-183 定价:40.00元/期、240.00元/年 邮编:435001

地址:湖北省黄石市黄石大道316号、大冶特殊钢有限公司《特殊钢》杂志社

电话:0714-6297386 0510-80673212

投稿方法及注意事项

为推进《特殊钢》期刊信息化,便于作者投稿查稿/专家审稿,由中信泰富特钢集团股份有限公司主管,大冶特殊钢有限公司主办的《特殊钢》杂志,实行作者网上投稿查稿、专家网上审稿的办刊模式。

1. 投稿注意事项

1.1 投稿前需按照本刊下载中心“杂志投稿模板及要求”对文稿进行修改。不尽之处可参考近期出版的《特殊钢》论文;

1.2 应在网络投稿时,准确、完整地填写所有署名作者的信息(姓名、单位(包括二级单位)应写全称、城市及邮编);无高级职称在读硕士研究生的文章需注明导师姓名,提供本人及导师的信息:姓名-出生年-性别-学历-职称--邮箱;

1.3 请填写中图分类号 TF 或 TG 类,中国图书馆分类法链接: <http://www.ztflh.com/>,通常在“工业技术”中的“TF—冶金工业”或“TG—金属学与金属工艺”栏目;

1.4 原稿必须是未在中外文正式刊物上发表的论文。本刊严禁一稿多投、重复内容多次投稿、不同文种重复投稿。一旦发现上述情况,稿件将按退稿处理,并将通知作者单位及材料冶金领域有关期刊。作者本人的稿件今后将不被录用;

1.5 来稿文责自负,编辑部有权对稿件做技术性、文字性修改,在征得作者同意后可以进行实质内容的修改。

2. 投稿方法

2.1 搜索《特殊钢》杂志官网或输入网站域名 <http://www.specialsteeljournal.com>,登录“特殊钢”网站。点击“作者登录”,进入“作者工作桌面”。

2.2 第一次在线投稿的作者需要先注册,取得用户名和登录密码。系统会自动将注册成功的信息发送至您提供的 E-mail 地址。以后投稿或查询稿件时均用此用户名和登录密码登录(登录后可修改用户名和登录密码)。如果忘记用户名和登录密码,可与编辑部联系,进行查询。投稿时登录或注册的作者被系统视为通信作者,有关稿件的事宜将与该人联系。此通信作者和论文第一页页脚备注的通信作者可以不同。

2.3 网上投稿前最好预先打开欲投稿件的电子文档,以便填写稿件信息时将文章标题、摘要、关键词等内容复制并粘贴到投稿系统中,减少录入的工作量。

2.4 用鼠标左键单击“投新稿件”的“向导式投稿(建议)”或者“列表式投稿”按钮,进入本刊的投稿系统。阅读“投稿确认书”并同意签署“著作权转让协议”后逐项填写稿件信息,然后上传稿件的电子文档(Word 或 PDF 格式)。对稿件信息进行核实,确认无误后点击“立即投稿”,此时系统会出现“投稿成功”的界面。随后,系统会自动发送收稿通知至您在注册时提供的 E-mail 地址,告知您的稿件编号等信息。

3. 网上查稿

3.1 稿件审查结果在 2 个月内通知作者,在此期间,作者不得将稿件投往他处。个别稿件可能送审时间较长,如果作者决定改投他刊或退稿,请通知编辑部后,再进行处理。编辑部决定录用稿件后,会及时通知作者;

3.2 搜索《特殊钢》杂志官网,或输入网站域名 <http://www.specialsteeljournal.com>,登录“特殊钢”网站。用鼠标点击“作者登录”,进入“作者中心”。

3.3 用鼠标点击“我的稿件”下面的不同按钮,即可看到您所投稿件的基本信息和目前的处理状态。

4. 稿件网上退修

4.1 编辑部收到作者的稿件入库后要经过三审,审稿过程中责任编辑会将稿件退修要求通过本刊网站向作者反馈,同时会发送提醒邮件至您提供的 E-mail 地址,请作者及时按要求修稿并重新提交修改稿。

4.2 修改稿请注明稿件编号,同时说明进行了哪些修改。每次修改的说明和修改稿都应留备份,以防丢失。修改稿需待编辑部讨论后,方能确定是否刊用。

4.3 退修稿件的修改时限为 10 日,编辑部将以作者返回修改稿日期作为投稿日期。如果需延期请事先联系编辑部,否则,超过 30 日,编辑部有权对稿件做出退稿处理。