

DOI:10.19344/j.cnki.issn1671-5276.2026.04.004

基于 MIC-RF 随机森林的热轧带钢力学性能预测

王力¹, 朱启宁¹, 杨洪凯², 张磊¹

(1. 沈阳化工大学 机械与动力工程学院, 辽宁 沈阳 110142;

2. 首钢京唐钢铁联合有限责任公司, 河北 唐山 063205)

摘要:为满足日益增长的热轧带钢力学性能预测精度要求,基于最大互信息系数和随机森林方法,以屈服强度和抗拉强度为研究对象,构建热轧带钢力学性能预测模型。利用 Pauta 法则对实际热轧数据进行预处理,去除数据中的异常值。采用最大互信息系数方法筛选重要影响特征作为模型的输入变量。利用随机森林算法训练模型,并与未进行特征选择的随机森林模型进行对比。结果表明,基于 MIC-RF 模型的屈服强度和抗拉强度预测误差均小于 1%,决定系数均在 0.99 以上,且预测精度优于未进行特征选择的随机森林模型。所提出的模型具有更高的预测准确率和较强泛化能力,可为带钢产品的设计和 optimization 提供直观可靠的参考。

关键词: MIC 最大互信息系数; RF 随机森林; 热轧带钢; 力学性能预测

中图分类号: TG334.9 **文献标志码:** A **文章编号:** 1671-5276(2026)04-0019-05

Prediction of Mechanical Properties of Hot-rolled Strip Steel Based on MIC-RF Algorithm

Wang Li¹, Zhu Qining¹, Yang Hongkai², Zhang Lei¹

(1. School of Mechanical and Power Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;

2. Shougang Jingtang United Iron & Steel Co., Ltd., Tangshan 063205, China)

Abstract: To meet the increasing requirements for prediction accuracy of mechanical properties of hot-rolled strip steel, a prediction model for the mechanical properties of hot-rolled strip steel was constructed based on the MIC-RF method, with yield strength (R_e) and tensile strength (R_m) as the research objects. The Pauta criterion was used to preprocess the actual hot-rolling data and remove outliers. The maximum mutual information coefficient method (MIC) was employed to screen important influencing features as input variables for the model. The random forest algorithm was used to train the model and was compared with a random forest model without feature selection. It is shown that the prediction errors for R_e and R_m based on the MIC-RF are both less than 1%, with coefficients of determination above 0.99, and the prediction accuracy is superior to that of the random forest model without feature selection. The proposed model has higher prediction accuracy and strong generalization ability, providing an intuitive and reliable reference for the design and optimization of strip steel products.

Keywords: MIC (maximum mutual information coefficient); RF (random forest); hot-rolled steel; mechanical property prediction

0 引言

近几年来,对钢材力学性能的研究已成为钢铁冶金企业重点关注的领域。通过对钢材力学性能的高精度预测,可对轧制工艺参数和化学成分进行反向调优,进而提高钢材生产质量,增加成材率,从而减少公司的损失。因此,对钢材的力学性能预测研究具有重要意义^[1-4]。

近年来,随着大数据的发展、计算机性能的提

高、人工智能的出现以及神经网络的提出和发展,利用这些有效的技术支持,进行热轧带钢力学性能的研究已成为当前的主流。如 2013 年, Zhao 等^[5]基于工业大数据和半参数单指数方法建立了热轧带钢力学性能预测模型,并与神经网络模型进行对比,证明了所建模型具有更强适应性。2018 年以来, Wu 等^[6-8]针对原始数据分布不均衡的问题,提出了采用均匀网格技术用于平衡数据,利用贝叶斯优化深度神经网络、人工神经网络等

基金项目: 国家自然科学基金资助项目 (U21A20117); 辽宁省人工智能领域科技创新项目 (重大专项项目) (2023JH26/10100002); 中国五矿科技专项项目 (2022ZX803)

第一作者简介: 王力 (1986—), 男, 讲师, 博士, 研究方向为板带轧制过程智能控制、模型优化、故障诊断等, wangli@syuct.edu.cn。

通信作者简介: 张磊 (1982—), 男, 讲师, 博士, 研究方向为智能控制、故障诊断等, zhang_lei@syuct.edu.cn。

方法对热轧带钢的力学性能进行高精度预测。文献[9-12]将大数据与冶金机制相结合,通过使用贝叶斯网和因果图等方法对重要特征进行筛选,构建一元无交互作用和二元有交互作用的广义可加性模型,并据此建立了大量含铌微合金钢的屈服强度、抗拉强度和延伸率的预测模型,并取得较好结果。2024年魏宝民等^[13]利用BP神经网络算法初步建立屈服强度、抗拉强度和延伸率的预测模型,同时利用平整轧制机制模型反算变形抗力,据此对力学性能进行二次修正。该研究实现了神经网络模型与物理冶金模型相结合的热镀产品力学性能预报,有效提高了实际工况下力学性能参数的预报精度。

以上方法对热轧带钢的力学性能预测起到了关键的作用,但在实际应用过程中,因轧制工艺参数和化学成分等特征数量过于众多,过多的特征变量对模型的预测精度具有较大影响。数据维度过高会造成数据冗余,继而使模型的预测精度大大降低。因此,本文提出了将最大互信息系数(maximum mutual information coefficient, MIC)法与随机森林(random forest, RF)算法相结合的方法。通过MIC法筛选出重要特征作为RF随机森林的输入变量,并据此建立精度可靠的热轧带钢力学性能预测模型,为力学性能的精准调控提供科学依据。

1 基本算法

1.1 RF 随机森林算法

RF 随机森林是一种基于集成学习的机器学习算法,通过有放回抽样和随机选择特征生成大量相互独立的决策树^[14-16]。每棵决策树都使用随机选择的数据样本和特征进行训练,从而降低了模型的方差,提高了泛化能力。在进行预测时,随机森林将所有决策树的预测结果取平均值作为最终的预测结果,这样可以有效减少过拟合的风险,提高模型的稳定性和准确性。其具体流程如图1所示。

1.2 MIC 最大互信息系数法

原始数据往往具有高维性,如果将所有特征参数直接作为模型的输入变量,则会因引入无关特征变量而降低模型的精度,出现过拟合等问题,因此需要对原始数据集进行重要特征选择。MIC法可针对变量之间的线性或非线性函数关系进行深入挖掘,且复杂程度低、鲁棒性高^[17]。具体流

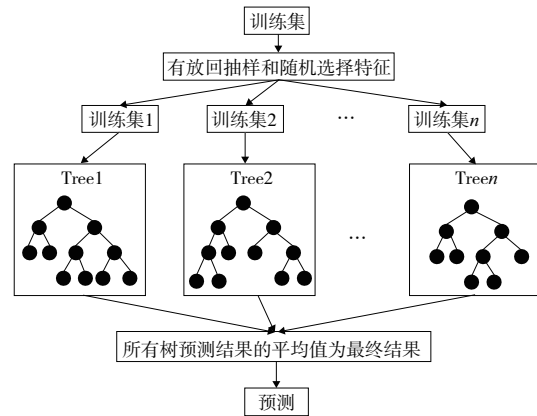


图1 随机森林算法流程图

程如下

1) 初假定 X 为数据集中某一特征变量, Y 为输出变量, 其中 $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$, 则定义 X, Y 间互信息为

$$I(x, y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy \quad (1)$$

式中: $p(x, y)$ 为联合概率密度; $p(x)$ 、 $p(y)$ 分别为边缘概率密度。

2) 在 X, Y 组成的数据散点图上绘制网格, 并计算各网格之间的互信息大小。使用不同的网格划分标准选取互信息的最大值, 计算公式为

$$\text{MIC}(x, y) = \max_{a \times b < B} \frac{I(x, y)}{\log_2 \min(a, b)} \quad (2)$$

式中: $\text{MIC}(x, y)$ 为 X 与 Y 最大互信息系数; a, b 分别为 X, Y 方向上划分网格数; B 为样本量的 0.6 次方。

3) 再次重复以上步骤, 直至计算出所有特征与输出变量 MIC 值。

2 数据预处理及重要特征筛选

2.1 实验数据来源

本文以某轧钢厂生产的 C-Mn 钢为例, 采用多种牌号钢的实际生产数据作为数据样本, 通过工厂控制系统的数据采集系统共获取到轧制数据 4 697 组。RDH 为粗轧出口厚度; FDH 为终轧厚度; RDT 为粗轧出口温度; FDT 为终轧温度; CT 为卷取温度; v 为层流冷却平均冷速; $P_{(F4-F6)}$ 为 F4 机架到 F6 机架总压下率; $w_{(C)}$ 、 $w_{(Si)}$ 、 $w_{(Mn)}$ 、 $w_{(P)}$ 、 $w_{(S)}$ 、 $w_{(V)}$ 、 $w_{(Ti)}$ 、 $w_{(Al)}$ 、 $w_{(Cu)}$ 分别为各合金元素的质量分数; Re 为屈服强度; Rm 为抗拉强度。其生产数据组成如表 1 所示。

表 1 C-Mn 钢生产数据组成表

变量	最小值	最大值	平均值
RDH/mm	28.45	40.61	36.92
FDH/mm	1.58	7.34	3.01
$RDT/^\circ C$	923.33	1 105.67	1 001.18
$FDT/^\circ C$	835.37	898.55	870.97
$CT/^\circ C$	491.93	696.50	593.77
$v/(^\circ C \cdot s^{-1})$	15.525	105.669	56.173
$P_{(F4-F6)}$	0.351 97	0.606 59	0.516 28
$w_{(C)}/\%$	0.015 9	0.180 8	0.079 5
$w_{(Si)}/\%$	0.004	0.378	0.085
$w_{(Mn)}/\%$	0.178	0.563	0.380
$w_{(P)}/\%$	0.007	0.095	0.022
$w_{(S)}/\%$	0.001 7	0.025 5	0.011 7
$w_{(V)}/\%$	0	0.005 4	0.002 5
$w_{(Ti)}/\%$	0	0.009 1	0.002 5
$w_{(Al)}/\%$	0.007 6	0.066 8	0.036 9
$w_{(Cu)}/\%$	0.003 7	0.322 5	0.039 2
Re/MPa	184	450	309
Rm/MPa	295	560	407

2.2 数据归一化

为解决特征间存在规模和量纲不同的问题,本文采用归一化方法将特征值转化为无量纲数值,使其变换后落入 $[0,1]$ 区间内^[18]。其计算公式如下

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

式中: x_i^* 为归一化后数据; x_i 为原始数据; $x_{\min}$ 为最小值; $x_{\max}$ 为最大值。

2.3 去除异常值

为消除异常值对模型精度的负面影响,本文采用 Pauta 准则去除异常值,该方法适用于数量较大的数据样本,能够有效剔除样本中存在的异常值^[19-20]。其计算公式如下

$$|a_i - \bar{a}| > 3\sigma \quad (4)$$

$$\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i \quad (5)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \bar{a})^2} \quad (6)$$

式中: a_i 为第 i 个样本值; $\bar{a}$ 和 σ 分别为所有样本的平均值和标准差; n 为样本总量。

Pauta 准则以 3 倍样本值的标准差作为检测

异常值的标准,满足这个条件即满足式(4)即可判断该值为异常值,需予以剔除。剔除异常值 399 组,剩余 4 298 组数据用于后期模型的建立。数据样本异常值分析结果如图 2 所示。

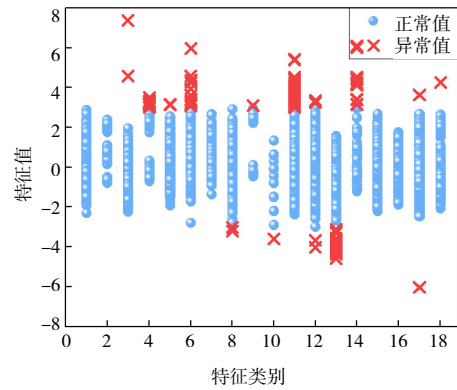


图 2 数据样本异常值分析

2.4 基于 MIC 最大互信息系数的重要因素筛选

根据 MIC 计算流程,分别计算特征变量与屈服强度和抗拉强度的 MIC 值。特征变量与屈服强度(Re)和抗拉强度(Rm)的 MIC 值统计如图 3、图 4 所示。一般认为,当 $MIC > 0.7$ 时两个变量之间具有较强相关性。图中 MIC 值大于 0.7 结果不足一半,为防止特征丢失改用平均值法对输入变量进行取舍。将大于均值的变量予以保留,将 MIC 小于均值的变量舍去。经上述规则筛选后,最终确定 Re 和 Rm 模型的输入变量如表 2 所示。

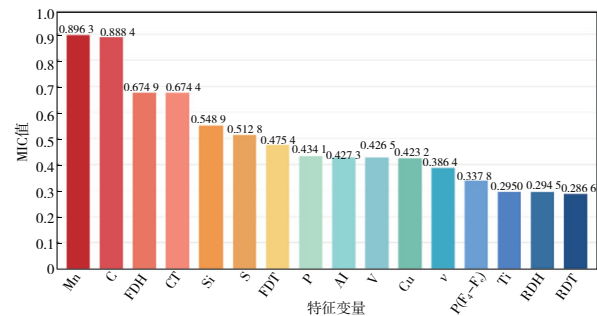
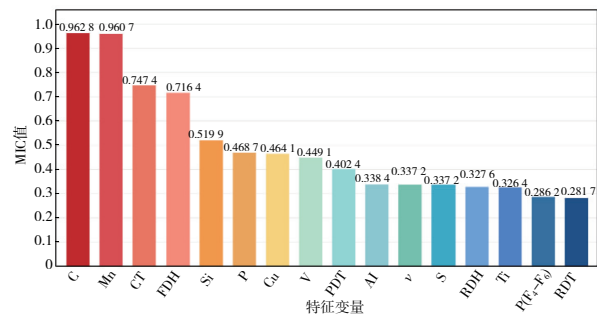
图 3 特征变量与 Re 的 MIC 值图 4 特征变量与 Rm 的 MIC 值

表 2 Re 、 Rm 预测模型的输入变量

输出变量	输入变量
Re 屈服强度	FDH 、 CT 、 $w_{(Mn)}$ 、 $w_{(C)}$ 、 $w_{(S)}$ 、 $w_{(Si)}$
Rm 抗拉强度	FDH 、 CT 、 $w_{(Mn)}$ 、 $w_{(C)}$ 、 $w_{(Si)}$

3 实验与结果

3.1 MIC-RF 模型的建立

将 4 298 组实测轧制数据按 7 : 3 划分为训练集和测试集,依据上述 MIC 筛选的模型输入变量,利用 MATLAB 分别建立 Re 和 Rm 预测模型。评价指标采用平均绝对百分比误差 (MAPE)、平均绝对误差 (MAE)、均方根误差 (RMSE) 和决定系数 (R^2),计算公式如下式所示

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

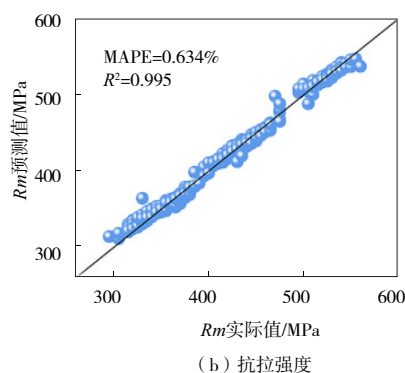
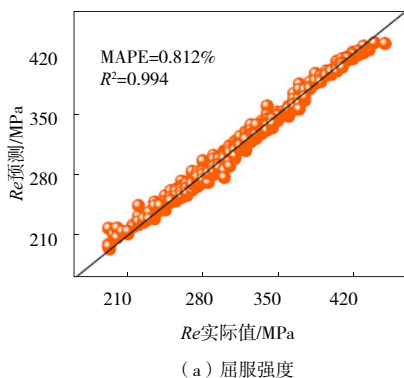


图 5 MIC-RF 实际值与预测值相关性

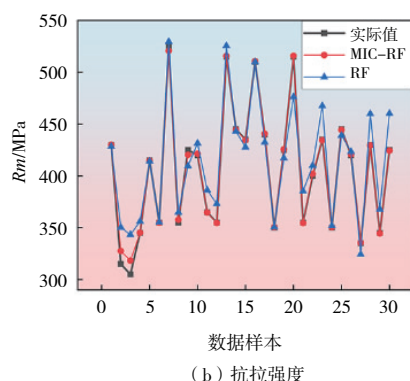
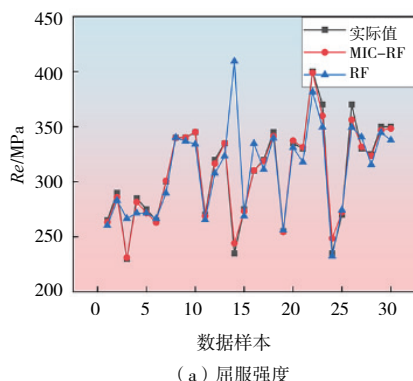


图 6 MIC-RF 模型与传统 RF 模型预测结果对比图

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

式中: n 为样本数量; $\hat{y}_i$ 为数据预测值; y_i 为真实值; $\bar{y}$ 为平均值。

同时,为验证经变量筛选后模型的预测性能,将变量筛选后的预测模型和未经变量筛选的预测模型分别与真实值进行对比。为控制变量,两种模型均采用相同的训练集和测试集来训练和测试模型,并采用上述指标对模型进行评价。

3.2 模型的验证与对比

将训练后得到的 MIC-RF 预测模型分别对其测试集样本进行预测,此时模型预测值与真实值相关性如图 5 所示。将变量筛选后的 MIC-RF 预测模型和未经变量筛选的传统 RF 预测模型分别与真实值进行对比如图 6 所示,其测试集误差统计对比结果如表 3 所示。

表 3 力学性能预测模型误差统计

力学性能	评价指标	MIC-RF	传统 RF
R_e 屈服强度	MAPE/%	0.812	2.065
	MAE	2.406	6.032
	RMSE	3.778	10.144
	R^2	0.994	0.956
R_m 抗拉强度	MAPE/%	0.634	1.309
	MAE	2.535	5.235
	RMSE	3.924	8.544
	R^2	0.995	0.975

由图 5 可知,基于 MIC-RF 的屈服强度和抗拉强度预测模型的平均绝对百分误差分别为 0.812%、0.634%,决定系数分别为 0.994、0.995。证明模型具有较高的精度以及较强的泛化能力。由表 3 和图 6 可知,本文所建 MIC-RF 预测模型在与未经变量筛选的传统 RF 预测模型对中,对于 R_e 和 R_m ,MIC-RF 模型比传统 RF 模型的误差分别低 1.253 个百分点和 0.675 个百分点,且泛化能力均提高了 3% 左右。证明经 MIC 变量筛选后的模型具有较高的预测精度及较强的泛化能力,且优化效果明显。综上所述,本文所建立的 MIC-RF 预测模型能够取得对钢材屈服强度和抗拉强度的精准可靠的预测效果,可为带钢力学性能预测提供一种新的方法。

4 结语

1) 对 C-Mn 钢热轧数据,采用 Pauta 法则进行异常值筛选,并利用 MIC 法进行特征选择,筛选重要影响特征作为模型的输入变量,为后续建立高精度模型提供了条件。

2) 利用 MIC-RF 算法对热轧带钢的力学性能建立了预测模型。结果表明:该模型对屈服强度预测误差为 0.812%,抗拉强度预测误差为 0.634%, R^2 均在 0.99 以上。该模型可对热轧带钢力学性能进行精准可靠的预测且具有一定泛化能力。

3) 将 MIC-RF 预测模型与未进行数据特征选择的传统 RF 随机森林模型进行对比。结果表明:对于屈服强度和抗拉强度,MIC-RF 预测模型的预测误差分别比未特征选择的传统 RF 预测模型低 1.253 个百分点、0.675 个百分点。MIC-RF 模型的预测精度明显优于未进行特征选择处理的

传统 RF 模型,具有预测性能稳定、预测精度高等优势。

参考文献:

- [1] Xu G, He J S, Lü Z M, et al. Prediction of mechanical properties for deep drawing steel by deep learning[J]. International Journal of Minerals, Metallurgy and Materials, 2023, 30(1): 156-165.
- [2] 刘振宇, 吴思炜, 丛津功, 等. 热轧钢材组织演变和力学性能预测——半经验模型、机器学习到工业大模型[J]. 轧钢, 2024, 41(5): 24-37, 50.
- [3] 蔡骛, 陈满泰, 左文康, 等. 高强度钢材低温力学性能试验研究与预测模型[J]. 上海交通大学学报, 2024, 58(11): 1707-1715.
- [4] 何晓霞, 魏苙越, 田彤, 等. 基于 LASSO 复合分位数回归的钢材力学性能预测[J]. 数学的实践与认识, 2023, 53(1): 122-129.
- [5] Zhao Y H, Weng Y, Peng N Q, et al. Prediction of mechanical properties of hot rolled strip by using semi-parametric single-index model[J]. Journal of Iron and Steel Research International, 2013, 20(7): 9-15.
- [6] Wu S W, Zhou X G, Cao G M, et al. High-dimensional data-driven optimal design for hot strip rolling of microalloyed steel[J]. Steel Research International, 2018, 89(7): 1800015.
- [7] Wu S W, Ren J K, Zhou X G, et al. Comparisons of different data-driven modeling techniques for predicting tensile strength of X70 pipeline steels[J]. Transactions of the Indian Institute of Metals, 2019, 72(5): 1277-1288.
- [8] Wu S W, Yang J, Cao G M, et al. Elevating prediction performance for mechanical properties of hot-rolled strips by using semi-supervised regression and deep learning[J]. IEEE Access, 2020, 8: 134124-134136.
- [9] Li W G, Yang W, Zhao Y T, et al. Modeling hot strip rolling process under framework of generalized additive model[J]. Journal of Central South University, 2019, 26(9): 2379-2392.
- [10] 夏志辉, 李维刚, 毛云飞. 基于 PCA-DWT-SVR 热连轧带钢板凸度预测[J]. 中国冶金, 2025, 35(3): 103-112.
- [11] 李维刚, 石林, 刘玮汲. 热连轧带钢厚度缺陷溯源研究及应用[J]. 中国冶金, 2024, 34(1): 99-108.
- [12] 马威, 李维刚, 赵云涛, 等. 基于深度学习的热连轧轧制力预测[J]. 钢铁研究学报, 2019, 31(9): 805-815.
- [13] 魏宝民, 王孝建, 白振华. 热镀锌机组产品力学性能预报技术[J]. 中国冶金, 2024, 34(1): 116-123.
- [14] 蔡超, 胡成翔. 基于 Huber 损失函数的稳健随机森林模型及应用[J]. 统计与决策, 2026, 42(5): 47-53.

(下转第 38 页)

的过程^[13],随着计算机和控制器的能力越来越强大,过程自动化级的很多功能“下沉”至基础自动化级来执行,融合形成“过程控制级”^[14];而管理的扁平化需求,使得制造执行系统和企业资源计划的分工界限变得不那么明显,融合形成“生产管理级”^[15]。所有的系统架构,总体上都是围绕企业的核心发展目标,实现各系统的功能定位和分工合作^[16]。需要依据生产需求和相关技术在不同发展阶段的特点,合理确定相关技术所能发挥的作用^[17],以及其在整个生产体系的地位,最终实现各项技术与现有的工业自动化系统相结合,形成有机的整体,有效地实现企业的发展目标。

参考文献:

- [1] 杜猛俊. 智能配电自动化系统通信网络架构设计与实现[D]. 北京:华北电力大学,2013.
- [2] 岳鹏. 智能用电信息系统通信机制及技术架构[J]. 电子制作,2018,26(增刊2):48-49,43.
- [3] 南国顺. 信息中心网络的服务提供机制和优化研究[D]. 北京:北京邮电大学,2017.
- [4] 吴洪涛,蒋天宇,常天佐,等. 中国机器人产业现状与创新发展趋势综述[J]. 机械制造与自动化,2025,54(1):1-6.
- [5] 史慧杰,王立平,王冬. 基于微服务的制造执行系统架构研究[J]. 现代制造工程,2024(1):45-50.
- [6] 吴启标,王亚刚,杨卫国. 基于改进S~3控制结构的批次过程控制系统架构研究[J]. 控制工程,2025,32(1):158-165.
- [7] 秦子晗,李伟东,冯美华,等. 冲击地压智能化防治系统架构及关键技术[J]. 煤炭科学技术,2022,50(8):1-7.
- [8] 李清,瞿盟津,于超. 面向新型工业化的国产工业软件架构[J]. 新型工业化,2023,13(9):31-38.
- [9] 谢安俊. 软件产业全球化与新型工业化[J]. 未来与发展,2004,28(4):28-31.
- [10] 齐永国. 煤矿机电自动化集控的发展与应用[J]. 企业技术开发,2013,32(6):73,75.
- [11] 杨克璧,汪建安,任恒,等. 汽车自动化人机柔性台架开发及应用流程方法研究[J]. 汽车工艺师,2025(7):6-11.
- [12] 赵艳莉,赵倩,李志强. 基于PLC的山核桃破壳自动化生产线控制系统研究[J]. 食品与机械,2023,39(1):111-115,194.
- [13] 田乐意. 煤矿用钻杆自动化生产线方案设计[J]. 煤矿机械,2024,45(6):120-122.
- [14] 任惠霞. 人工智能技术在电池自动化生产线中的应用分析[J]. 储能科学与技术,2024,13(11):4056-4058.
- [15] 柴钰庚,董兵斌,张兴,等. 面向未来的冶金机械自动化生产线设计与发展趋势[J]. 机械研究与应用,2025,38(1):160-162,167.
- [16] 周吉林. A公司汽车冲压自动化生产线项目技术风险控制研究[D]. 济南:山东大学,2024.
- [17] 周磊,张礼华,陈景铭,等. 基于PLC的可降解塑料片材成型自动化生产线设计[J]. 机床与液压,2024,52(10):116-122.

收稿日期:2024-04-24

(上接第23页)

- [15] 王小妹,王惠英,吉顺平. 基于随机森林算法的图书借阅量预测研究[J]. 信息记录材料,2026,27(5):181-183.
- [16] 何春宁,乔宁,陈海东,等. 基于深度随机森林算法的电力负荷分时段预测[J]. 信息技术,2026(2):78-83.
- [17] 李晓平,刘科材,庞婷婷,等. 基于最大互信息系数结合弹性网络特征提取的原酒质量评价方法[J]. 食品与发酵工业,2025,51(16):172-181,I0010,I0011.
- [18] 陈信强,樊家畅,吴华锋,等. 融合可逆实例归一化及Informer编解码的船速预测[J]. 上海海事大学学报,2025,46(4):1-9.
- [19] 艾博,孙伟,宗俊吉,等. 基于拉依达准则改进的测量重复性评定方法[J]. 电子工业专用设备,2024,53(6):43-45,57.
- [20] 邓乙丁,李艳玲,徐颖,等. 大坝安全监测数据异常识别模型簇研究[J]. 人民长江,2024,55(4):230-238.

收稿日期:2024-03-20