

DOI:10.3969/j.issn.1003-5060.2023.11.001

基于随机森林的离合器制造过程质量控制方法研究

沈维蕾, 张鑫洋, 杨雪春

(合肥工业大学 机械工程学院, 安徽 合肥 230009)

摘要:传统控制图作为统计过程控制(statistical process control, SPC)的核心工具,对影响制造过程的系统性因素识别效率较低。文章针对传统控制图对系统性因素识别不充分的问题,提出基于随机森林的控制图模式识别模型,利用网格搜索法进行参数优化,建立基于随机森林算法流程和控制图模式识别模型以识别影响过程失控的系统性因素;以汽车离合器为例,将基于随机森林的模式识别算法应用到离合器制造过程中,并与支持向量机(support vector machine, SVM)、逻辑回归(logistic regression, LR)等机器学习算法相比较,结果验证了随机森林模型对控制图模式识别的可行性和有效性。

关键词:统计过程控制(SPC);随机森林;离合器;模式识别;机器学习

中图分类号:TH164 **文献标志码:**A **文章编号:**1003-5060(2023)11-1441-07

Research on quality control method of clutch manufacturing process based on random forest

SHEN Weilei, ZHANG Xinyang, YANG Xuechun

(School of Mechanical Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: As the core tool of statistical process control (SPC), the traditional control chart has limitations in terms of the recognition of systematic factors that affect the manufacturing process. Aiming to solve the problem of insufficient systematic factors recognition of the traditional control chart, this paper proposes a control chart pattern recognition model based on random forest. The grid search method is used to optimize the parameters, and a model based on random forest algorithm flow and control chart pattern recognition is established to identify the systematic factors that affect the process out of control. Taking the automobile clutch as an example, the pattern recognition algorithm based on random forest is applied to the clutch manufacturing process, and then compared with the support vector machine (SVM), logistic regression (LR) and other machine learning methods to verify the feasibility and effectiveness of the random forest model for the control chart pattern recognition.

Key words: statistical process control (SPC); random forest; clutch; pattern recognition; machine learning

汽车在日常生活中不可或缺,近年来,人们对汽车的需求持续增长,汽车产量呈现飞速增长的趋势^[1]。离合器由于使用的特殊性,其市场容量比汽车整体的市场容量大^[2],提高离合器的质量对提高汽车整体质量、降低经济损失有重要意义^[3-4]。目前针对改进汽车传动系统质量的研究,大多从离合器设计方面进行改进,对生产制造过

程的研究并不多。因此,本文通过收集生产数据,并对制造过程进行统计过程分析,以研究如何提高制造过程质量的稳定性。

文献[5]针对以小样本、多变量为主要特征的油田产量预测问题,对比发现随机森林预测的结果最接近油田真实产量;文献[6]采用傅里叶变换红外光谱与随机森林结合的新方法进行试验,结

收稿日期:2022-10-12;修回日期:2022-12-15

基金项目:国家自然科学基金资助项目(51975003)

作者简介:沈维蕾(1969—),女,安徽怀远人,博士,合肥工业大学副教授,硕士生导师。

果表明该方法对低浓度泰乐菌素的预测结果比较准确;文献[7]实现并设计开发了植物叶片病害分类模型,对比发现随机森林分类器分类效果最好;文献[8]就滚刀故障诊断问题,提出基于谱图小波变换与随机森林相结合的故障诊断方法,试验结果表明该方法能达到 90% 以上的识别精度;文献[9]研究了滚动轴承工作故障信息获取困难的问题,提出一种基于随机森林的故障识别方法,建立相关故障识别模型,并与多种分类器相比较,结果表明提出的方法准确度更高;文献[10]提出将随机森林与频繁模式相结合的分类方法,该方法可通过最大化来控制每个子模型的欠拟合与过拟合问题;文献[11]就面部识别问题,提出使用聚合方法与随机森林相结合的方法,实际应用证明该方法比以前的方法更优越;文献[12]针对多类疾病的分类问题,提出结合属性评价者方法和实例过滤方法改进的随机森林算法,实践结果表明,针对多类疾病数据集,改进后的算法性能更佳;文献[13]提出基于随机森林的糖尿病预测模型,结果表明该模型的预测结果可信度较高;文献[14]建立基于随机森林的机器学习算法模型,从而预测

药物依从性,对过去 8 年的数据进行回顾性分析,结果表明该方法可应用于多因素的预测;文献[15]提出基于随机森林的心脏病预测方法,通过与多目标粒子群算法相结合生成多样化和精确的决策树,结果表明该方法可以提高预测精度。

综合发现,随机森林在多个领域应用良好,在分类上识别结果准确度高。相较于其他机器学习方法,随机森林具有运算量小、实现简单的优点。因此,本文将研究随机森林在控制图模式识别方面的应用。

1 控制图模式识别

休哈特控制图是统计过程控制 (statistical process control, SPC) 中最重要的工具之一,可以实现对生产过程的监视与控制,用于反映指定过程是否按计划运行,体现可能存在的一些不规则进程变化的原因。控制图中的正常模式表示进程处于正常状态,没有任何故障或变化。除正常模式外,还有周期模式、阶跃上升模式、阶跃下降模式、逐渐增加模式、逐渐减少模式等异常模式,如图 1 所示。

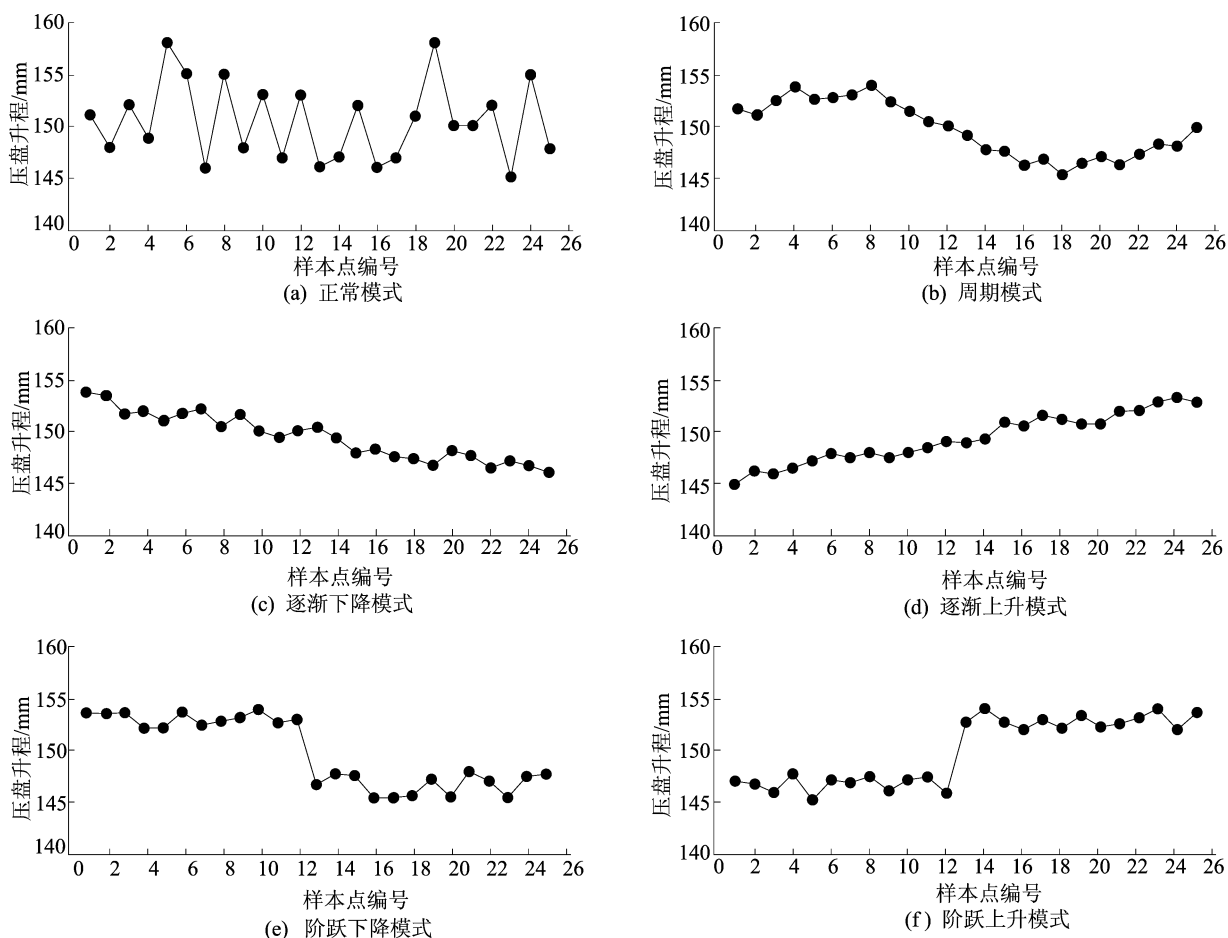


图 1 控制图模式分类

1) 正常模式下控制图的产品质量观测值符合正态分布规律,在控制限内随机分布,上下不断波动。

2) 周期模式的图像中产品质量观测值围绕中心线上下呈周期性变化趋势。

3) 逐渐下降模式和逐渐上升模式表现为质量观测值呈现下降或者上升趋势,这种趋势是持续发生的。

4) 阶跃下降模式的质量观测值会在某一观测点突然整体向下偏移,且偏移过后质量观测值不再上升;阶跃上升模式则是在某一观测点质量观测值整体突然向上偏移,且偏移过后不再回落。阶跃模式最显著的特点是断崖式偏移。

准确地识别控制图模式(control chart patterns, CCP)可以监控生产过程是否正常,防止不合格品的产生。

2 随机森林算法

在机器学习方法中,随机森林是一种比较经典的算法。文献[16]提出了随机森林方法,该方法通过随机的方式建立一个由决策树组成的森林,每一棵树之间相互独立。决策树通过拆分预测变量并递归划分数据集来描述因变量与 1 个或多个自变量之间的关系^[17-18],其实质是对决策树算法的一种改进,将多个决策树合并在一起,每棵树的建立依赖于一个独立抽取的样本集。

因此,随机森林是一个由单株分类器 $\{h(\mathbf{X}, \theta_k); k=1, 2, \dots\}$ 组成的分类器,其中 θ_k 是独立同分布的随机变量。在输入 \mathbf{X} 后,每棵决策树只投 1 票给它认为最合适的分类标签,最后选择投票多的那个分类标签作为 \mathbf{X} 的分类。

引入随机变量 θ_k 是为了控制每棵树的生长,通常针对于第 k 棵决策树引进随机变量 θ_k ,它与前面的 $k-1$ 个随机变量是独立同分布的,利用训练集和 θ_k 生成第 k 棵树,也就等价于生成一个分类器 $h(\mathbf{X}, \theta_k)$,其中 \mathbf{X} 为一个输入向量。

对于给定的一系列分类器 $h(\mathbf{X}, \theta_1), h(\mathbf{X}, \theta_2), h(\mathbf{X}, \theta_3), \dots, h(\mathbf{X}, \theta_k)$,随机地选择一些样本,设其中的 \mathbf{X} 为样本向量, \mathbf{Y} 为正确分类的分类标签向量。定义边际函数为:

$$m_g(\mathbf{X}, \mathbf{Y}) = a_{vk}(I(h(\mathbf{X}, \theta_k) = y)) - \max_{j \neq y} a_{vk}(I(h(\mathbf{x}) = j)) \quad (1)$$

其中: $I(\cdot)$ 表示示性函数; $a_v(\cdot)$ 表示取平均值。边际函数表示在正确分类 \mathbf{Y} 之下 \mathbf{X} 的得票数目超过其他分类的最大得票数目,该值越大表明分类

的置信度越高。

泛化误差公式为:

$$P_E = P_{\mathbf{X}, \mathbf{Y}}(m_g(\mathbf{X}, \mathbf{Y}) < 0) \quad (2)$$

其中, \mathbf{X}, \mathbf{Y} 表示概率的定义空间。

根据大数定律中的辛钦定理,当决策树数目增加时,对于所有的序列 θ_k 和 P_E 都会收敛到下式,对应于大数定律里的频率收敛于概率,即

$$P_{\mathbf{X}, \mathbf{Y}}\{P_\theta[h(\mathbf{X}, \theta) = y] = y - \max_{j \neq y} P_\theta[h(\mathbf{x}, \theta) = j] < 0\} \quad (3)$$

因此随机森林不会随着决策树的增加而过度拟合,并且有一个有限的泛化误差值。随机森林算法在小样本、多特征的数据集上表现良好,在学习过程中具有很好的抗噪声能力,分类识别结果准确度高。

3 基于随机森林的识别模型

3.1 随机森林模型的搭建

随机森林外部应用 Bagging 框架,基模型为决策树。Bagging 框架中的子训练集是通过训练集自助采样得到的,具有相互独立性,测试集得到的预测结果是对基模型的结果综合。Bagging 框架下的综合预测一般采用投票法,随机森林框架如图 2 所示。

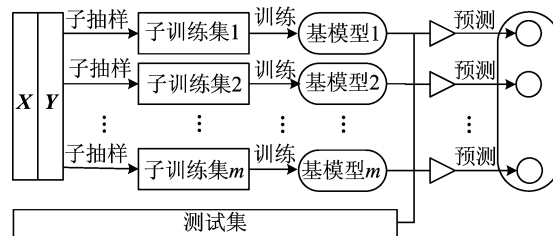


图 2 随机森林框架

2 种采样方法保证了算法的随机性,因此随机森林的基模型不需要进行减枝,可以做到尽可能地生长也不过拟合。故随机森林在各类数据集上表现良好,可以最大可能地保留原始数据特征,不需要进行数据特征选择,这也是它的优势。

随机森林中最重要的是树的生成,每棵树的生成遵从一定的规则。

1) 假设模型总训练集的大小为 N ,则对每棵决策树,都是通过自助采样法从训练集中抽取一定数量的样本作为训练样本,总共抽取 N 个样本作为该树的训练集,重复 k 次,对应生成 k 组训练样本集。

2) 每个特征的样本维度为 M ,指定一个常数

$m \in M$, 从 M 个特征中随机选取 m 个特征。

3) 利用 m 个特征让每棵树生长到最大程度, 不需要剪枝, 尽可能保留数据特征。

最终得到应用随机森林模型进行分类的算法流程, 如图 3 所示。

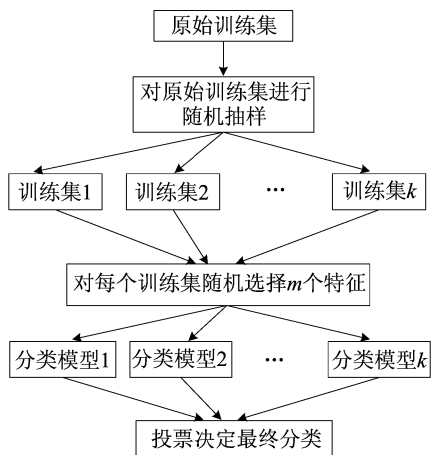


图 3 随机森林分类算法流程

3.2 基于随机森林的识别模型

基于随机森林的控制图模式识别模型如图 4 所示。该识别模型主要工作步骤如下:

1) 确定生产工艺中的关键工序, 并获取关键参数相关数据; 运用 SPC 有目的地收集生产数据; 对生产数据进行处理, 并根据生产数据特征选择合适的控制图。

2) 当生产异常模式样本数据不足时, 可以根据生产数据关键参数特征进行仿真实验, 从而获取充足的生产数据; 从仿真数据中抽取适量的生产数据样本, 对随机森林模型进行训练, 抽取 k 个样本, 同时生成 k 个数据模型, 完成模型训练。

3) 使用该模型对预测集数据样本进行预测, 通过预测结果进行分析, 证实其可行性和有效性, 并分析其识别精度。

4) 通过模式输出结果确定生产质量控制模式, 判断生产状况是否正常, 并进行质量诊断分析。

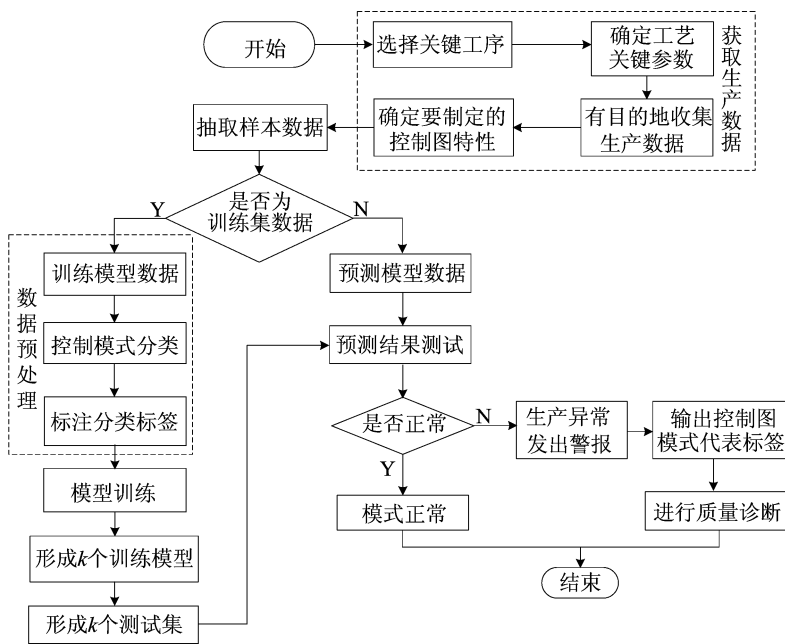


图 4 基于随机森林的控制图模式识别模型

3.3 模型评价指标

模型的评估指标有准确率、精确度、召回率和 F_1 值。其中准确率也是最简单的指标, 但存在明显的缺陷, 且存在错误率。测试集中被判定为正样本的数量记为 T_P , 判定为负样本数量记为 F_N ; 负样本被判定为负样本的数量记为 T_N , 被判定为正样本的数量记为 F_P 。本文采用精确度、召回率以及精确度与召回率的加权调和平均进行模型评价^[19]。

1) 精确度 (Precision) P_1 也叫查准率, 定义为:

$$P_1 = \frac{T_P}{T_P + F_P} \quad (4)$$

2) 召回率 (Recall) R 也叫查全率, 定义为:

$$R = \frac{T_P}{T_P + F_N} \quad (5)$$

3) F_1 值来自精确度与召回率的加权调和平均, 定义为:

$$F_1 = \frac{2P_1R}{P_1+R} \tag{6}$$

4 应用实例

SPC 通过控制图理论对关键工序进行过程诊断与控制,可以避免大量不合格品的出现。离合器是汽车传动系统中的重要元件,汽车离合器有摩擦式离合器、液力变矩器(液力耦合器)、电磁离合器等几种。从装配过程来看,离合器的装配分为两大部分,即盖及压盘部分和从动盘部分。通过现场调查发现,压盘升程既是盖及压盘总成特殊特性,也是工序零件特殊特性。因此本文基于 SPC 对离合器装配过程的关键质量特性——压盘升程进行控制图模式识别,找出引起质量问题的系统性因素,并对出现的质量问题及时进行识别和纠正。

压盘升程是指压盘离开摩擦片后的轴向后移距离,是离合器能否正常分离的重要参数。若压盘升程偏小,将发生离合器半分离半接合的现象,即分离不彻底;不彻底将使离合器磨损过快,换挡产生冲击,甚至可能导致离合器损坏,在行驶过程中无法换挡,行驶速度不受控。压盘升程的质量合格与否直接影响离合器的功能。

根据特殊特性选择压盘升程,从压盘升程的计算原理、数据特点上按照控制图选取流程,最终选择均值极差控制图。压盘升程的技术标准为 1.30~1.65 mm。因为生产实际过程中异常模式生产样本数据不足,获取足够的异常模式样本所耗费的成本过大,所以通过采用蒙特卡洛方法生成样本数据,确定了 6 种基本控制图模式的参数,利用 MATLAB 仿真获取。

为满足模型训练样本的需求,模型应用的控制图采用均值控制图。本文将所得数据处理为均值控制图所需的样本数据,最终得到每种均值控制图模式的样本量为 50 组,其中每一组包含 25 个数据均值样本点,共 300 组生产数据。从样本数据结果计算得出,压盘升程的均值控制上限为 1.540 mm,控制中心限为 1.495 mm,控制下限为 1.450 mm。

对所获得的样本数据进行预处理,并将 6 种不同模式分别标注,标签为 1,2,⋯,6;每种模式内的特征值有 25 个,标注为 d_1, d_2, \dots, d_{25} 。

本文总共利用 300 组数据,为避免数据分布不均等造成模型结果的误差,采用五折交叉验证法对模型进行评估。五折交叉验证法即将样本数

据随机、均匀地分为 5 份,轮流用其中 4 份训练模型,余下的 1 份作为测试集,以此来测试模式的准确率,最后取 5 次模型的准确率均值作为模型的准确率。

4.1 基于随机森林的离合器质量控制

基于随机森林的离合器控制图模式识别模型结果评价采用五折交叉验证法,其中一组测试结果的混淆矩阵如图 5 所示。

10	0	0	0	0	0
0	10	0	0	0	0
0	0	9	0	0	1
0	0	0	10	0	0
0	0	0	0	10	0
0	0	0	0	0	10

图 5 混淆矩阵结果

从图 5 的混淆矩阵可以直观地看出,实际结果与预测结果基本吻合,预测结果中仅有 1 例实际类别为类别 3 的模式判断为类别 6,导致输出结果错误。

模型预测结果数据如下所示:

1) 实际结果为[1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6];

2) 预测结果为[1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 6 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6];

3) 精准度为 0.983 333 333 333 333 3;

4) 召回率为 0.983 333 333 333 333 3;

5) F_1 值为 0.983 333 333 333 333 3。

五折交叉验证所得结果见表 1 所列。

表 1 随机森林方法下五折交叉验证结果

组别	准确率	精准度	召回率	F_1 值
第 1 组	0.983	0.983	0.983	0.983
第 2 组	1.000	1.000	1.000	1.000
第 3 组	0.983	0.983	0.983	0.983
第 4 组	0.983	0.983	0.983	0.983
第 5 组	1.000	1.000	1.000	1.000
平均值	0.989	0.989	0.989	0.989

从表 1 可以看出,接受者操作特性曲线(receiver operating characteristic curve, ROC)平均值为 0.989。平均值为模型的最终评估数据,代表模型的准确度。

ROC 曲线如图 6 所示。

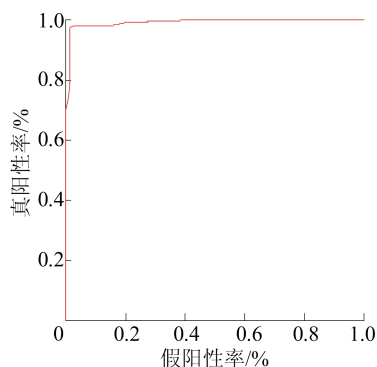


图 6 ROC 曲线

4.2 与其他机器学习算法的对比

随着训练数据的增加,模型的预测精度必然有所提升。为验证随机森林的模型性能,本文采用同样的数据样本,将随机森林算法与不同机器学习算法所建模型的结果进行对比分析。分别构建基于支持向量机(support vector machine, SVM)的控制图模式识别模型和基于逻辑回归(logistic regression, LR)算法的控制图模式识别模型,并采用离合器的相关数据进行模型训练与测试。

SVM 五折交叉验证结果见表 2 所列。由表 2 可知,该模型准确率为 0.953。

表 2 SVM 算法下五折交叉验证结果

组别	准确率	精准度	召回率	F_1 值
第 1 组	0.917	0.917	0.917	0.917
第 2 组	0.950	0.950	0.950	0.950
第 3 组	0.983	0.983	0.983	0.983
第 4 组	0.933	0.933	0.933	0.933
第 5 组	0.983	0.983	0.983	0.983
平均值	0.953	0.953	0.953	0.953

LR 五折交叉验证结果见表 3 所列。由表 3 可知,该模型准确率为 0.654。

表 3 LR 算法下五折交叉验证结果

组别	准确率	精准度	召回率	F_1 值
第 1 组	0.667	0.667	0.667	0.667
第 2 组	0.617	0.617	0.617	0.617
第 3 组	0.717	0.717	0.717	0.717
第 4 组	0.650	0.650	0.650	0.650
第 5 组	0.617	0.617	0.617	0.617
平均值	0.654	0.654	0.654	0.654

最终获取 3 种不同机器学习算法的模型预测结果指数,见表 4 所列。

表 4 不同机器学习算法模型预测结果指数对比

算法	准确率	精准度	召回率	F_1 值
SVM	0.953	0.953	0.953	0.953
LR	0.654	0.654	0.654	0.654
随机森林	0.989	0.989	0.989	0.989

从表 4 可以看出,基于机器学习算法的控制图模式识别模型中,同样的数据样本条件下,随机森林要优于其他算法。可知,随机森林具有非常明显的优势,它不需要过多的模型训练就能达到一个极高的精确度,从效率上要高于其他算法,适用于现有的小批量多品种生产模式,能够有效地提高生产质量控制。

5 结 论

本文针对控制图模式识别问题搭建了基于随机森林的控制图模式识别模型,通过离合器的实例应用验证其可行性与实用性,提高了生产人员在生产过程中对产品质量的控制,从而为减少产品质量问题,提前预知异常的发生提供技术支持。本文的主要工作如下:

1) 针对控制图控制线内系统性因素识别能力不足的问题,提出了基于随机森林的控制图模式识别分类模型,搭建了随机森林算法的学习框架,并采用网格搜索对随机森林模型进行参数优化,建立了基于随机森林的控制图模式识别模型。

2) 以离合器制造过程为应用实例,将基于随机森林的控制图模式识别模型应用于离合器装配关键质量特性——压盘升程中,并与 SVM 和逻辑回归的控制图模式识别模型进行比较,通过五折交叉验证进行模型评估;结果表明,在控制图分类识别方面随机森林算法比其他机器学习算法准确率最高、性能更优越。

[参 考 文 献]

- [1] 吴佳星. 双离合器的研究分析[J]. 信息记录材料, 2018, 19(10):225-226.
- [2] 李红云. 汽车零部件再制造技术性评价模型的建立及应用[J]. 现代工业经济和信息化, 2020, 10(11):84-85.
- [3] 李一斌. 长春市汽车零部件产业发展政策研究[D]. 长春: 吉林大学, 2020.
- [4] 严正峰. 汽车离合器行业发展研究[J]. 汽车与配件, 2011(42): 24-26.
- [5] 曹冲, 程林松, 张向阳, 等. 基于多变量小样本的渗流代理模型及产量预测方法[J]. 力学学报, 2021, 53(8):2345-2354.

(下转第 1500 页)

- 2018,56(2):268-281.
- [10] ZHANG C H,JI E Y,WANG B T, et al. Research on a critical hydraulic gradient of piping in noncohesive soils [J]. *Soils Advances in Civil Engineering*, 2021, 2021:6217101.
- [11] ROHANINEJAD M,ZARGHAMI M. Combining Monte Carlo and finite difference methods for effective simulation of dam behavior[J]. *Advances in Engineering Software*, 2011,45(1):197-202.
- [12] BARBETTA S,CAMICI S,BERTUCCIOLI P, et al. Levee body seepage: a refinement of an expeditious procedure for fragility curves and vulnerability diagrams assessment[J]. *Hydrology Research*,2017,48(3):763-775.
- [13] WANG F W,OKEKEUSTIN A C U,KOGURE T, et al. Assessing the internal structure of landslide dams subject to possible piping erosion by means of microtremor chain array and self-potential surveys[J]. *Engineering Geology*, 2018,234:11-26.
- [14] MING P,LU J,CAI X, et al. Multi-particle model of the critical hydraulic gradient for dike piping[J]. *Soil Mechanics and Foundation Engineering*,2020,57(3):200-210.
- [15] 梁越,陈建生,陈亮,等. 双层堤基管涌发生发展的试验模拟与分析[J]. *岩土工程学报*,2011,33(4):624-629.
- [16] 黄翔. 三峡库区马家沟滑坡抗滑桩变形机制研究[D]. 合肥:合肥工业大学,2018.
- [17] 杨启贵,林学锋,郭志华. 极端冰雪灾害对边坡工程稳定性影响分析研究[J]. *人民长江*,2010,41(24):67-71.
- [18] 胡建林,孙利成,崔宏环,等. 基于修正摩尔库伦模型的深基坑变形数值分析[J]. *科学技术与工程*,2021,21(18):7717-7723.
- [19] 胡建林,孙利成,崔宏环,等. 修正摩尔库伦模型下的深基坑变形数值分析[J]. *辽宁工程技术大学学报(自然科学版)*,2021,40(2):134-140.
- [20] 刘鹏,李祥祥,李剑锋,等. 临地铁结构物桩基施工技术研究[J]. *公路*,2019,64(12):89-96.
- [21] FU Z Q,SU H Z,WEN Z P. Multi-scale numerical analysis for linear elastic behavior of clay concrete[J]. *International Journal of Solids and Structures*,2020,203:23-45.
- [22] 邹育麟,何川,胡雄玉,等. 富水区隧道合理防排水型式及注浆加固参数研究[J]. *防灾减灾工程学报*,2014,34(4):485-491.
- [23] 王宇,谷艳昌,王士军,等. 基于雷诺数的砂砾石管涌过程判别[J]. *水利水运工程学报*,2022(6):113-120.
- [24] 陶高梁,李进,崔惜琳. 不同颗粒级配的砂土渗流破坏特性[J]. *土木工程与管理学报*,2019,36(2):90-97.
- [25] 魏勇,赵安文,许开立. 尾矿库坝基管涌破坏试验[J]. *金属矿山*,2015(9):157-160.

(责任编辑 胡亚敏)

(上接第 1446 页)

- [6] GOMES M D F A,ALMIR C M L,ELIZABETH A D M B, et al. Infrared spectroscopy combined with random forest to determine tylosin residues in powdered milk [J]. *Food Chemistry*,2021,365:130477-130477.
- [7] SINGH A K,SREENIVASU S,MAHALAXMI U S B K, et al. Hybrid feature-based disease detection in plant leaf using convolutional neural network, bayesian optimized SVM, and random forest classifier [J]. *Journal of Food Quality*,2022,2022:1-16.
- [8] DONG X,LI G L,JIA Y C, et al. Multiscale feature extraction from the perspective of graph for hob fault diagnosis using spectral graph wavelet transform combined with improved random forest[J]. *Measurement*,2021,176:109178.
- [9] 王兰兰,朱捷,周正平,等. 基于随机森林的滚动轴承故障辨识方法研究[J]. *机电工程*,2021,38(12):1599-1604.
- [10] HATWELL J,GABER M,AZAD R, et al. CHIRPS: Explaining random forest classification [J]. *Artificial Intelligence Review*,2020,53(8):5747-5788.
- [11] AISHWARYA K,KUMAR B S,RAJU S V, et al. Facial recognition using aggregation and random forest classification method [J]. *Journal of Physics: Conference Series*, 2019,1362(1):12078.
- [12] ARCHANA C,SAVITA K,RAJ K, et al. An improved random forest classifier for multi-class classification[J]. *Information Processing in Agriculture*, 2016, 3 (4): 215-222.
- [13] 杨美洁,唐建军. 基于随机森林算法的糖尿病预测研究[J]. *医学信息学杂志*,2019,40(9):47-49.
- [14] KAMAL S,URATA J,CAVASSINI M, et al. Random forest machine learning algorithm predicts virologic outcomes among HIV infected adults in Lausanne, switzerland using electronically monitored combined antiretroviral treatment adherence[J]. *Aids Care*,2021,33(4):530-536.
- [15] ASADI S,ROSHAN S,KATTAN M W, et al. Random forest swarm optimization-based for heart diseases diagnosis [J]. *Journal of Biomedical Informatics*,2021,115:103690.
- [16] LEO B. Random forests [J]. *Machine Learning*, 2001, 45(1):5-32.
- [17] THOMAS M H,JULIAN F P K,FRED A H, et al. End-to-end learning of decision trees and forests[J]. *International Journal of Computer Vision*,2019,128:997-1011.
- [18] SEYED A N,KOUROSH A,ALIREZA D, et al. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping [J]. *Water Resources Management*, 2017,31(9):2761-2775.
- [19] 赵澄,陈君新. 基于隐马尔可夫模型的反射型 XSS 检测技术[J]. *浙江工业大学学报*,2019,47(4):442-447.

(责任编辑 胡亚敏)