

基于随机森林的钛加工表面质量评定研究

李晟,周超

(福州大学 机械工程及自动化学院,福建 福州 350116)

摘 要:针对不同工艺加工的零件表面质量未有统一评价标准且三维参数众多未有统一规范的问题,提出了一种基于随机森林算法筛选用以评定表面质量的三维参数方法。采用三维轮廓仪,分别测量采用超声滚压、车削、磨削、抛光、铣削5种工艺加工的钛材表面形貌,基于随机森林算法,通过学习选用的高度类及间距类共9个三维参数作为研究特征,以达到准确分类表面的目的,并筛选出用于评定钛材加工表面质量的参数集。研究表明:基于随机森林算法,算术平均高度、均方根高度及偏斜度在分类过程中有最高的重要性,且在三者组成的三维空间中表面区分效果良好,作为参数集能够合理、可靠地评定钛材加工表面质量。

关键词:表面形貌;随机森林;特征重要性;参数评定

中图分类号:TH161⁺.1; TP301.6 **文献标识码:**A **文章编号:**1671-5276(2020)06-0036-03

Study of Surface Quality Evaluation of Titanium Processing Based on Random Forest

LI Sheng, ZHOU Chao

(School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350116, China)

Abstract: Aiming at the problem that there is no uniform evaluation standard and specification for surface quality of mechanical parts machined by different processes and 3D parameters, a method based on random forest algorithm used to screen 3D parameters for surface quality evaluation is proposed in this paper. The surface topography of titanium material processed by ultrasonic rolling, turning, grinding, polishing and milling is measured by 3D surface profilometer. Through learning, nine 3D parameters for height class and distance class are selected as the research features, which are used to classify different machined surfaces accurately and screen out 3D parameter sets for evaluating the surface quality of the titanium material. The result shows that based on the random forest algorithm, the arithmetic average height, root mean square height and skew are of the highest importance in the classification process, and the effect for the surface discrimination is good in the 3D space composed of them, which can be used to reasonably and reliably evaluate the processing surface quality.

Keywords: surface topography; random forest; feature importance; parameter evaluation

0 引言

钛及其合金具有耐腐蚀、导热率低及生物相容性好等特点,在航天、航空、医疗等领域广泛应用,是具有特殊战略意义的高性能金属材料^[1]。采用不同工艺加工的钛合金表面具有不同的几何形貌,其对零部件的摩擦、磨损、润滑等性能的影响越来越大,已成为决定零部件性能的重要因素之一^[2-3]。

不同工艺加工的零件表面形貌存在结构性差异,正确选用评定参数对定量表征加工表面质量从而进一步指导改进加工工艺具有重要意义。由于粗糙表面的三维本质,传统的二维轮廓评定参数不能全面表征表面质量,越来越多的研究工作者选用三维参数评价表面质量^[4]。ISO 25178-2: 2012给出了5类共40多个三维参数,必须针对特定应用从中选取合适的参数集,方能准确、全面地评定表面质量。例如,吴乙万等^[5]利用三维参数分析超精密光学表面,认为 S_q 、 S_{dr} 、 S_{dq} 能很好地表征光学表面性能。杨洁等^[6]基于机器视觉对不同工艺处理的金属样块

进行研究,认为 S_q 、 S_{sk} 、 S_{ds} 可作为三维表面粗糙度评价的可靠参数;赵登超等^[7]选取三维参数中的高度类、空间类和混合类中的12个参数进行聚类分析,发现 S_a 、 S_q 对于识别不同抛光工艺的紫铜表面的贡献度最大。

本文采用超声滚压、车削、磨削、抛光、铣削5种不同工艺加工钛样本表面,使用三维轮廓仪测量表面几何形貌;基于机器学习中的随机森林算法,以表面的三维参数 S_a 、 S_q 、 S_{sk} 、 S_{ku} 、 S_p 、 S_v 、 S_z 、 S_{al} 、 S_{tr} 作为研究特征,进行表面的自动分类、特征筛选(上述表征参数的意义及算法,参见文献[5]),发现 S_a (算术平均高度)、 S_q (均方根高度)和 S_{sk} (偏斜度)组成的参数集具有表征表面质量的最高重要性,并可精准区分不同加工表面。

1 随机森林算法

随机森林(random forest, RF)算法由LEO Breiman^[8]提出,将CART决策树与Bagging算法相结合,形成一种集成学习算法(ensemble learning)。CART决策树是二元分类器^[9],

基金项目:福建省自然科学基金项目(2016J01225)

第一作者简介:李晟(1994—),男,福建龙岩人,硕士研究生,研究方向为表面形貌建模分析及超分辨。

其通过两个随机过程进行构建:1)有放回地从训练样本数据集中随机抽取固定数目的样本作为训练样本集合;2)每棵树使用的特征集是从总数据集中随机选取的少数特征。通过组合多棵决策树,最终求平均值获得预测结果。

CART 决策树的构建由非叶节点开始,选择特征的随机子集中的最佳分割点将数据二分类,且不断生长,当数据到达叶节点后不再划分。为将特征元素更合理地分布在非叶节点上,CART 决策树通常采用基尼系数使信息增益最大化:

$$I_G(D_p, s) = I_G(D_p) - \frac{N_l}{N_p} I(D_l) - \frac{N_r}{N_p} I(D_r) \quad (1)$$

式中: I_G 为基尼系数; s 为待分类特征; D_p 为父节点; N_l 为左子节点样本数量; N_p 表示父节点中样本数量; N_r 为右子节点样本数量; D_l 为左子节点; D_r 为右子节点。

对于特征 s , 在决策树生长过程中, 利用式(1)计算出节点划分前后的不纯度之差, 便得到不纯度衰减量。平均所有决策树得到的平均不纯度衰减量, 最终得到该特征的重要性^[10]。对每个特征进行此操作, 便得到所有特征的重要性, 从而进行排序、筛选。

2 实验材料及方法

2.1 钛表面加工及数据获取

本文选用直径为 20 mm 的钛棒, 分别采用车削、超声

表 1 钛棒加工表面三维参数

样本编号	$S_a/\mu\text{m}$	$S_q/\mu\text{m}$	S_{sk}	S_{ku}	S_v	S_z	S_p	$S_{al}/\mu\text{m}$	S_{tr}	类别
1	0.457	0.604	-0.019	4.211	-2.403	4.757	2.354	0.032	0.045	1
101	0.486	0.691	-1.050	8.997	4.891	7.122	2.231	0.033	0.046	2
201	0.237	0.305	0.270	3.580	-0.930	2.276	1.345	0.029	0.040	3
301	0.069	0.098	0.845	9.884	-0.365	1.122	0.757	0.008	0.058	4
401	0.407	0.513	0.099	3.111	2.014	4.157	2.142	5.553	0.550	5

在数据集中, 包含 500 个样本表面及其 9 个三维参数, 记作 500×9 维的矩阵, 即 $X \in R^{500 \times 9}$:

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_9^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_9^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ x_1^{(500)} & x_2^{(500)} & \cdots & x_9^{(500)} \end{bmatrix} \quad (2)$$

其中每一个 $x_i^{(j)}$ 代表三维参数的具体值。

类似地, 用 1 个 500 维的列向量表示 5 种加工工艺的类形标记:

$$y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(500)} \end{bmatrix} \quad (3)$$

其中每一个 $y^{(i)}$ 代表 1 个样本的类标, 取值分别为 1—超声滚压; 2—车削; 3—磨削; 4—抛光; 5—铣削。

3 实验结果及分析

3.1 预测结果分析

将数据集按照 3 : 7 的比例分为训练集和测试集, 保

滚压、铣削、磨削以及抛光 5 种工艺加工其圆端面, 每种工艺加工 2 个试样, 总计 10 个试样。车削使用外圆车刀, 转速为 300 r/min, 进给量为 80 mm/min, 背吃刀量为 0.2 mm; 球头超声滚压的加工深度为 0.2 mm, 进给量为 250 mm/min; 铣削使用球头铣刀, 转速为 10 000 r/min, 进给速度为 550 mm/min, 切削深度为 0.15 mm; 磨削使用 1500 号砂纸; 抛光使用丝绒抛光布, 粒度为 1.5 μm 的钻石膏, 抛盘转速为 1 400 r/min。采用三维轮廓仪测量样品表面中范围为 10 mm $\times$ 10 mm 的区域, 采样间隔为 4 μm , 每个表面测得 2 500 $\times$ 2 500 个高度点数据。

对测量的原始数据做预处理。根据文献[11], 经小波变换滤除测量基准面、高频噪声和异常值点; 在每个测量表面随机选取 50 个 128 $\times$ 128 个采样点组成的正方形区域, 总计 500 个表面样本。

2.2 数据处理

编写相关粗糙表面三维参数的计算程序。为适用模型训练, 将计算所得所有三维参数制成数据集。表 1 给出了表面三维参数 S_a (算术平均高度)、 S_q (均方根高度)、 S_{sk} (偏斜度)、 S_{ku} (峰态)、 S_v (最大谷深度)、 S_z (最大高度)、 S_p (最大峰高度)、 S_{al} (自相关长度) 和 S_{tr} (纹理特征比), 具体几何定义见文献[4], 并展示了数据集中部分数据, 每一行代表一个样本表面。

证两者同分布。为得出合理、可靠的结果, 需优化算法模型, 因此绘制决策树数量的验证曲线, 分析最佳决策树数量取值, 如图 1 所示。

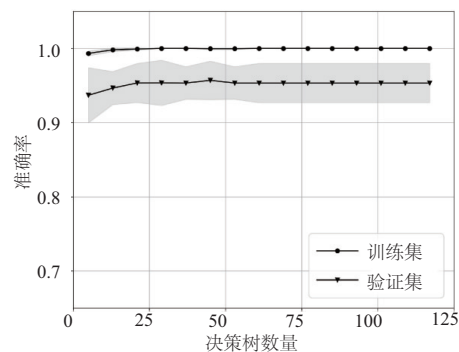


图 1 决策树数量的验证曲线

由图 1 (灰色部分为方差) 可见, 随着决策树数量上升, 模型预测准确率总体上升直至平坦。决策树数量上升, 虽然预测准确率上升, 但模型复杂度升高, 过拟合愈加严重, 同时选用不同决策树数量带来的分类以及重要度的计算成本随之增加。因此需在计算成本与降低过拟

合之间权衡。利用 K 折交叉验证法对算法进行评估,选择合理的决策树数量,量化预测准确率与泛化性能,如图 2 所示。

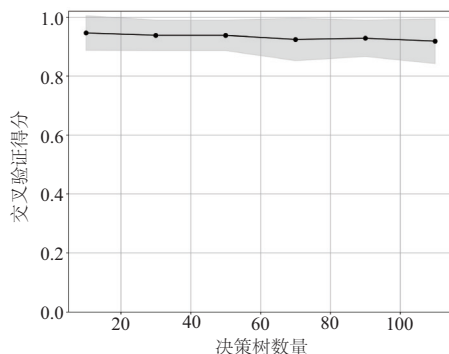


图2 决策树数量得分曲线

由图 2 可知,当决策树数量 <10 ,准确率高但方差也高,说明过拟合程度较高; >60 时,预测准确率下降而方差却升高,说明过拟合程度也较高;而在 30~40 之间时,准确率与方差的差别很小。因此考虑到计算效率与计算成本,选择 30 棵决策树较为合理。

利用优化好的随机森林模型对数据进行拟合并绘制学习曲线,分析模型性能如图 3 所示。

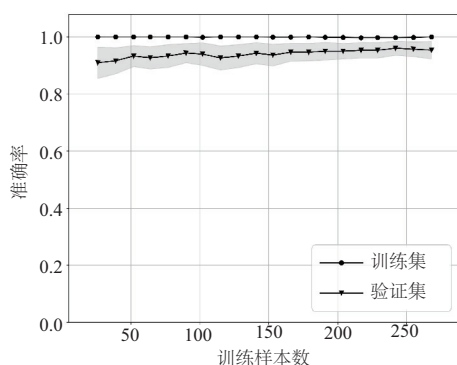


图3 学习曲线

由图 3 可见,随着训练样本数增加,模型在训练集上拟合良好,而在验证集上则是初始准确率上升较快,而后趋于平稳上升,在样本数 250 之后趋于稳定;模型的泛化性能稳步提升,最后与训练准确率只相差 2.23%,说明模型对训练数据有轻微过拟合。

3.2 三维参数筛选及分析

利用优化好的模型,计算 9 种三维参数对分类表面的重要度并排序,得到图 4 所示结果。

由图 4 可知, S_a 的重要性得分最高,为 0.23,其次是 S_q 和 S_{sk} ,分别为 0.15 和 0.14。从得分情况看,选择 S_a 、 S_q 和 S_{sk} 作为评定参数集较为合理。为验证合理性,以 S_a 、 S_q 和 S_{sk} 作为空间三坐标轴,对分类结果做可视化展示,分析合理性,详见图 5。

由图 5 可知:对于超声滚压、抛光、铣削及车削、磨削、铣削,数据在 S_a 、 S_q 和 S_{sk} 组成的三维空间中呈线性可分态,说明分类效果良好;滚压和车削呈线性不可分态,可能

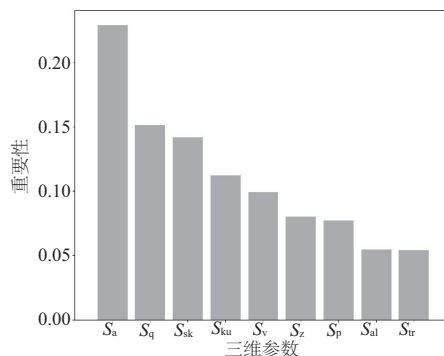


图4 特征重要性及排序

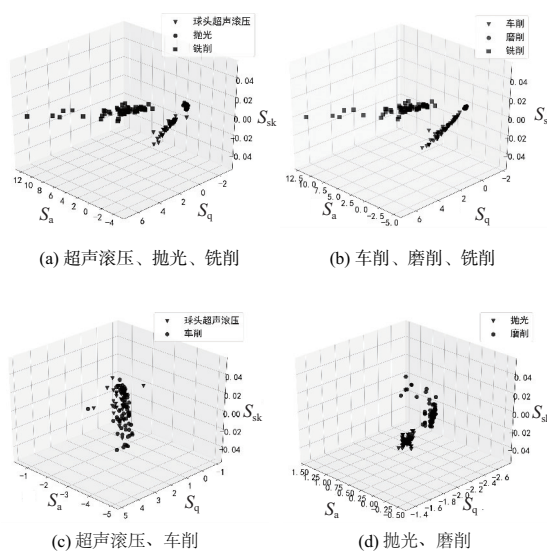


图5 三维参数分类结果

是由于滚压在车削后加工,未能完全去除残余特征;抛光和磨削之间参数差距较小,可能是因为抛光的切削深度小,两者留下的特征较为接近。从表面分类来看, S_a 、 S_q 和 S_{sk} 有良好的分类效果,作为评定参数集是合理的;从样品表面分析,由于 128×128 个点组成的正方形区域面积较小,表面纹理特征较为简单,因此 S_a 、 S_q 和 S_{sk} 可以更好地评定表面质量,同时有计算简便的优势。

4 结语

本文提出采用随机森林算法筛选三维参数用以评定钛加工表面的方法。结果表明:只需向随机森林算法提供加工试样表面的小部分区域和部分三维参数,就能自动、准确分辨不同表面;随机森林算法能够定量计算三维参数在分类表面过程中的贡献度大小,对于 5 种加工工艺而言, S_a 、 S_q 和 S_{sk} 有最大的重要性,分别达到了 0.23、0.15 和 0.14;由 S_a 、 S_q 和 S_{sk} 组成的空间中,超声滚压、抛光、铣削(车削、磨削、铣削)工艺的整体区分效果良好,说明此 3 种参数即可分类不同加工工艺,因此选用 S_a 、 S_q 和 S_{sk} 参数集定量评定钛加工表面质量是合理的,并可为提高加工质量、优化加工工艺提供定量依据。

(下转第 60 页)

根据图9和图10(a)可知,当 n 为0.8时,最大温差及温度方差达到最小,即型板表面温度均匀性最好,此时温度方差为 25.98 K^2 。而随着 n 的增大或减小,温度方差也随之增大或减小。周光炯等^[12]研究表明,温度边界层与速度边界层成正比相关,且 $\delta_t \propto x^{0.8}$ 。当 n 为0.8时,即 $s \propto x^{0.8}$ 时,型板厚度变化规律与温度边界层厚度变化规律同步,因此型板表面温度均匀性达到最佳。

3 结语

本文利用FLUENT进行热压罐固化过程中温度场模拟。研究发现型板表面低温区域出现在距迎风端115 mm处,约占面积25%。为有效提高型板表面温度均匀性,本文提出一种对型板不同区域变厚度方法,分别对型板低温区域减厚和高温区域加厚。主要包括3种厚度方法:梯形、阶梯形、连续变厚度,通过有限元仿真对改善后型板分别进行模拟型板表面温度分布均匀性,分析得到以下结果:

1) 梯形型板在突然变厚度区域存在明显的温度起伏现象,其中400 mm与600 mm尤为明显,因此可通过对距迎风端30%~33.3%处进行加厚处理。

2) 阶梯型板未出现明显温度起伏现象,当距离迎风端900 mm区域加厚时,温度均匀性达到最佳状态,温度方差降低近70.25%。

3) 连续变厚度型板采用 $s \propto x^{0.8}$ 函数关系时,型板厚度变化规律与温度边界层厚度规律一致,此时温度均匀性最佳,结果表明此时温度方差降低了84.37%。

参考文献:

- [1] 包正强. 热固性树脂基复合材料构件的固化变形仿真研究[D]. 广汉:中国民用航空飞行学院,2019.

- [2] MI Yingying, YAN Qingzhi, LI X, et al. Effects of temperature induced thermal expansion and oxidation on the Charpy Impact property of C/C composites[J]. Journal of Wuhan University of Technology-Mater, Science Citation Index Editor, 2015, 30(3): 473-477.
- [3] 赵一鸣. 复合材料构件热压罐成型模具温度场模拟与结构改进研究[D]. 南京:南京航空航天大学,2018.
- [4] 张铖. 大型复合材料结构热压罐工艺温度场权衡设计[D]. 哈尔滨:哈尔滨工业大学,2009.
- [5] 吴思保,礼嵩明,鹿海军. 芳纶纤维/碳纤维复层结构复合材料变形研究[J]. 玻璃钢/复合材料,2019(6): 90-94.
- [6] 张铖,梁宪珠,王永贵,等. 热压罐工艺环境对于先进复合材料框架式成型模具温度场的影响[J]. 材料科学与工程学报,2011(4): 547-553.
- [7] XIE G N, LIU J, ZANG W H, et al. Simulation and improvement of temperature distributions of a framed mould during the autoclave composite curing process[J]. Journal of Engineering Thermophysics, 2013, 22(1): 43-61.
- [8] 林家冠,杨睿,王廷霞,等. 大型复合材料构件热压罐成型温度分析与均匀性改善研究[J]. 玻璃钢/复合材料,2015(5): 61-65.
- [9] 傅承阳,李迎光,李楠垭,等. 飞机复合材料制件热压罐成型温度场均匀性优化方法[J]. 材料科学与工程学报,2013, 31(2): 273-276.
- [10] 张旭生,甘忠,张海燕. 热压罐时效成形模具温度场优化研究[J]. 中国制造业信息化,2011,40(19): 30-32,37.
- [11] 杨世铭,陶文铨. 传热学[M]. 4版. 北京:高等教育出版社,2006.
- [12] 周光炯,严宗毅,许世雄,等. 流体力学 下册[M]. 2版. 北京:高等教育出版社,1993.

收稿日期:2019-10-11

(上接第38页)

参考文献:

- [1] D B, C W J. Perspectives on titanium science and technology[J]. Acta Materialia, 2013, 61(3): 844-879.
- [2] 张绪寿,余来贵,陈建敏. 表面工程摩擦学研究进展[J]. 摩擦学学报,2000(2): 156-160.
- [3] LARSEN Basse J. Surface engineering and the new millennium[J]. Surface Engineering, 1998, 14(2): 81-84.
- [4] 何宝凤,魏翠娥,刘柄显,等. 三维表面粗糙度的表征和应用[J]. 光学精密工程,2018, 26(8): 1994-2011.
- [5] 吴乙万,任志英,高诚辉,等. 超精密光学镜片三维表面形貌参数评定方法研究[J]. 计量学报,2017, 38(4): 410-415.
- [6] 杨洁,李乐. 基于机器视觉的表面粗糙度测量与三维评定[J]. 光学技术,2016, 42(6): 491-495.

- [7] 赵登超,刘明,周超,等. 线性判别分析在表征抛光布抛光效果中的应用研究[J]. 机电工程,2018, 35(8): 793-798.
- [8] LEO Breiman. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [9] TIBSHIRANI R. Bias, variance, and prediction error for classification rules[M]. University of Toronto: Department of Statistics, 1996.
- [10] STROBL C, BOULESTEIX A, KNEIB T, et al. Conditional variable importance for random forests[J]. BMC Bioinformatics, 2008, 9(1): 307.
- [11] 赵仕宇,周超,詹艳然. 5052 铝合金薄板数控渐进成形制件表面质量分析[J]. 塑性工程学报,2018, 25(6): 34-40.

收稿日期:2019-07-25