

doi: 10.7690/bgzd.2023.01.011

基于机器学习的固体火箭发动机燃烧室壳体旋压数据预测

糜旗, 吕晓雷, 杨锋, 曹婷

(中国航天科技集团第八研究院上海航天动力技术研究所, 上海 201109)

摘要: 为提高固体火箭发动机燃烧室壳体旋压加工的良品率, 提出一种旋压数据预测模型。通过对旋压加工数据集进行分析处理和特征提取, 并运用机器学习建立预测模型, 五折交叉验证评估模型, 可有效地预测中部跳动和Q处跳动。预测结果表明: 该模型为后期优化旋压加工工艺, 提高燃烧室壳体良品率提供了基础。

关键词: 旋压; 固体火箭发动机; 机器学习; 模型; 跳动

中图分类号: TJ7 **文献标志码:** A

Prediction of Spinning Data of Solid Rocket Motor Combustion Chamber Shell Based on Machine Learning

Mi Qi, LYU Xiaolei, Yang Feng, Cao Ting

(Shanghai Space Propulsion Technology Research Institute, No. 8 Research Academy,
China Aerospace Science and Technology Corporation, Shanghai 201109, China)

Abstract: In order to improve the yield of solid rocket motor combustion chamber shell, a prediction model of spinning data was proposed. Through the analysis and feature extraction of the spinning processing data set, the prediction model is established by using machine learning, and the five-fold cross-validation evaluation model can effectively predict the middle runout and Q runout. The prediction results show that the model provides a basis for optimizing the spinning process and improving the yield of combustion chamber shell.

Keywords: spinning; solid rocket motor; machine learning; model; runout

0 引言

上海航天动力技术研究所(以下简称动力所)是我国航天领域集固体火箭动力系统与推进剂配方设计、研制、生产、试验为一体的专业研究所, 随着国防安全需求的日益增加, 动力所也加快了固体火箭发动机研制的速度, 不仅数量种类繁多, 而且性能指标也不尽相同。

固体火箭发动机主要分成燃烧室和喷管等零部件, 其中燃烧室最重要的组成部分即为燃烧室壳体。加工固体火箭发动机燃烧室壳体时, 国内大多选择无缝钢管为毛坯, 采用强力旋压的成形工艺加工毛坯^[1]。

强力旋压工艺系逐点变形成形工艺。旋压时旋压毛坯夹紧在芯模上, 芯模带动毛坯一起旋转^[2]。芯模与旋轮之间存在间隙, 间隙决定了毛坯加工完毕后的壁厚。芯模自转并沿其母线方向与旋轮做相对运动(筒形件立式旋压)或芯模自转, 旋轮沿芯模母线方向与芯模做相对运动(筒形件卧式旋压)^[2-3]。

旋压时, 由于直接用锻环旋压成形, 无纵向焊

缝, 故相同壁厚的旋压筒较之卷焊筒承载能力有所提高。在相同的载荷条件下, 旋压筒的壁厚较薄, 从而使消极重量减轻。并且由于不需要纵向焊缝焊接设备, 可简化圆段的制造工艺^[2]。此外由于旋压成形的筒体无纵向焊缝, 是减少低应力爆破的一种有效措施。

动力所采用三旋轮卧式强力旋压机对直径 $\Phi 70$ mm的毛坯进行强力旋压时, 由于毛坯尺寸存在差异, 存在旋压过程中实际减薄率波动问题, 会造成金属材料回弹量以及旋压三向应力波动, 从而影响旋压壳体成形后的形位精度。

毛坯加工完毕后, 会采用中部跳动和Q处跳动来表示测量的行位精度, 当中部跳动大于0.8, Q处跳动大于0.6时, 通常认为旋压壳体存在质量问题, 会严重影响后续的制造加工工序。

为了提高强力旋压的壳体良品率, 尝试采用机器学习方式建立模型, 预测中部跳动和Q处跳动, 为将来能够发现优化对象和范围, 降低中部跳动和Q处跳动提供研究基础。

收稿日期: 2022-09-01; 修回日期: 2022-10-20

作者简介: 糜旗(1979—), 男, 上海人, 硕士, 高级工程师, 从事信息安全、系统架构、逆向工程、机器学习研究。

E-mail: zeror@163.com.

1 机器学习模型

1.1 XGBoost 模型

XGBoost(extreme gradient boosting)是一种集成学习算法,其原理是由多棵决策树迭代计算^[4],最后所有决策树计算结果累计在一起得到最终答案^[5]。高效地实现 GBDT (gradient boosting decision tree)算法并进行了算法和工程上的许多改进,为更快更准地下降梯度而选择使用二阶导数,并将选取损失函数与优化算法进行分离,可以根据实际情况选择损失函数;因此,也使得 XGBoost 更具适用性,分类与回归都可以兼顾使用。

1.2 LightGBM 模型

LightGBM(light gradient boosting machine)是一个实现 GBDT 算法的框架^[6],支持高效率的并行训练,并且具有更快的训练速度、更低的内存消耗、更好的准确率及支持分布式可以快速处理海量数据等优点^[7-9]。LightGBM 采用分裂策略为 Leaf-Wise,遍历当前全部叶子节点,寻找到最大收益的节点进行分裂,但在递归中极易陷入较高的深度,出现过拟合,因此需要限制最大深度。另外 LightGBM 将父节点直方图减去兄弟节点直方图,计算出子节点直方图,实现差加速^[10]。

2 旋压数据预测模型

为了解决遇到的实际问题,采集毛坯加工过程中的原始数据,原始数据中包含了毛坯几何尺寸数据,设备时序数据和壳体跳动数据。拟使用机器学习等数据挖掘技术,通过数据驱动,分析出几何参数、时序数据与跳动数据的对应关系,通过构建预测模型,可以实现形变溯源、参数优化和控制成本。

2.1 数据分析处理

将原始数据集分成训练集与测试集,数据结构完全一致。训练集包含共 168 个毛坯的旋压加工前毛坯几何尺寸数据,每个毛坯旋压加工中设备时序数据和旋压加工后壳体跳动数据。其中毛坯几何尺寸数据在旋压过程中保持不变,时序数据的采样频率为 2 s,每次毛坯加工完的时长为 8 min 左右。

采取向前填充的方法对旋压加工中缺失的毛坯几何尺寸数据和壳体跳动数据进行填充。填充完毕后,整个数据集共分 3 部分。第 1 部分为毛坯几何尺寸数据,包含了加工时间、加工顺序、工件编号,毛坯几何尺寸等 13 个字段;第 2 部分为壳体跳动数

据,包含中部跳动、Q 处跳动和备注 3 个字段;第 3 部分为设备时序数据,大致为 18 个字段。如表 1 所示,0-12 为第一部分;13-15 为第二部分;17-34 为第三部分。

表 1 数据集结构

序号	字段	非空值数量	类型
1	加工时间	168 non-null	object
2	加工顺序	168 non-null	float64
3	工件编号	168 non-null	float64
4	小头壁厚 5±0.03	168 non-null	float64
5	Unnamed: 4	168 non-null	float64
6	内孔中 39.4±0.02	168 non-null	float64
7	斜坡壁厚 2.5±0.03	168 non-null	float64
8	Unnamed: 7	168 non-null	float64
9	大段壁厚 2.45±0.03	168 non-null	float64
10	Unnamed: 9	168 non-null	float64
11	大端内控φ 69.84	168 non-null	float64
12	Unnamed: 11	168 non-null	float64
13	总长 352±0.3	168 non-null	float64
14	中部跳动≤0.8	168 non-null	float64
15	Q 处跳动≤0.6	168 non-null	float64
16	备注	168 non-null	object
17	Unnamed: 16	168 non-null	float64
18	W3. DATAPOINT. X_R	35 389 non-null	float64
19	W3. DATAPOINT. X_L	35 389 non-null	float64
20	W3. DATAPOINT. ZR	35 287 non-null	float64
21	W3. DATAPOINT. Z_L	35 286 non-null	float64
22	W3. DATAPOINT. X2_R	35 389 non-null	float64
23	W3. DATAPOINT. X2_L	35 389 non-null	float64
24	W3. DATAPOINT. X3_R	35 389 non-null	float64
25	W3. DATAPOINT. X3_L	35 384 non-null	float64
26	W3. DATAPOINT. FEED_ACT	35 389 non-null	float64
27	W3. DATAPOINT. FEED_SET	35 389 non-null	float64
28	W3. DATAPOINT. SPINDLE ACT	35 389 non-null	float64
29	W3. DATAPOINT. SPINDLE SET	35 389 non-null	float64
30	W3. DATAPOINT. SPINDLE PER	35 389 non-null	float64
31	W3. DATAPOINT. CODE L1	35 369 non-null	object
32	W3. DATAPOINT. CODE L2	34 541 non-null	object
33	W3. DATAPOINT. CODE_L	335 389 non-null	object
34	W3. DATAPOINT. C_R	35 389 non-null	float64
35	W3. DATAPOINT. C_L	35 389 non-null	float64

训练集共存在 34 个字段,包含了加工时间、加工顺序、工件编号、毛坯几何尺寸、设备时序数据、跳动数据和质检信息等。

数据集呈现出 3 个特点:大部分数据类型为 object 和 float;部分时序数据的字段存在较多缺失值,选择向前填充的方法,数值型数据通过计算均值后填充;个别字段存在单一值对象。

2.2 特征提取

经过前期对数据的简单分析后,尝试从训练集中筛选出更好的数据特征,优质的数据预处理可以高效率高质量的训练模型,好处在于:1) 减少训练数据大小,加快模型训练速度;2) 减少模型复杂度,避免过拟合;3) 特征数少,有利于解释模型;4) 如果选择到正确的特征子集,模型准确率可能会

提升。

在原始数据集中，数据常会呈现出缺失错位的现象，往往无法直接使用，可以采取归一化、标准化和正则化等方法来改善，以提升后期模型的训练效果。

在进行特征提取的过程中，首先严格遵循了真实加工过程中的 3 个工艺逻辑：1) 加工顺序，反映了刀具的磨损情况，随着加工次数增加磨损也随之严重；2) 加工前工艺参数，反映了毛坯原始尺寸以及尺寸极差；3) 旋压加工工艺参数设置值，直接影响了加工精度。

其次在数据处理过程中，采用可视化数据探索性分析(exploratory data analysis, EDA)模式，对 4 个量级最大的特征进行缩放，以缩小模型偏差影响，同时去除了加工时长不足 7 min 的异常样本。

最终形成 2 类特征：1) 固有特征，为毛坯几何尺寸数据，直接使用无需计算，如表 2 所示；2) 统计特征，为设备时序数据统计计算值，包括均值、标准差、最小值、最大值、5%分位数、25%分位数、50%分位数、75%分位数、95%分位数，部分特征数据如表 3 所示。

表 2 固有特征

序号	加工时间	加工顺序	工件编号	小头壁厚_1	小头壁厚_2	内孔	科坡壁厚_1	科坡壁厚_2
0	2021.5.18	1.0	210625.0	5.03	5.02	39.42	2.53	2.51
1	2021.5.18	2.0	210626.0	5.01	4.99	39.43	2.51	2.50
2	2021.5.18	3.0	210627.0	5.00	5.03	39.44	2.51	2.53
3	2021.5.18	4.0	210628.0	5.02	5.02	39.45	2.51	2.52
4	2021.5.18	5.0	210629.0	5.02	5.02	39.43	2.52	2.53

序号	大端壁厚_1	大端壁厚_2	大端内孔_1	大端内孔_2	总长	备注	中部跳动≤0.8	中部跳动≤0.6
0	2.45	2.45	69.850	69.84	351.96	NaN	0.330	0.269
1	2.47	2.46	69.850	69.84	352.04	NaN	0.143	0.090
2	2.46	2.46	69.880	69.84	352.02	NaN	0.278	0.249
3	2.46	2.47	69.840	69.87	351.98	NaN	0.350	0.315
4	2.48	2.46	69.860	69.87	351.99	NaN	0.263	0.233

表 3 统计特征

序号	均值	标准差	最小值
0	19.491 549	33.505 566	3.471 047
1	19.321 568	32.933 074	3.432 000
2	18.595 967	30.268 088	3.473 721
3	18.574 734	30.201 627	3.473 465
4	18.967 537	31.625 226	3.470 000
...

序号	5%分位数	25%分位数	50%分位数
0	3.493 128	15.960 722	16.000 578
1	3.548 186	15.870 606	16.019 197
2	3.552 233	16.005 408	16.029 599
3	3.546 579	16.004 441	16.028 968
4	3.497 479	15.971 906	16.007 743
...

2.3 模型构建

预测模型的评估指标采用均方根误差(root mean squard error, RMSE)，也称为标准误差，如式(1)所示，其中 n 为测量次数； $(X_{obs, i} - X_{model, i})$ 为 1 组预测值与真实值的偏差，先计算预测值与真实值之间偏差的平方，再除于预测次数 n ，最后计算商的平方根^[11-14]。在模型构建时，真实值为旋压加工后的实际测量值。

$$RMSE = \sqrt{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2 / n} \quad (1)$$

由于 RMSE 对极大和极小的误差值特别敏感，因此成为评估模型预测精度的标准，RMSE 值越大表示预测精度越低，反之则表示预测精度越高，表

达了预测值与真实值的偏离程度^[15]。

交叉验证也被称为循环验证。在机器学习建模过程中，经常用来评估模型。拟采用五折交叉验证法来预测模型。

先将原始数据集平均分成 5 组子集数据，分 5 次，每次选取不同的一组子集数据作为验证集，另外 4 组子集数据则作为训练集^[16-17]。每次训练都可以产生一个模型，循环 5 次后会产生 5 个模型。然后分别将 5 个模型在验证集中进行评估，会得到 5 个不同的 RMSE 评估结果。最后将 5 个 RMSE 加和平均就能计算出交叉验证误差，如图 1 所示。

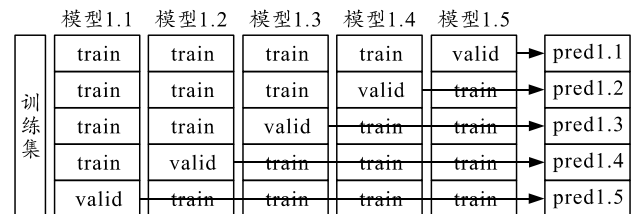


图 1 五折交叉验证

当数据集中只存在有限的数据时，选择交叉验证是一种较为有效的方法，评估结果也与模型在测试集上的表现较为接近^[18-19]，也可被用作模型优化。

共选择了 2 类不同模型，分别为 LightGBM 和 Xgboost，模型调参后，2 类模型对应的 RMSE 最优

结果为 0.105 5 和 0.123 3，因此选择 LightGBM 为最终使用模型。

基于 LightGBM 模型可进行特征重要度评价，采用各特征在 LightGBM 模型训练过程中带来的平均信息增益为特征重要度打分，LightGBM 特征重要度排序如图 2 所示。

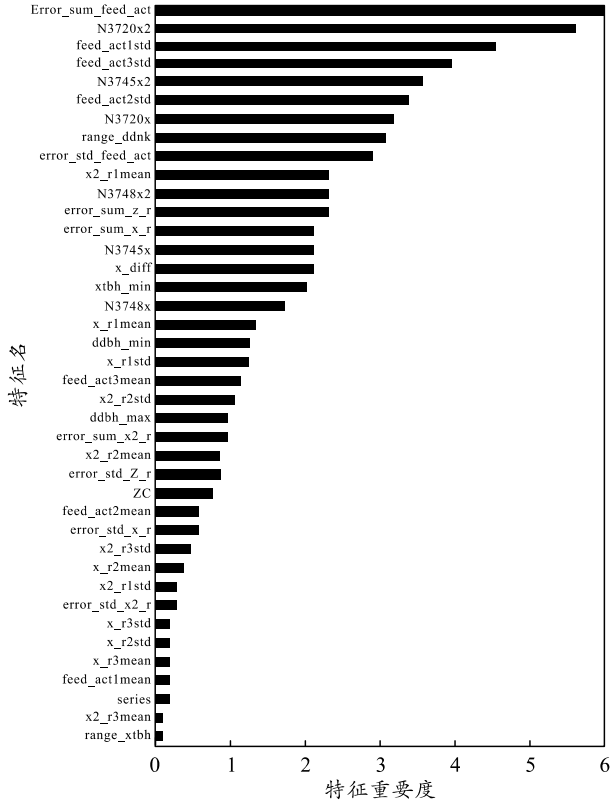


图 2 特征重要度排序

根据特征重要度排序结果，采用交叉验证的方式可以确定最优特征数。当特征数为 10 的时候 RMSE 分值最低，特征数过少或过多都将导致模型预测误差大幅增大，如图 3 所示。

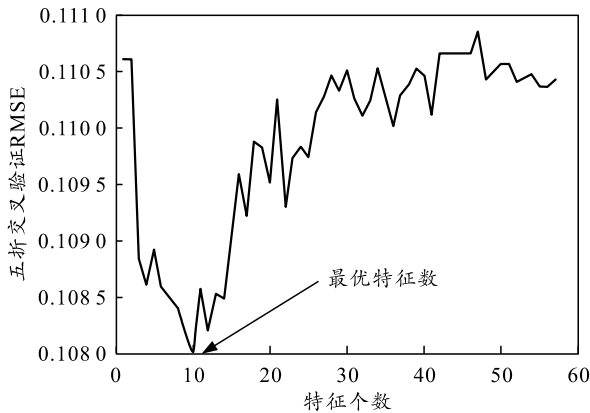


图 3 最优特征数分布曲线

2.4 模型预测效果

综合以上的特征工程和机器学习，最终由

LightGBM 建立构建模型。在全新的测试集上，对训练完毕的模型进行最终评估，中部跳动预测结果如图 4 所示，Q 处跳动预测结果如图 5 所示，RMSE 最终得分为 0.098 9。

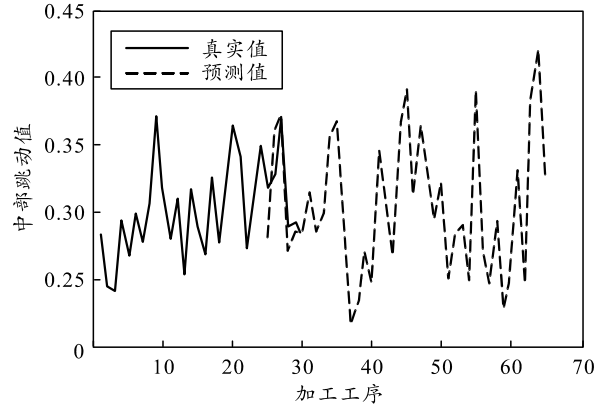


图 4 中部跳动模型预测

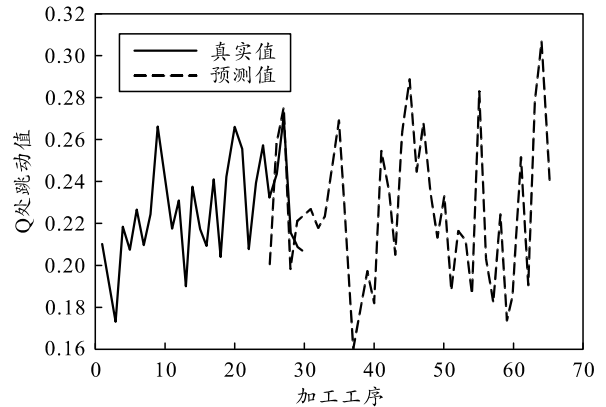


图 5 Q 处跳动模型预测

通过构建多重机理特征与统计特征，融合工艺参数与时序的多源异构数据，可以有效地预测强力旋压后壳体的中部跳动和 Q 处跳动。

3 结束语

在强力旋压的特定加工场景中，可以通过机器学习建立模型，模型具有较好的泛化能力，可以定量分析各特征与壳体跳动的相关性。下一步工作将对训练集开展无约束的多目标优化，采用 NSGA-II 算法，发现优化对象和优化范围，进一步降低中部跳动和 Q 处跳动。

参考文献：

[1] 魏江峰, 常智勇, 莫蓉, 等. 固体火箭发动机燃烧室壳体旋压毛坯制造工艺技术研究[J]. 机械科学与技术, 2009, 28(1): 56-59.

[2] 邱新杰. 固体火箭发动机壳体的旋压和旋前坯预成形工艺的改进[J]. 宇航材料工艺, 1985(6): 48-49.