



团队：Micro-i 上海微亿智造科技

愿景：以工业人工智能及大数据技术助力工业企业智能化转型

成员介绍：

姓名	单位	学校	学历
王闯	上海微亿智造 /Micro-i	上海交通大学	硕士
顾徐波		MIT/上海交通大学	在读博士
宋怡然		UNC/上海交通大学	在读硕士



任务A 虚拟量测

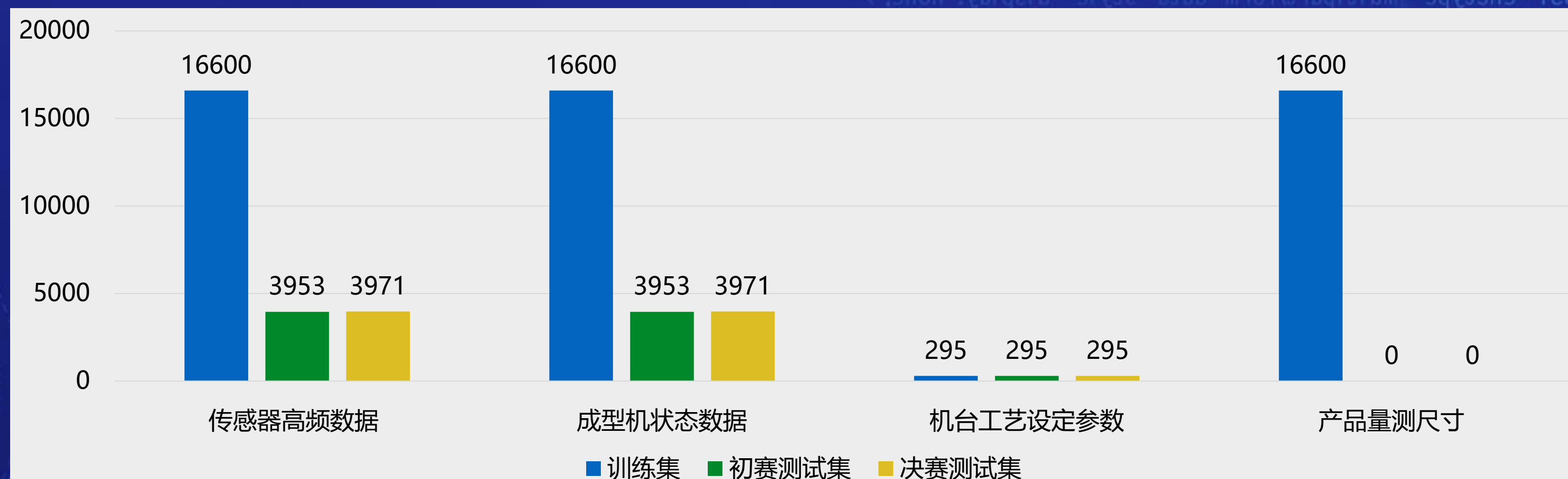
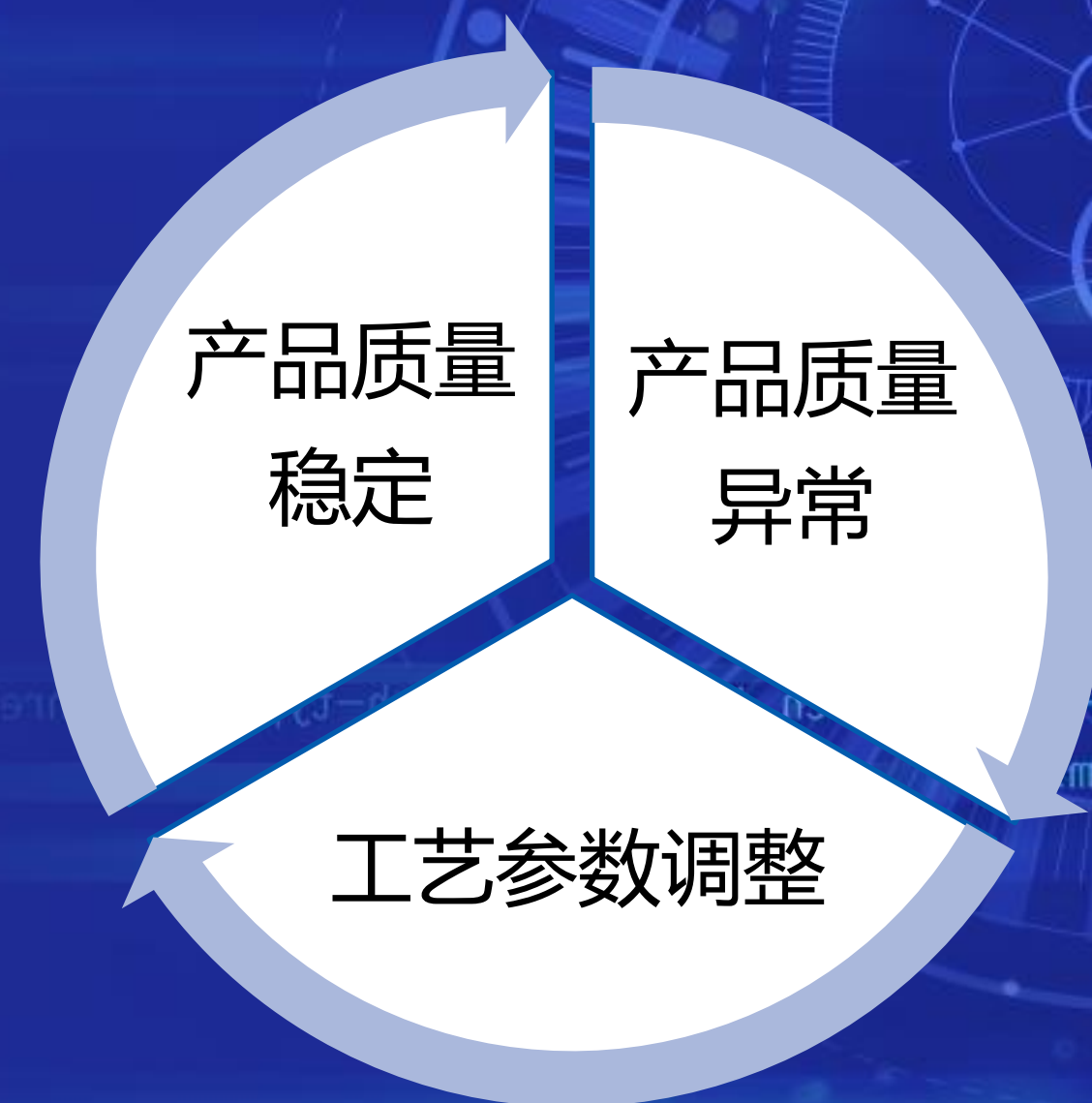
任务B 调机优化

后续建议及思考



虚拟量测-任务分析

- **目的：**针对成型工艺品质异常中**尺寸超规问题**进行虚拟量测
- **背景：**注塑产品生产过程中，由于所处环境、工况的变化导致系统状态迁移，从而导致产品质量由稳定转向异常，需要**人为工艺参数调整**，使系统回归稳定生产状态。
- **数据：**产品质量数据
生产过程数据（传感器高频数据 成型机状态数据）
机台工艺设定参数

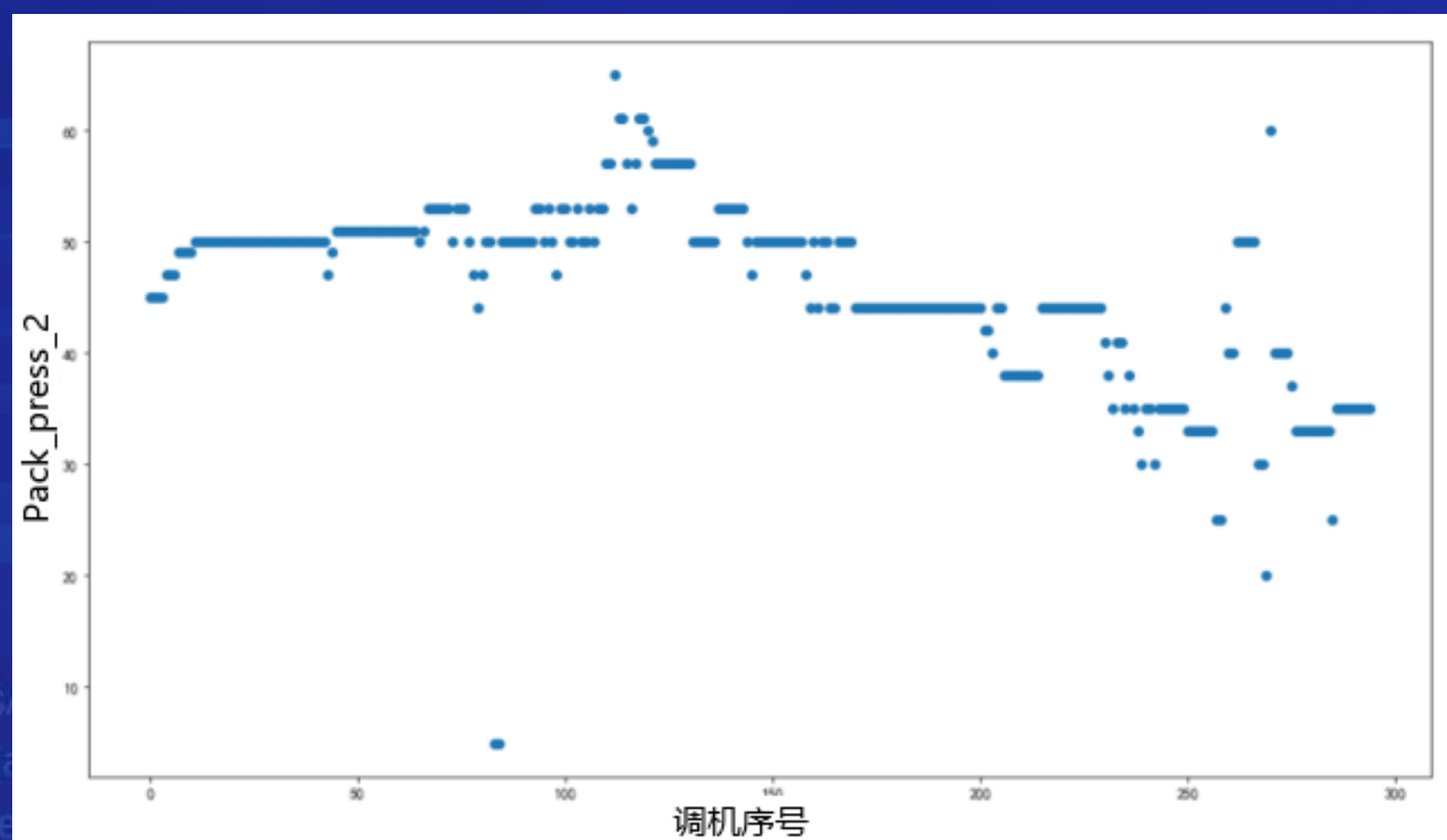




虚拟量测-EDA

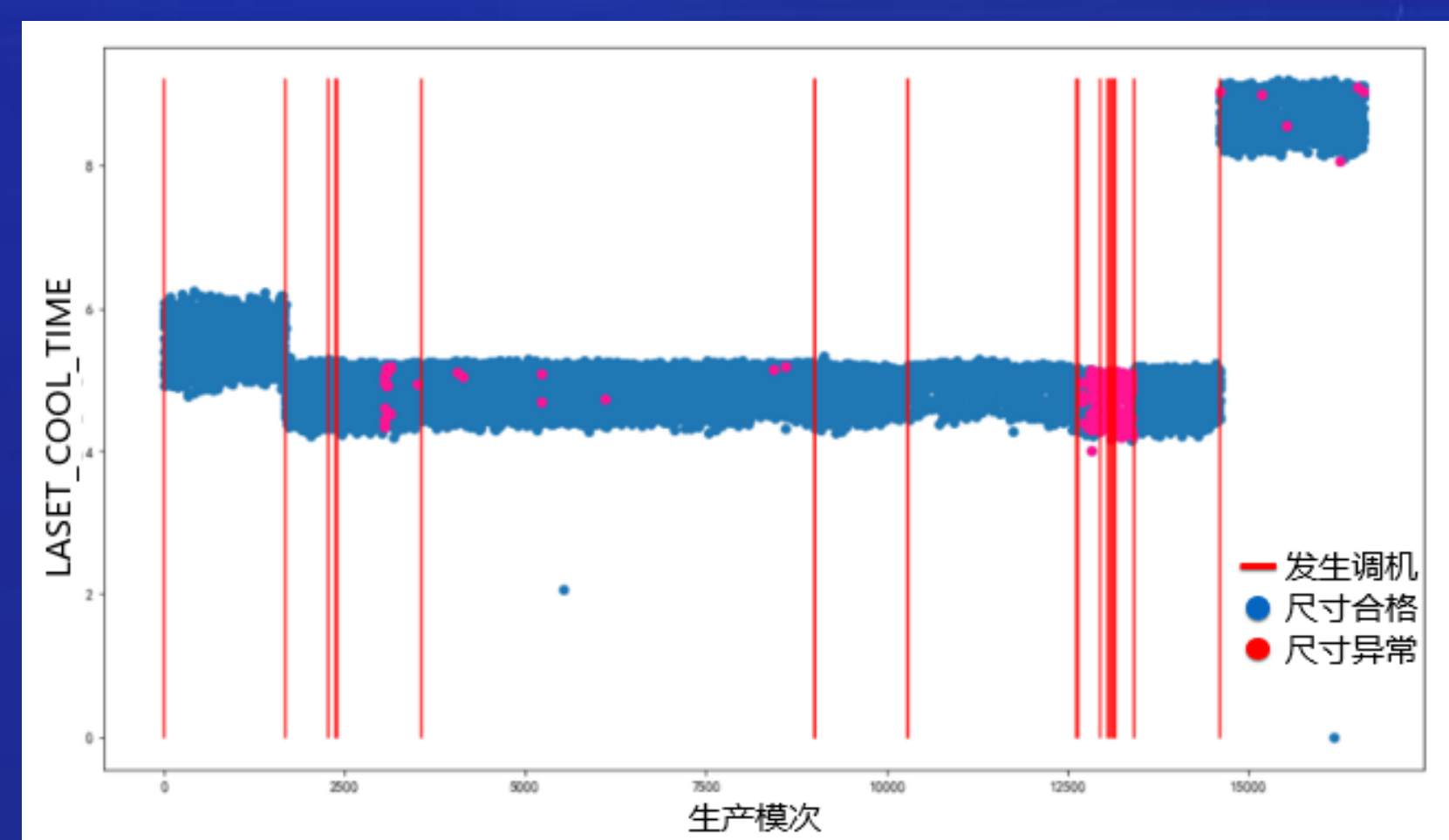
机台工艺参数

- 数据离散
- 参数变化时记录



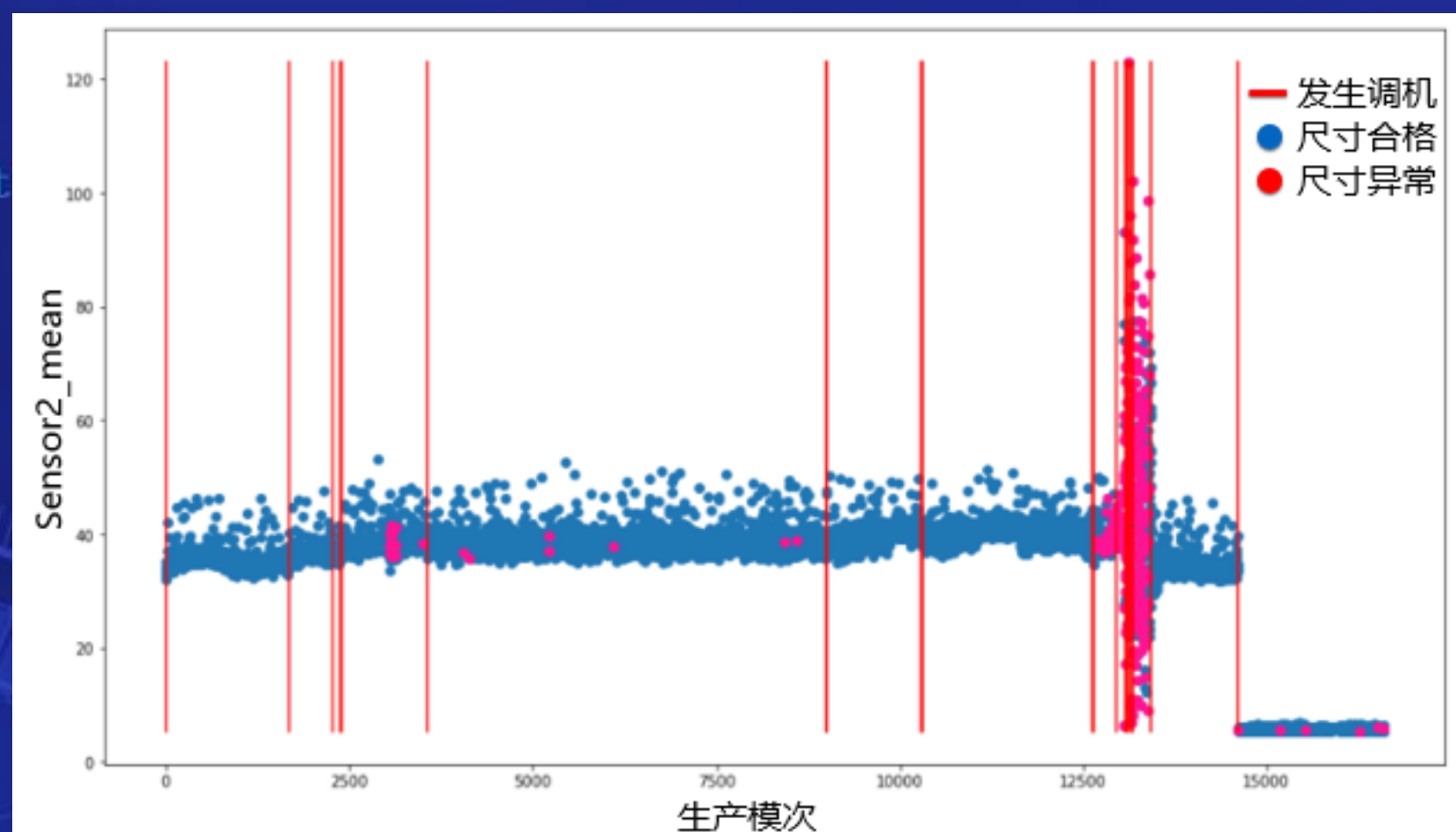
成型机状态参数

- 数据离散
- 每条数据对应一模次



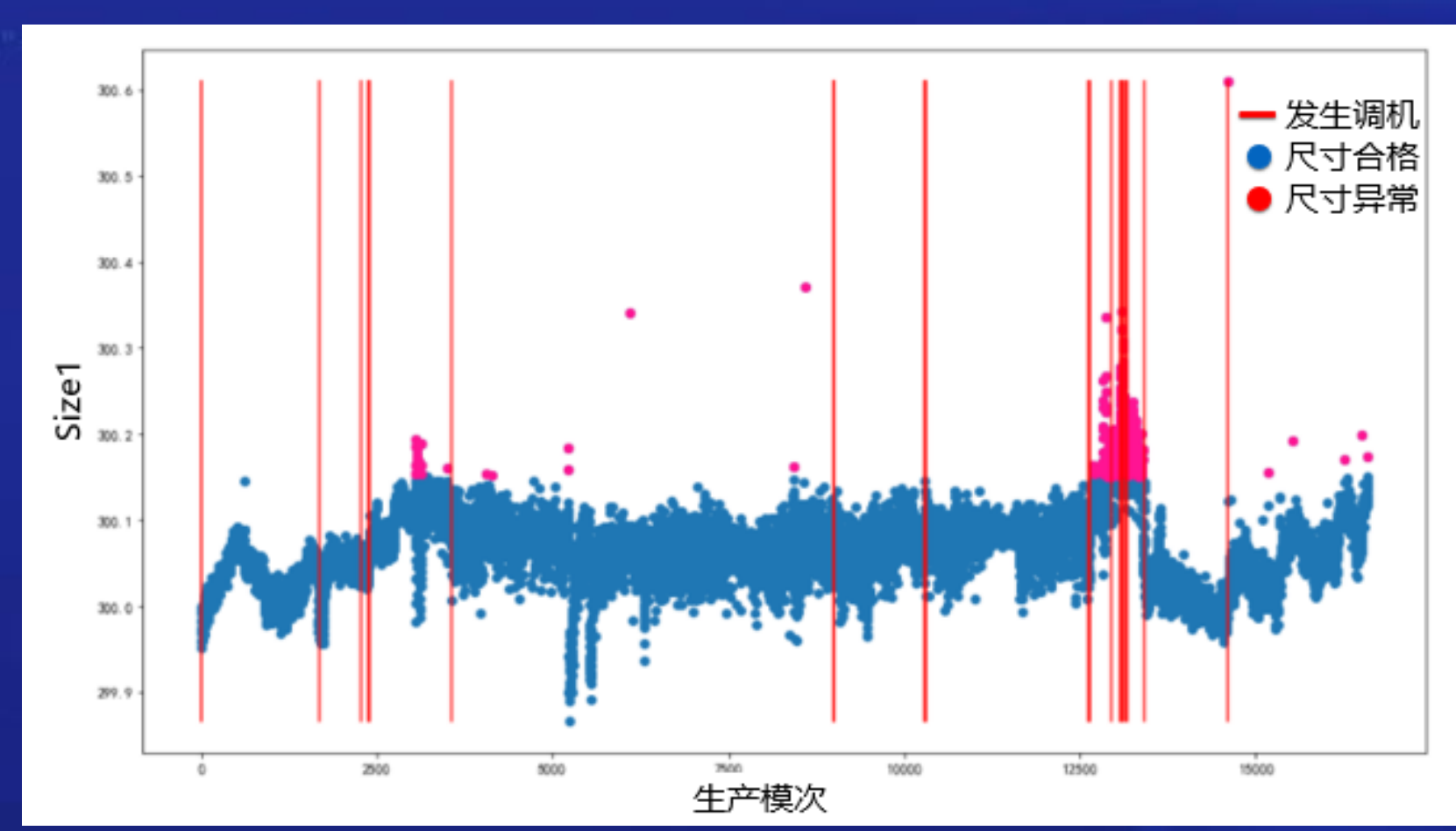
传感器数据

- 连续波动
- 取均值处理



尺寸数据

- 数据连续
- 分段波动





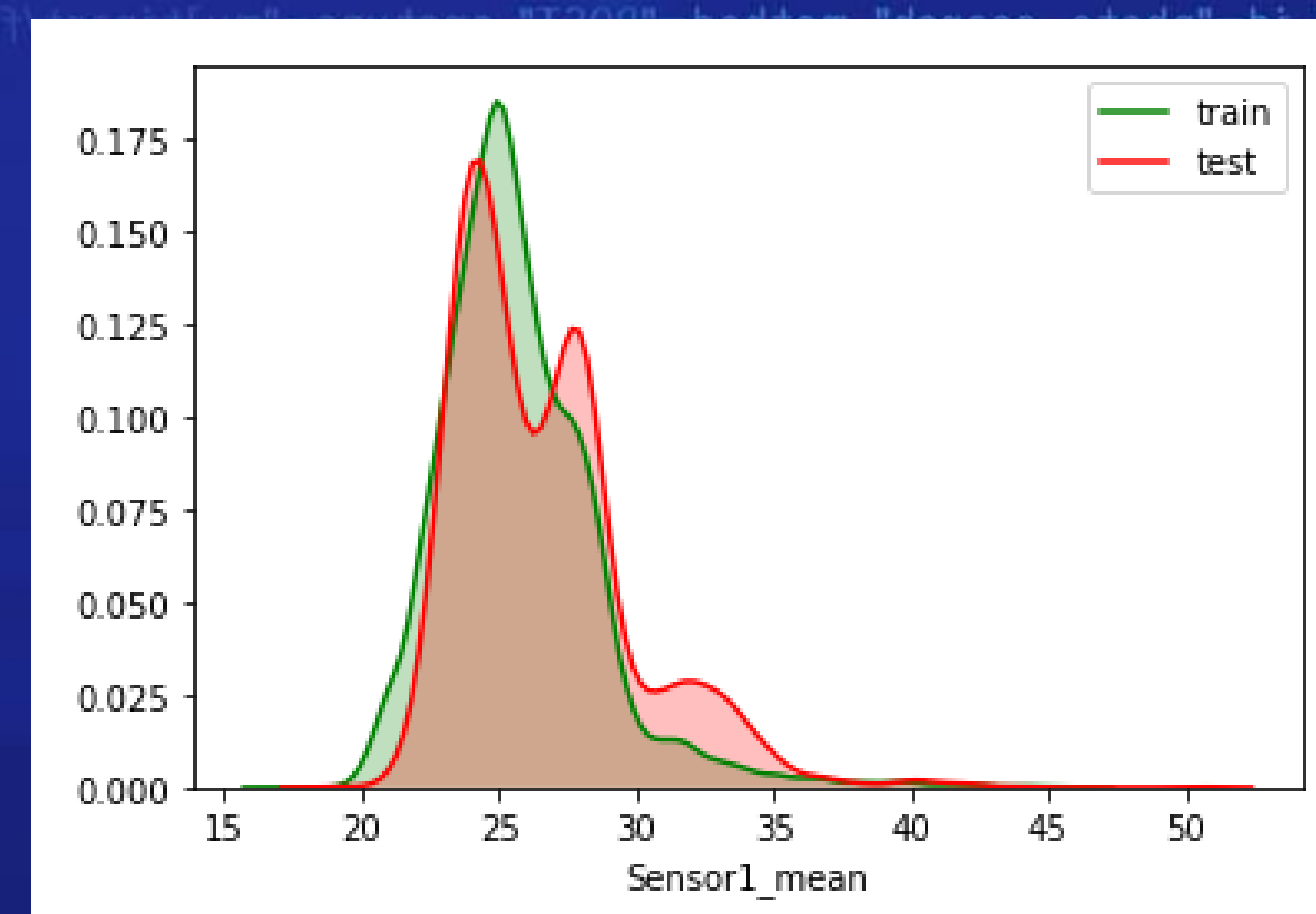
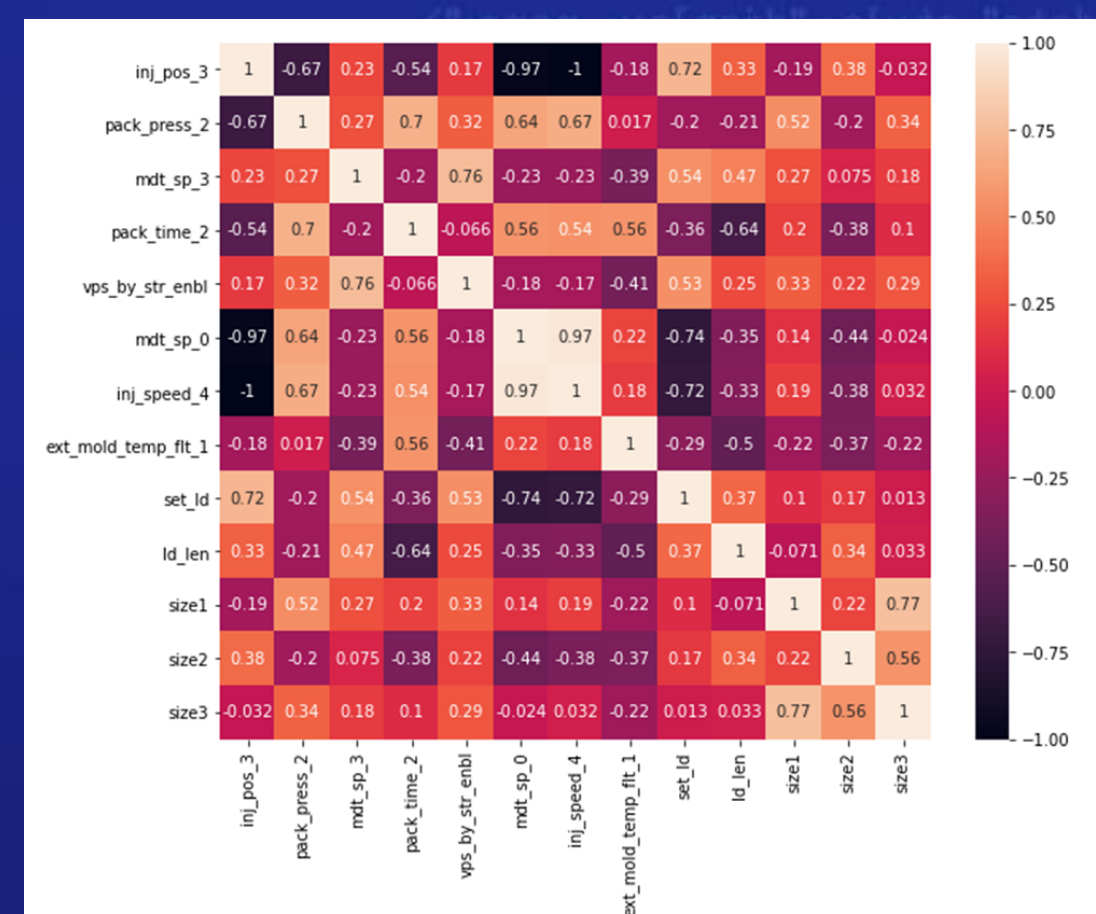
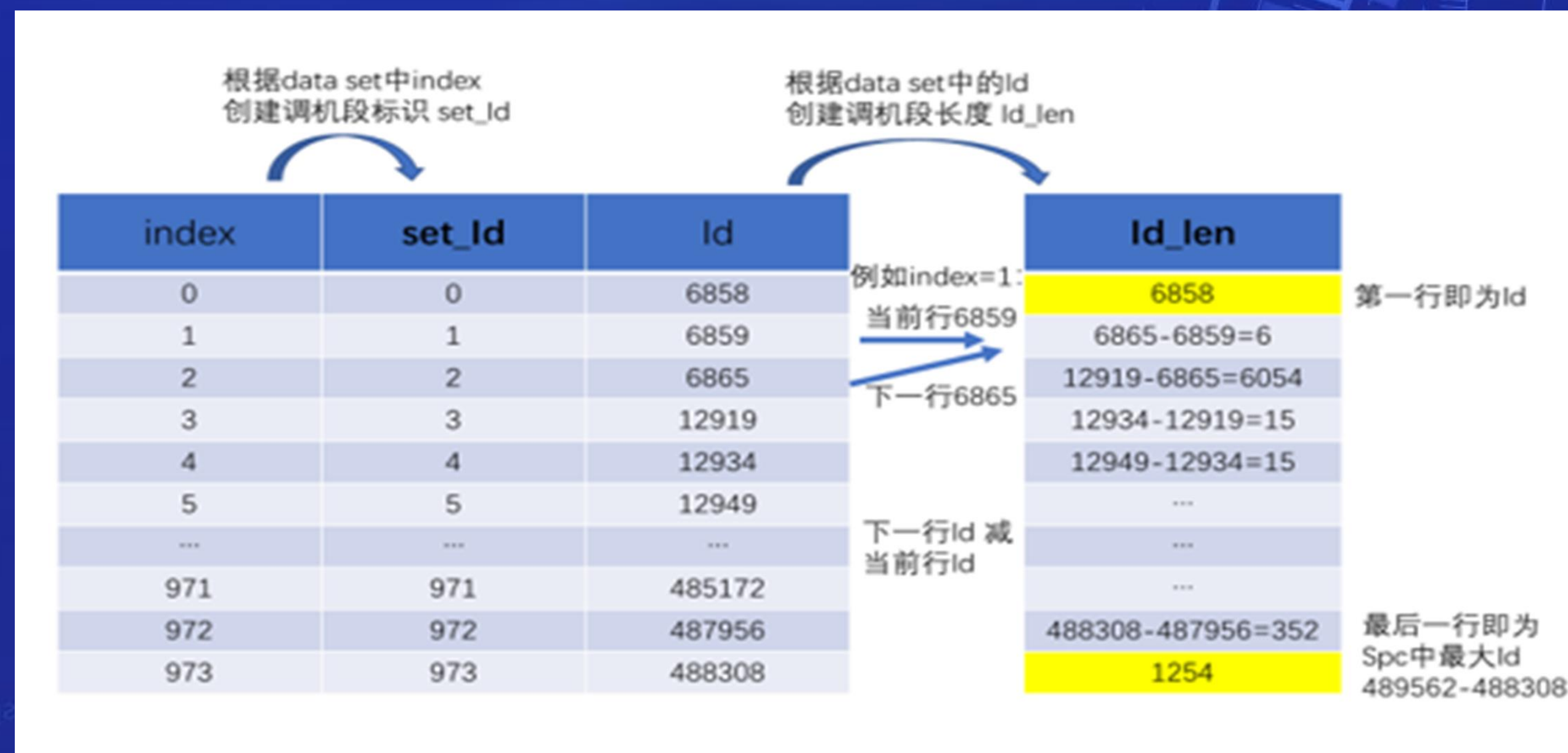
虚拟量测-特征工程

Step 1 特征构造

- 打标：创建调机段Id特征，对成型机，传感器等数据打上标签
- 分组+特征提取：根据调机段对尺寸数据data size分组，统计不同调机段尺寸中位数、最大值等
- 均值：对高频传感器数据取均值处理

Step 2 特征筛选

- 数据相似性：特征与目标尺寸相关性系数
- 数据分布：特征在训练集和测试集分布





虚拟量测-技术架构

目标尺寸: lgbm模型

传感器、成型机和调机参数特征

Size衍生特征

根据相关性系数,
以及特征train和test
分布情况

基本特征筛选

创建调机段特征
及其他衍生特征

创建不同调
机段各size的
衍生特征

删除空特征及
单一值特征

对传感器信
号取均值

删除空特征及
单一值特征

传感器数据
Data sensor

成型机数据
Data spc

调机参数
Data set

尺寸参数
Data size



虚拟量测-验证过程

Step 3 模型验证

- 五折交叉
- 关注预测尺寸分布

注：
1. 基本特征*表示对传感器数据分不同工艺段抽取特征
2. stacking基学习器为 dbdt , ada , rf , svr , lgbm , xgb , 第二层学习器为lgbm

特征方案	模型	初赛得分
基本特征* 200维	stacking	3.7e7
基本特征根据分布筛选 45维	lgbm	4.4e6
基本特征根据分布筛选 + 调机参数衍生特征 46维	lgbm	2.0e6
基本特征根据分布筛选 + 调机参数、尺寸衍生特征 49维	lgbm	5.5e5
基本特征根据相关性、分布筛选 + 调机参数、尺寸衍生特征 22维	lgbm	7.5e4
基本特征根据相关性、分布筛选 + 调机参数、尺寸衍生特征 22维 (尺寸衍生特征中, 缺省值采用前值填充)	lgbm	1.2e5
基本特征根据相关性、分布筛选 + 调机参数、尺寸衍生特征 22维 (尺寸衍生特征中, 缺省值采用前值填充)	stacking	2.5e6
基本特征根据相关性、分布筛选 + 调机参数、尺寸衍生特征 22维 (对size2删除尺寸衍生特征, 对size3使用两段模型预测)	lgbm	2.8e4 (决赛3.1e4)



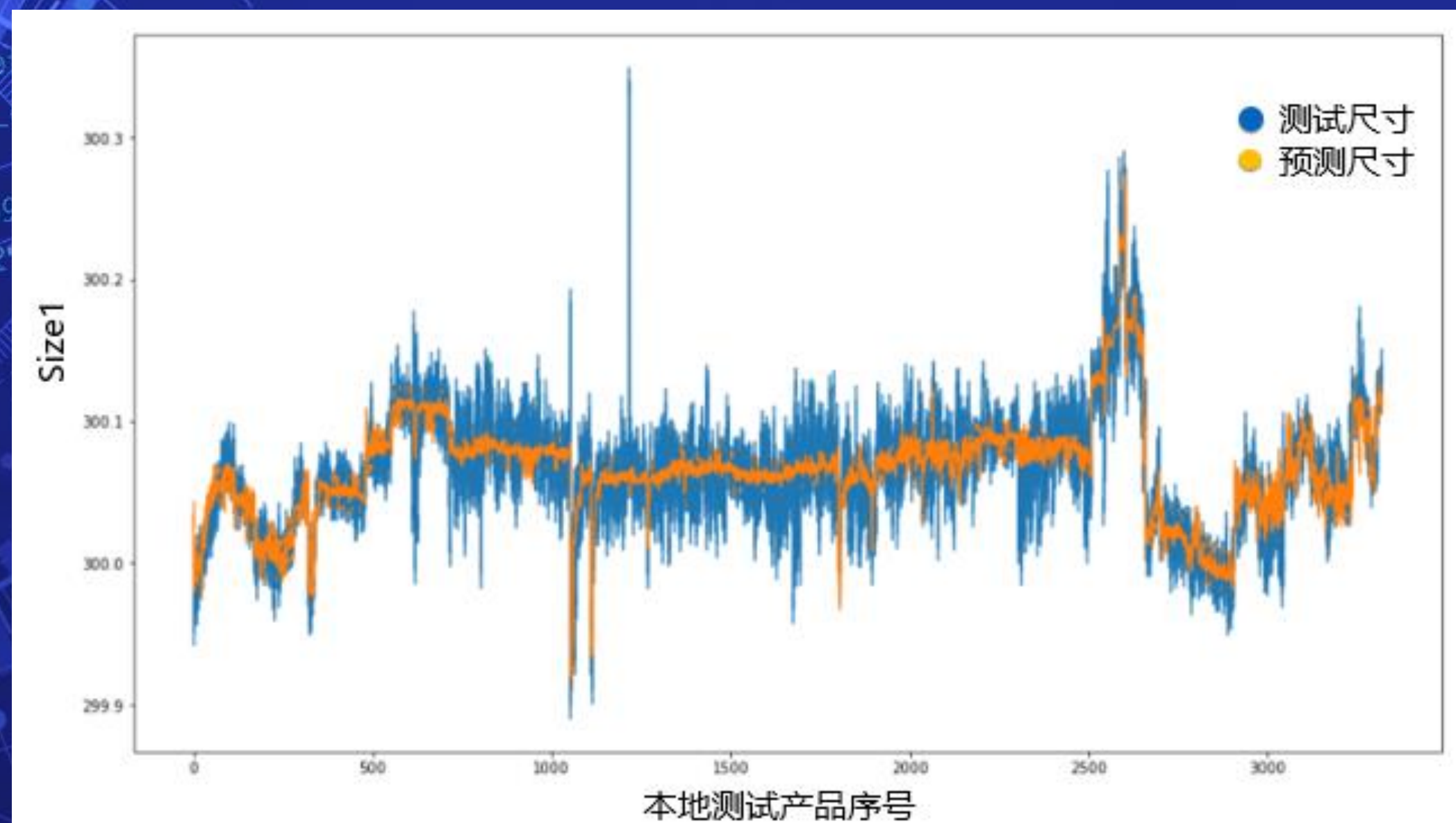
虚拟量测-方案特点

Step 4 模型预测

- 在本地验证集上，模型预测值对于尺寸趋势捕捉准确，但波动较小
- 预测值与实际值误差在0~0.04之间

模型优势

- 特征数量少，最终选用特征22维
- 模型简单，最终采用lgbm单模型
- 尺寸变化趋势捕捉准确，模型稳定性良好



22维特征

9维
高频传
感器特
征

7维
成型机
状态特
征

3维
调机参
数，包
括1维衍
生特征

3维
尺寸参
数衍生
特征



任务A 虚拟量测

任务B 调机优化

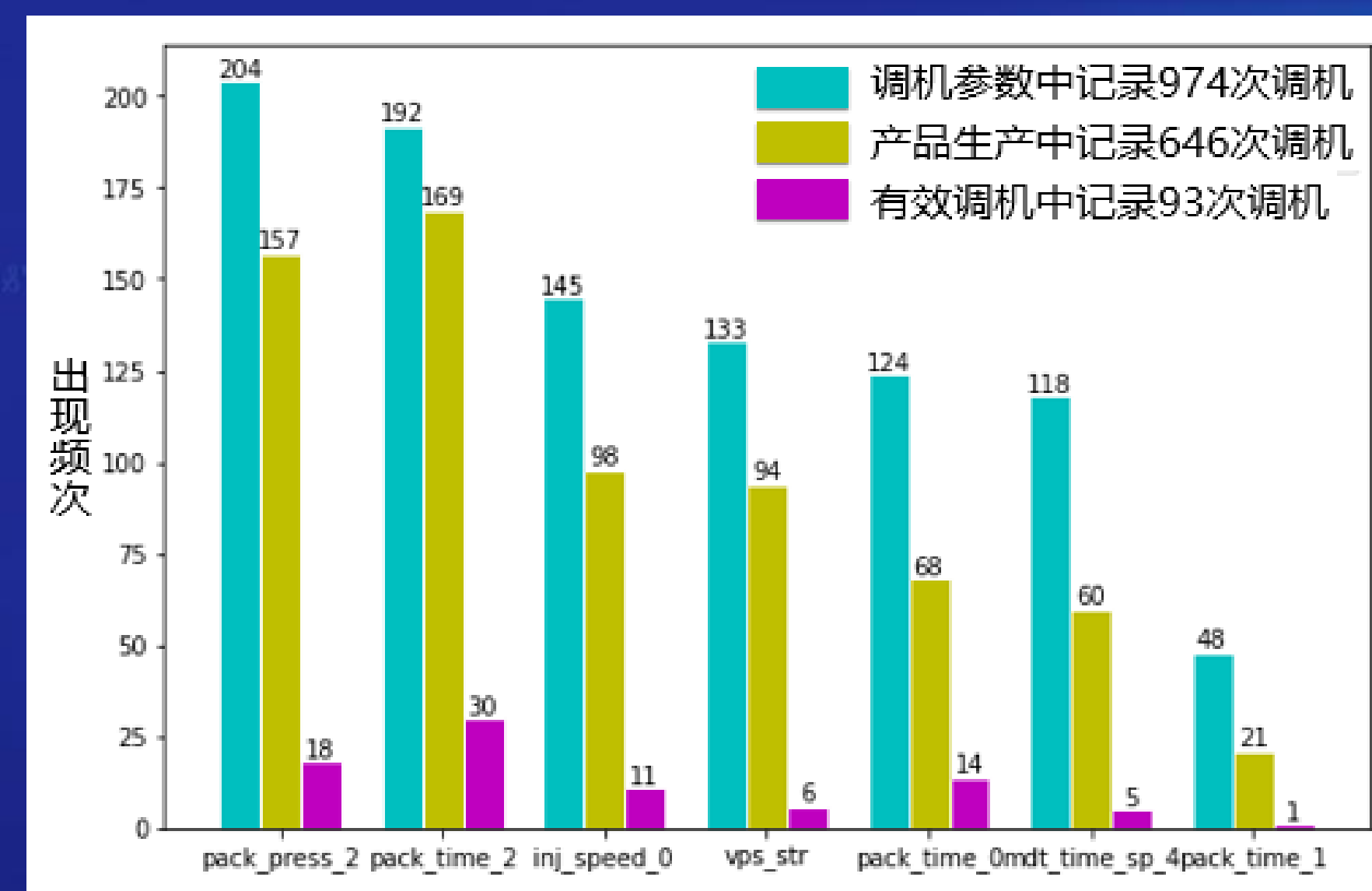
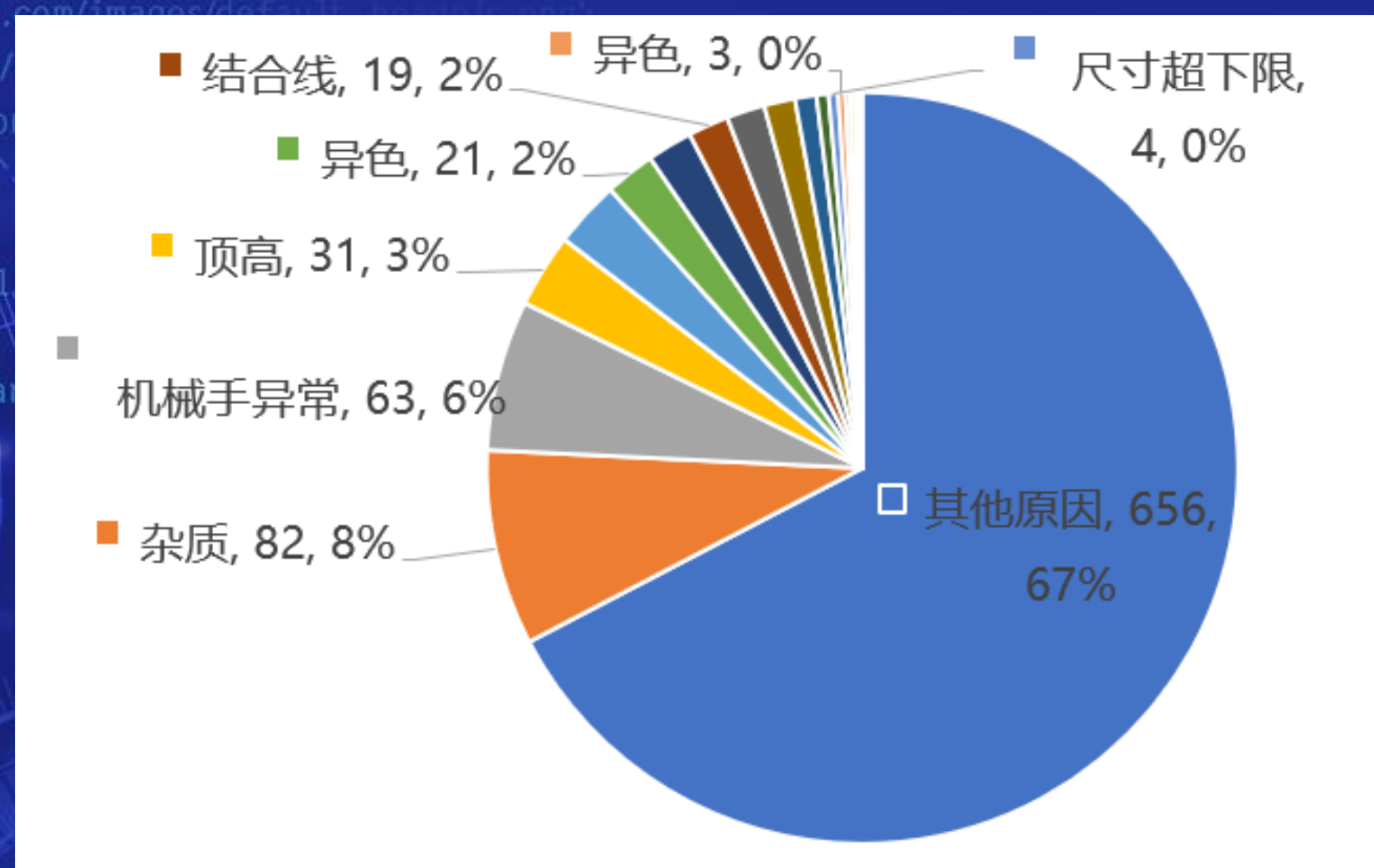
后续建议及思考



调机优化-人为操作

观点分析

- 异常调机的原因**：主要是杂质和机械手异常，而非尺寸超规。所以生产环境的稳定性，清洁状况及设备可靠性的提高，对于减少调机次数至关重要
- 有效调机比例低**：在进行的646次调机中，有效调机段93段，说明注塑机调机难度较大，但调节参数主要集中在前6个



注：有效调机定义，调机参数共记录974次调机，在提供的350283模次生产中，共646次，平均每次调机生产 $350283/646 \approx 542$ 次，将同一调机段内生产模次数大于542的调机段称为有效调机，共93段

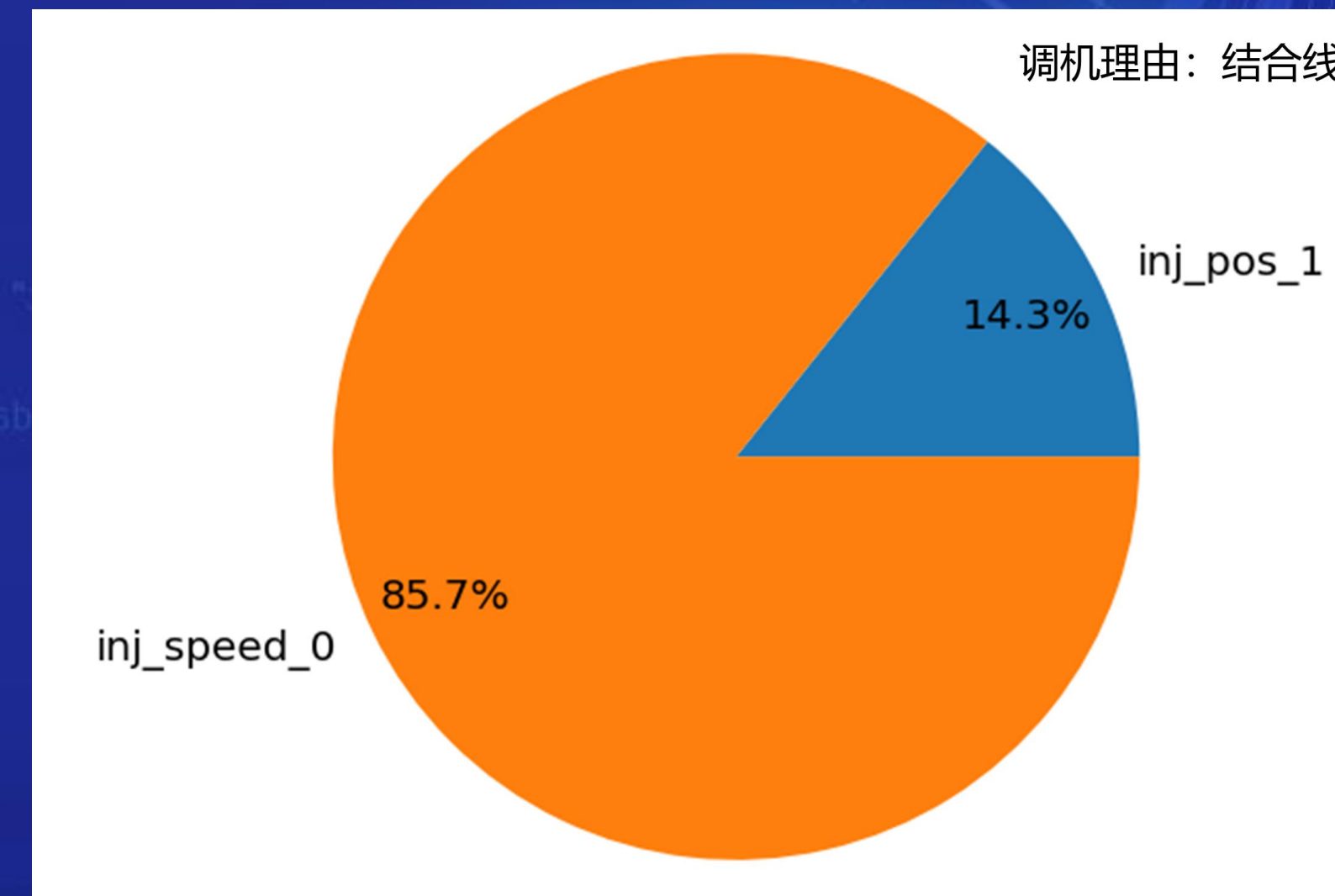
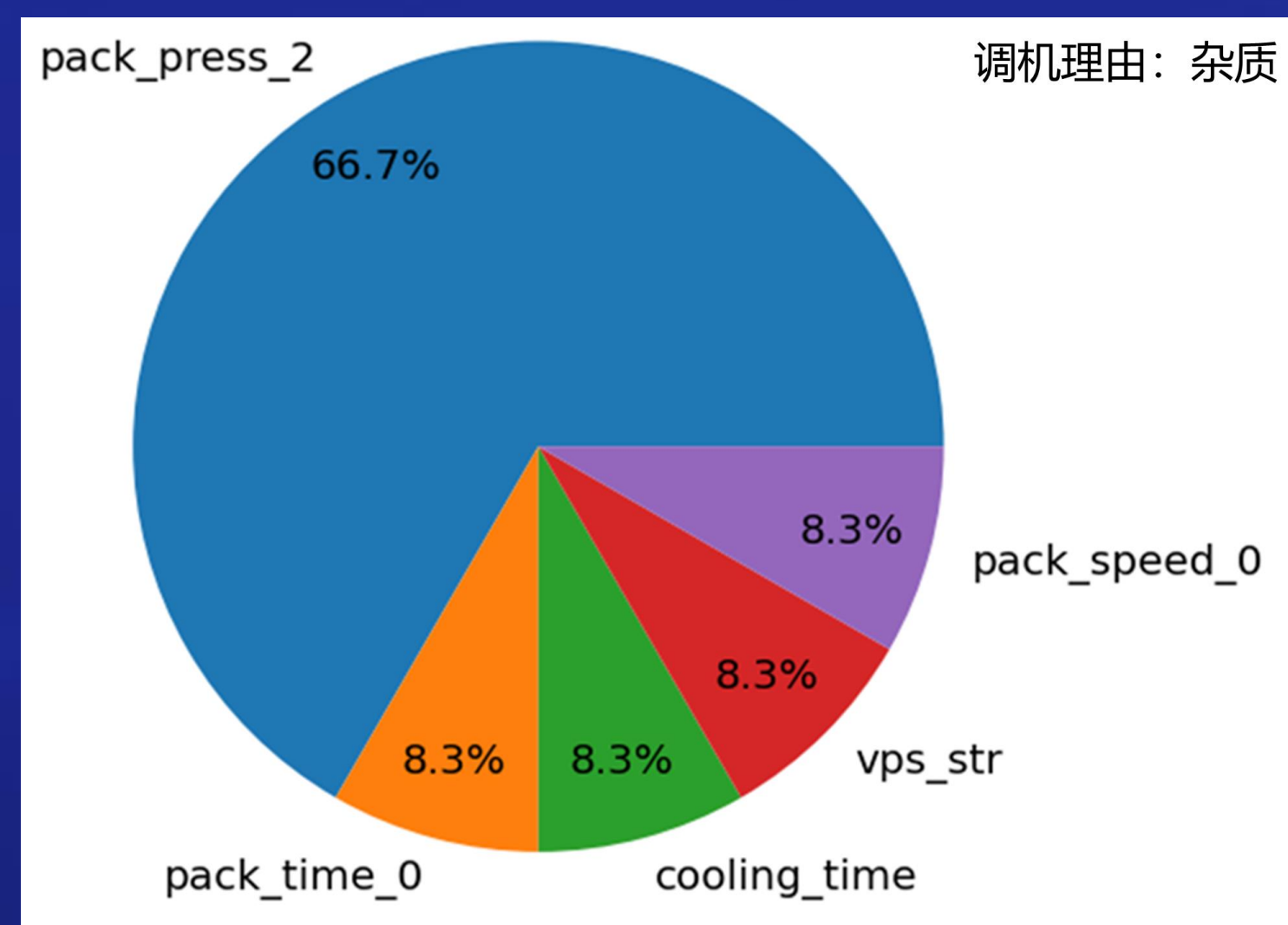
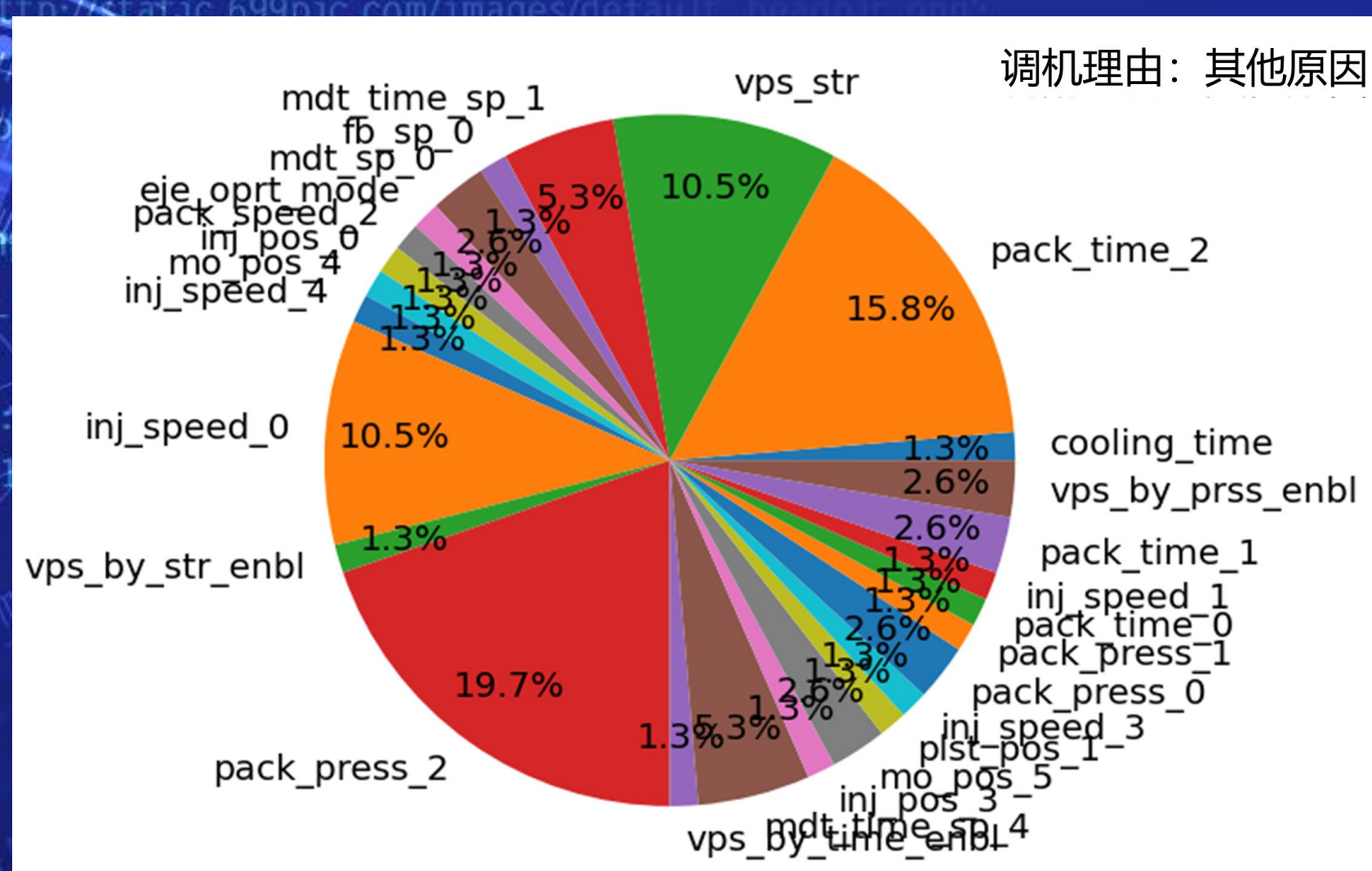


调机优化-人为操作

观点分析

数据进一步处理：只对有效调机的93次调参记录进行统计，对比93次有效调机段之间的调机参数变化情况

3. 特定原因下调参：根据调机原因对调机参数进行统计，可以发现在特定原因下调参规律，如当出现杂质问题时，66.7%的操作更改pack_press_2



注：有效调机定义，调机参数共记录974次调机，在提供的350283模次生产中，共646次，平均每次调机生产350283/646≈542次，将同一调机段内生产模次数大于542的调机段称为有效调机，共93段

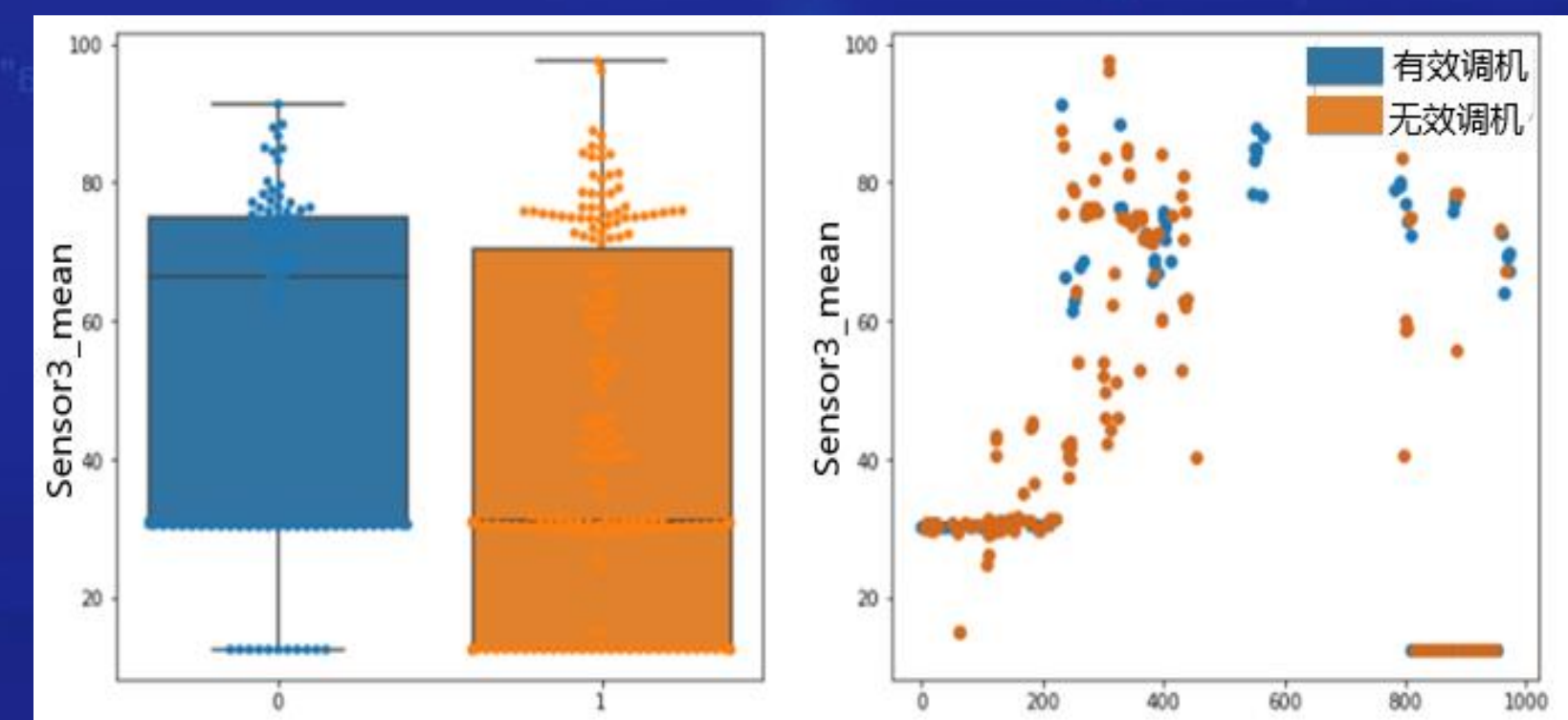
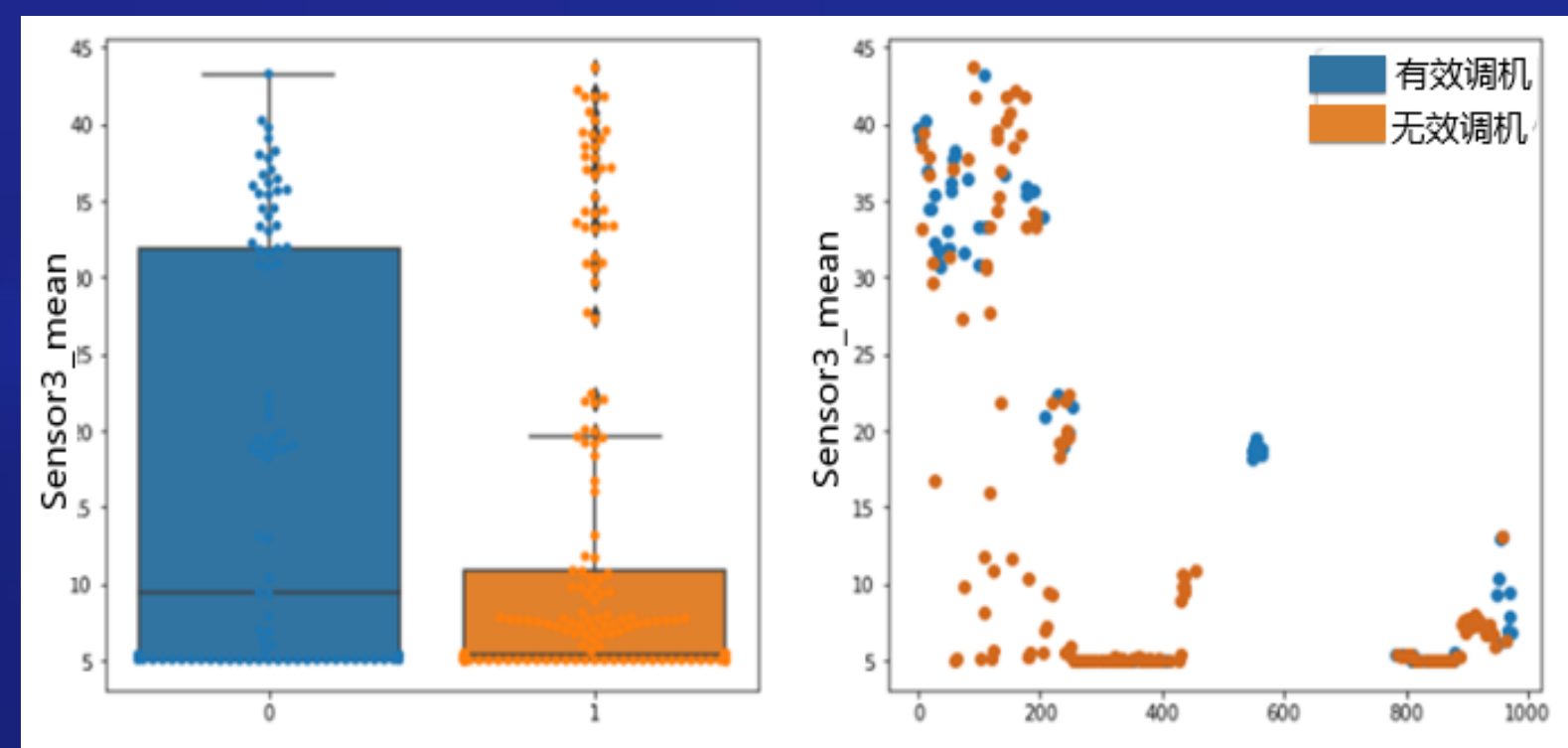
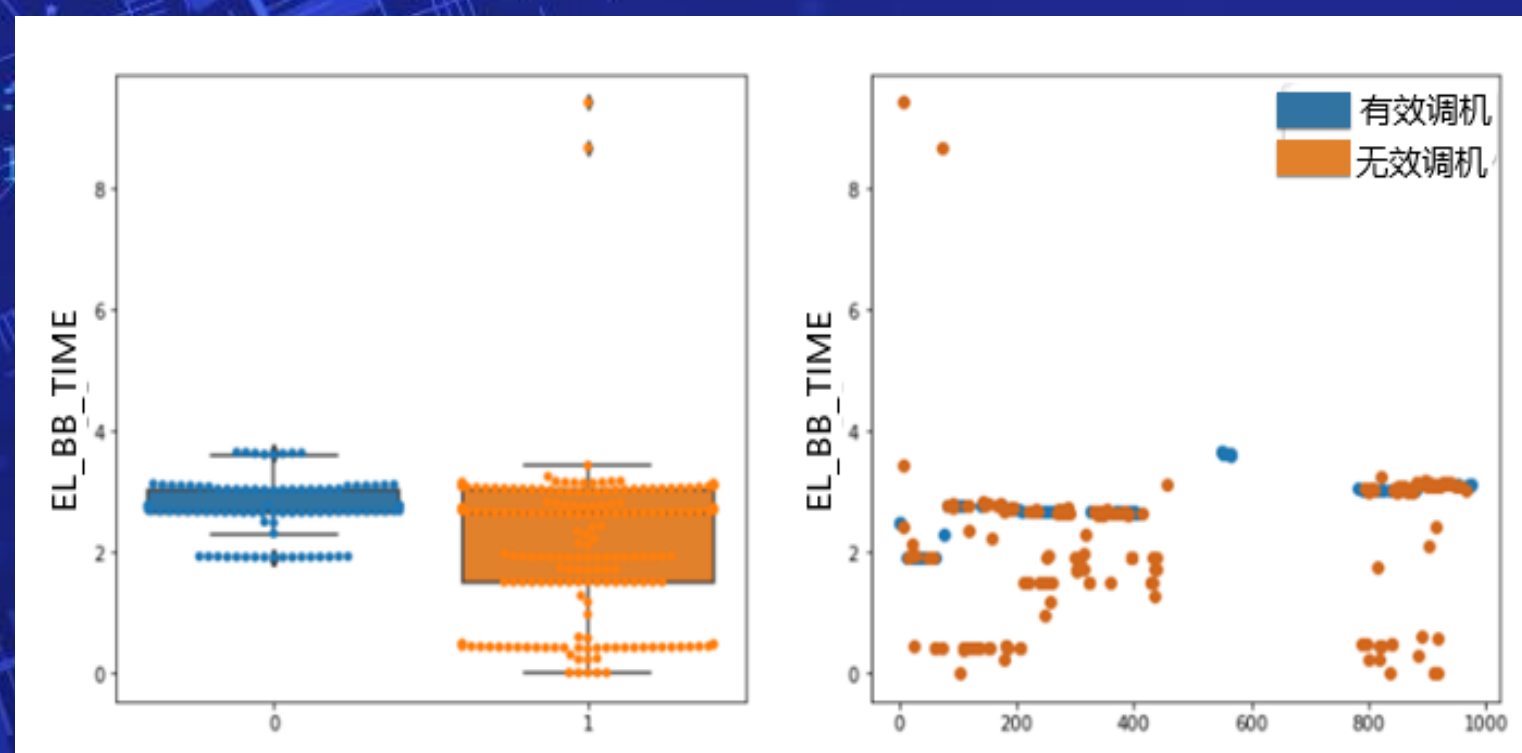


调机优化-设备运行

设备运行分析

通过参数分布差异，识别设备运行状态，将有效调机和无效调机段中参数分布差异分为3类：

1. 无效调机段参数波动明显，有效调机段参数波动较弱的特征，共11个特征
2. 无效调机段参数波动不明显，共2个特征，均来自传感器均值
3. 无效调机段与有效调机段参数中位数差异明显，共2个特征，为传感器均值

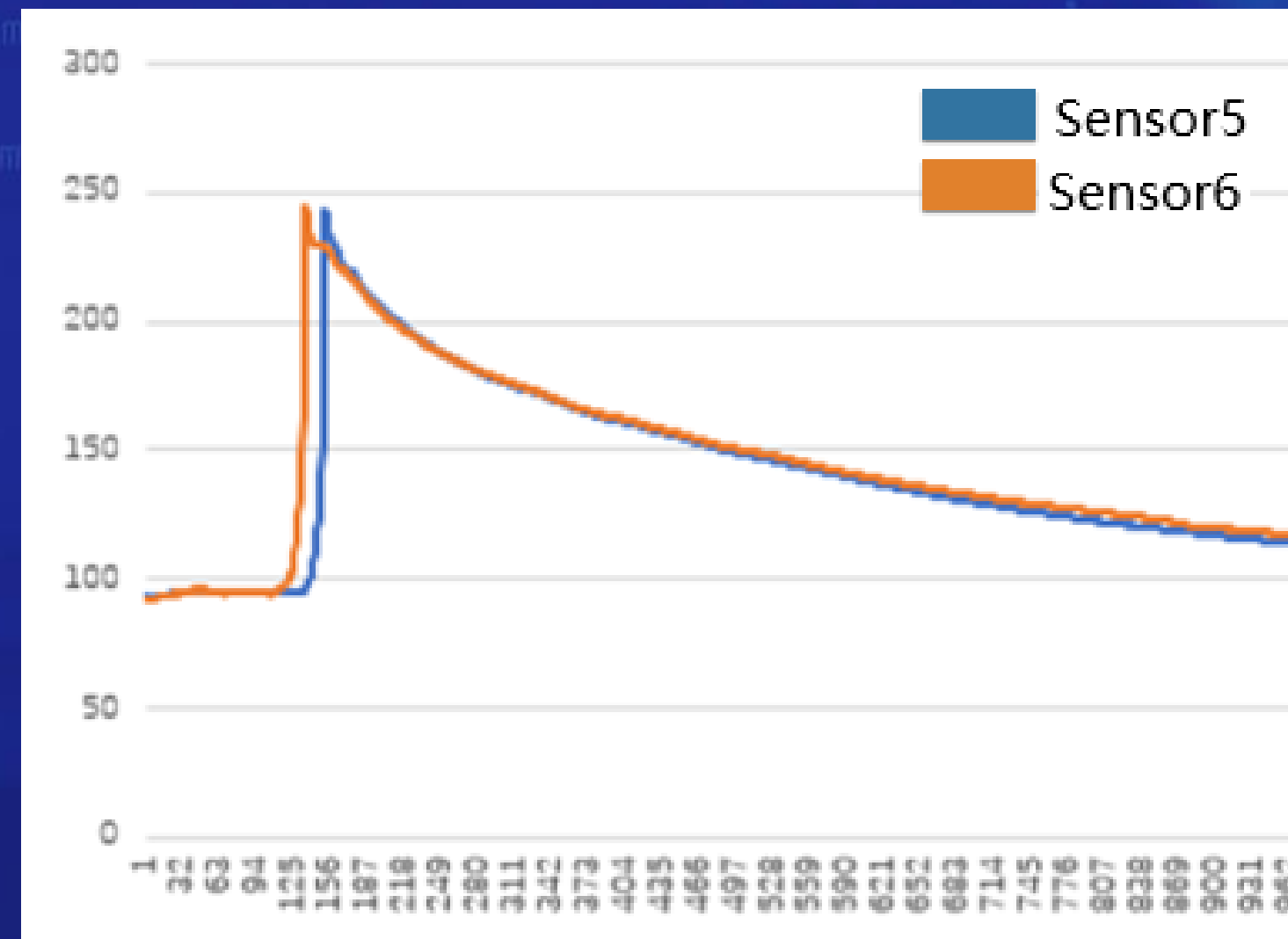
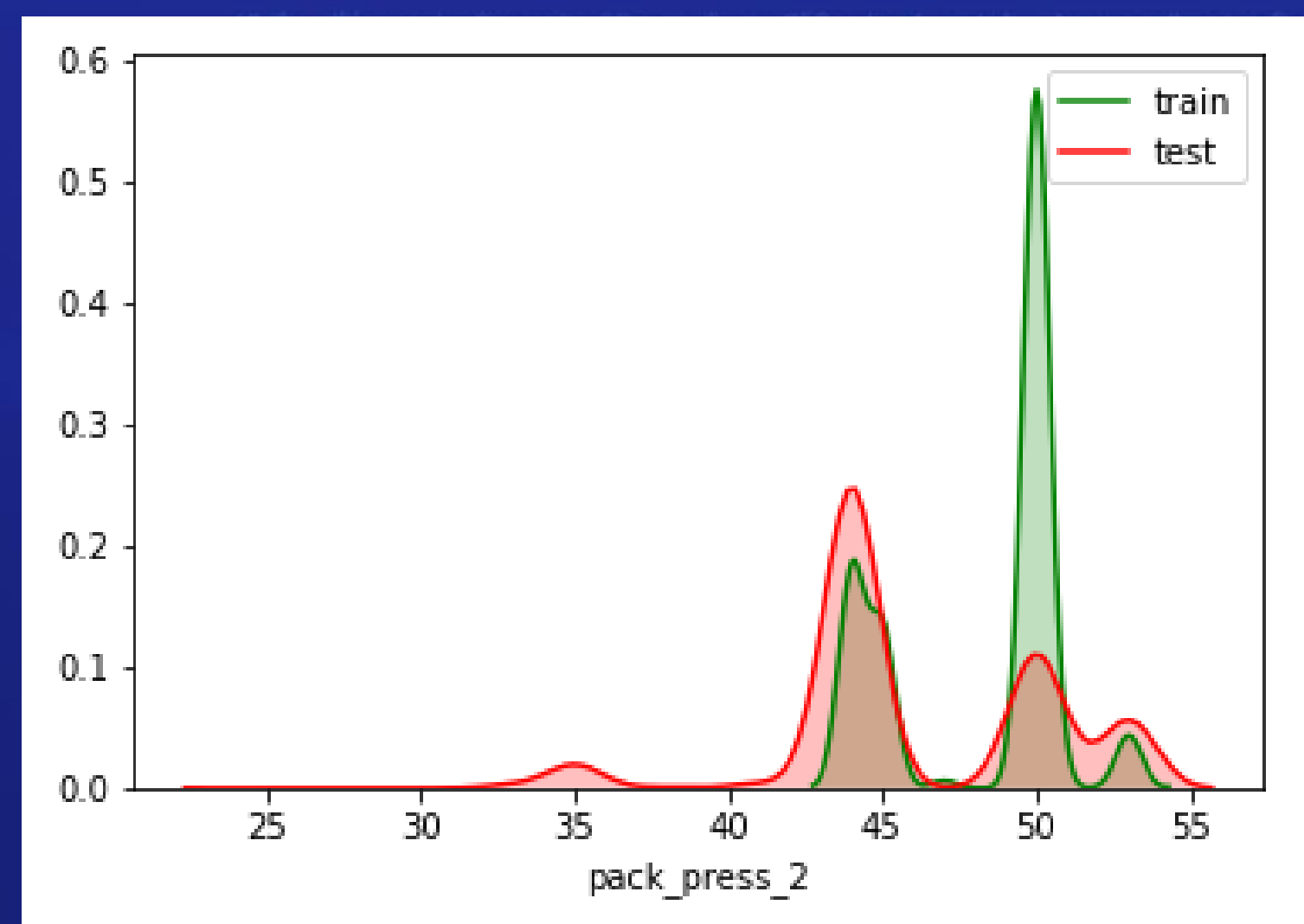
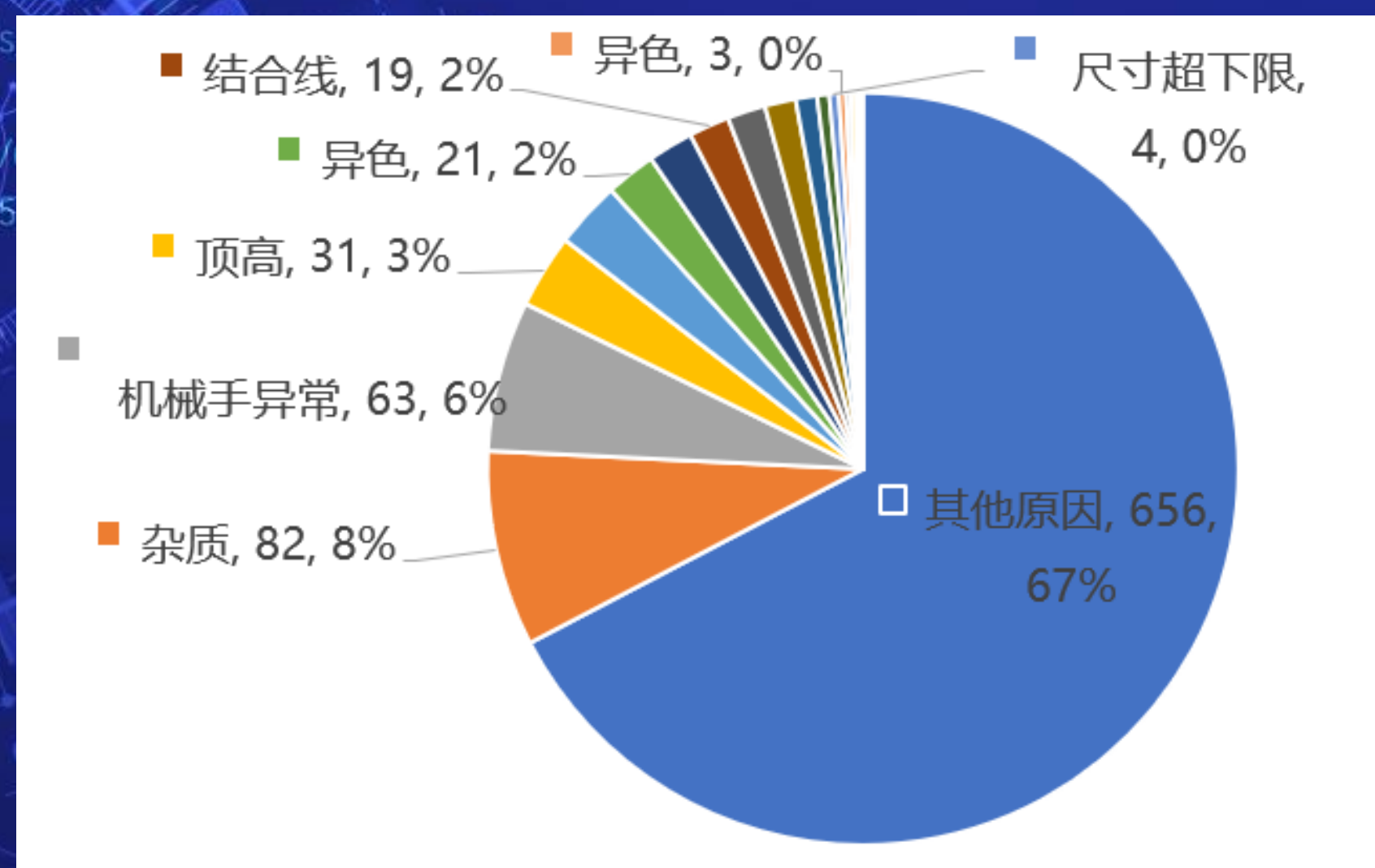


注：无效调机定义，两次调机之间生产模次在2~5之间，共207段，617模次产品数据



调机优化-数据质量

- **原因缺失**: Bad Quality问题, 所提供调机记录中, 67%的调机没有记录原因, 不利于对特定调机原因的操作进行评价分析
- **数据差异**: Bad Quality问题, 任务A时所提供数据, 部分重要参数在训练集和测试集中差异过大, 反而不利于模型结果预测
- **数据冗余**: Broken 数据冗余问题, 高频传感器中sensor5和sensor6差异较小, 可适当改善传感器布置





任务A 虚拟量测

任务B 调机优化

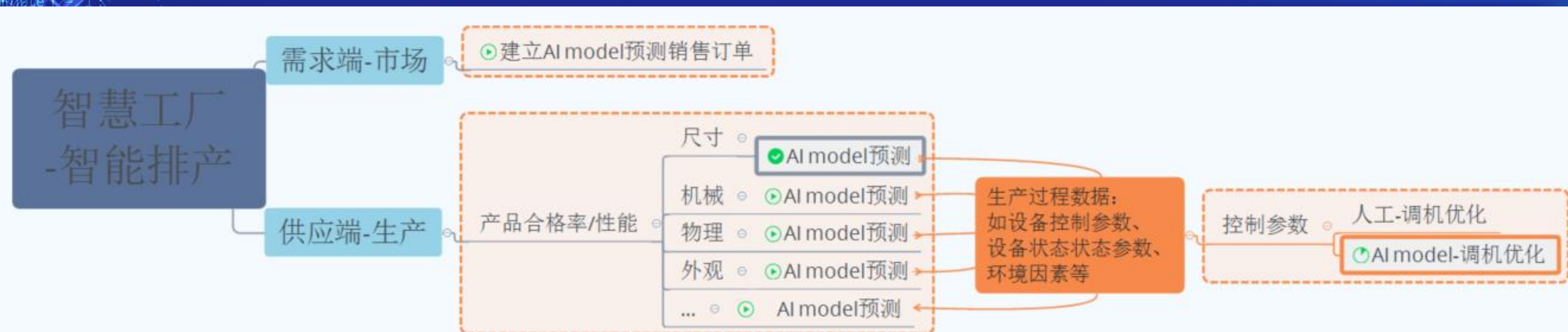
后续建议及思考



后续建议及思考

工业大数据的应用——智慧工厂

- 思考1：利用已有数据，对产品机械、外观等考核项进行全面预测，从而进行产品全面评估
- 思考2：目前调机参数依赖人工经验，进一步通过对调机数据更加全面的收集和分析，建立 AI model自动进行调机优化
- 思考3：对市场需求端的订单建立预测模型，获取销售订单





THANKS