

注塑成型工艺的虚拟量测和调机优化_第四届工业大数据竞赛题目发布

转载 kk_land 2020-09-10 21:45:46 703 收藏 2

编辑 版权

文章目录

- 1 题目描述
- 2 比赛任务
- 3 数据描述
- 4 评分规则
- 5 答案提交

REF: 官网

1 题目描述

注塑成型作为做常见的一种塑料制品加工工艺，它所加工的产品在生活中随处可见，例如电子产品、汽车配件、玩具以及其他众多消费品。由于成型系统较为复杂并且对环境较为敏感，注塑成型加工过程中的不稳定因素很容易导致产品不良的发生，造成经济损失。所以我们建立注塑成型大数据，来感知这些不可见的干扰因素，然后通过分析建模解决甚至避免现场痛点问题。比如成型过程的异常检测预警及不良品的识别，有助于减少甚至避免不合格品的产生，对于管控产品质量、降低生产成本有重要的作用。同时，针对异常产生现场人员因经验差异导致调机无法规范化的问题，如果能够根据成型过程数据和异常事件进行建模分析，改进调机策略，将会节省大量的时间成本和经济成本。

提示：为方便选手理解注塑成型工艺过程，在数据下载中提供了工艺简介，供选手参考。

2 比赛任务

本赛题需要选手完成以下两项任务：

- 任务A – 虚拟量测（初赛及决赛）
要求选手针对成型工艺品质异常中尺寸超规问题进行虚拟量测。根据训练集所提供所有模次产品的过程数据和相对应的实际量测值（标签）进行虚拟量测模型建模，然后对测试集中的产品进行尺寸预测，即虚拟量测。
- 任务B – 调机优化（仅进入决赛的选手参加）
此任务中会提供更多维度（品质异常记录，调机记录）和更长时间跨度的数据。

要求选手适当结合任务A中所得到的知识对所有数据，包括预赛数据和决赛数据，进行数据挖掘，并按照模板所要求的格式从以下几个方面完成一份调机优化报告：

1. 数据集提供了每个产品的生产过程数据，还提供了工艺参数的调机记录和部分原因。请结合这些数据分析现场调机操作的有效性与合理性形成调机优化策略。（有效性：工艺参数的调整是否解决了问题。合理性：是否存在更好的处理方式。）
2. 任务A数据集中每个模次都提供了尺寸测量信息，而任务B数据集中的模次大多没有精确测量的尺寸信息，只有定期巡检的数据用来作为品质数据的标签数据，那么请问我们是否能评估数据集中标签缺失时段的生产状况。请基于数据分析的证据给出判断。如果能，请分析说明具体品质状况；如果不能，请说明理由。
3. 工业大数据的质量一直是影响建模表现的重要因素，请结合任务A和任务B，至少从3个方面基于实证指出并说明，数据质量目前存在的问题以及如何进一步提升数据质量（如增加更多的传感器类型，改变采样方式，减少某些非关键数据采样等等）。
4. 请利用数据挖掘所得到的结果对注塑成型未来智能化升级提出自己的看法。鼓励结合做题过程中的感悟对智能制造以及工业4.0未来方向的思考。

任务B的补充说明可在决赛数据下载包中获得。

建议参赛选手将两项任务综合考虑。并鼓励参赛选手在尝试不同建模分析方法的同时，能够结合数据场景所对应的物理含义进行挖掘，将工业大数据建模与成型加工制造过程相结合，提高模型的准确性、可解释性和可信度，从而实现知识发现，赋能实际的加工制造过程。

3 数据描述

本次竞赛的数据集包含以下多种来源：

- a) 传感器高频数据：该数据来自于模温机及模具传感器采集的数据，文件夹内每一个模次对应一个csv文件，单个模次时长为40~43s，采样频率根据阶段有20Hz和50Hz两种，含有24个传感器采集的数据；
- b) 成型机状态数据（data_spc）：该数据来自成型机机台，均为表征成型过程中的一些状态数据，每一行对应一个模次，数据维度为86维；
- c) 机台工艺设定参数（data_set）：文件夹中含有注塑成型的81种工艺设定参数； d) 产品测量尺寸(size)：文件夹内含有每个模次产品的3维尺寸； e) 现场巡检数据（初赛不提供此类数据）；
- f) 电机停机记录（初赛不提供此类数据）；
- g) 炮筒温度记录（初赛不提供此类数据）。

各数据字段含义可在数据下载包中获得。

数据概况：

数据来源	任务 A 虚拟量测（单位：模）				任务 B 开放题
		训练集	初赛测试集	决赛测试集	
a 传感器高频数据	✓	16600	3953	3971	✓
b 成型机状态数据 data_spc	✓	16600	3953	3971	✓
c 机台工艺设定参数 data_set	✓	295			✓
d 产品测量尺寸 Size	✓	16600	---	---	---
e 现场巡检数据	---	---	---	---	✓
f 电机停机记录	---	---	---	---	✓
g 炮筒温度记录	---	---	---	---	✓

4 评分规则

评分规则会综合考虑任务A、任务B报告以及答辩表现。选手最后的总分会按照以下比例综合打分：

	任务 A：虚拟量测	任务 B：报告	答辩表现
占比	50%	30%	20%

任务A：（初赛、决赛A题）

根据加工要求，实际尺寸在一定范围内为合格尺寸，所给出数据的真实尺寸 st_{ij} 中包含超规尺寸（三个尺寸上下限见下表）。对于预测尺寸 sp_{ij} 是否存在超规问题的漏报及误报现象，评分函数都会有相应的惩罚。

其中真实尺寸超规，预测尺寸未超规为漏报现象；真实尺寸正常，预测尺寸超规为误报现象。

尺寸	合格范围上限 UL（Upper Limit）	合格范围下限 LL（Lower Limit）
size1	300.15	299.85
size2	200.075	199.925
size3	200.075	199.925

任务A得分为每个模次三个维度尺寸加权，n为测试集内含有模次个数，分数越低，则预测结果越准确。（分数显示为负数时，请忽略前面的负号。当分数过大，超过INT32位时，将显示“BR”标识以示超限Beyond Range。分数详情中以科学计数法方式呈现分数。）

https://blog.csdn.net/weixin_43531163

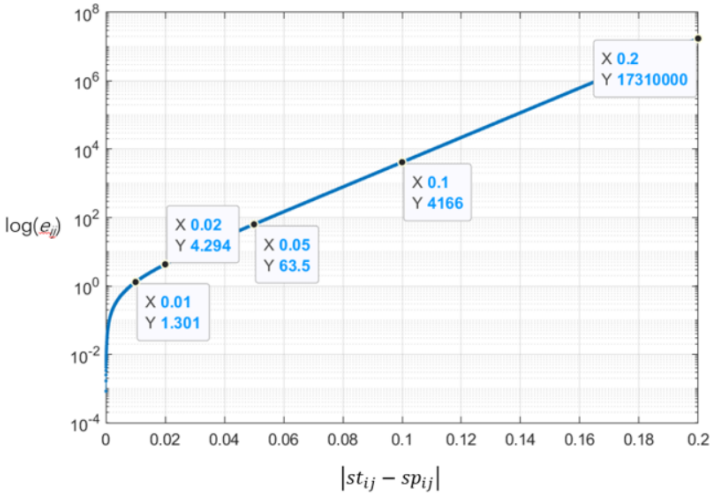
$$Score = \frac{1}{n} \sum_j \sum_i^3 (\alpha_{ij} \times \omega_{ij} \times e_{ij})$$
$$e_{ij} = e^{\frac{|st_{ij} - sp_{ij}|}{0.012}} - 1$$
$$\omega_{ij} = \begin{cases} 10 & (LL \leq st_{ij} \leq UL) \& (sp_{ij} > UL \text{ or } sp_{ij} < LL) \\ 10 & (st_{ij} > UL \& sp_{ij} \leq UL) \text{ or } (st_{ij} < LL \& sp_{ij} \geq LL) \\ 1 & \text{其余情况} \end{cases}$$
$$\alpha_{ij} = 100 \times b_{ij} + 1$$
$$b_{ij} = \begin{cases} |st_{ij} - LL| & st_{ij} < LL \\ |st_{ij} - UL| & st_{ij} > UL \\ 0 & LL \leq st_{ij} \leq UL \end{cases}$$

https://blog.csdn.net/weixin_43531163

其中：e为自然底数。

e_{ij} 为预测误差惩罚分数：误差惩罚分数和预测值与真实值之差的绝对值呈指数关系。从现场实际情况出发，我们允许模型在尺寸的预测上存在一定的误差，但是更重要的是即时精准地反映整体尺寸趋势的变化。尺寸误差 $|st_{ij} - sp_{ij}|$ 在0.01范围内，惩罚分数很小；但当尺寸误差超过0.01之后，惩罚分数将以指数速度增加。

e_{ij} 的函数图像为：



https://blog.csdn.net/weixin_43531163

w_{ij} 为误报漏报系数：该系数为每一模次根据是否存在超规的漏判及误判的情况，所对应的评分函数权重。以尺寸上下限为界限，模型应当具备准备发出报警促进现场生产的能力，而存在过多的误报或者漏报均会降低生产效率，所以评分函数会对误报/漏报的情况加以一定的惩罚。

α_{ij} 为样本异常系数：我们希望确保模型对于现场异常状况（尺寸超规）的预测精度比正常尺寸更加准确，所以评分函数会根据异常样本超规的严重程度赋予相应的惩罚权重。

任务B:

报告的评分细则:

	占比	评审内容
1. 调机操作评估	30%	基于数据分析给出证据 结合生产背景讨论 有效性说明 合理性说明
2. 生产状况评估	30%	结论 基于数据分析给出证据 结合生产背景讨论
3. 数据质量评估	30%	3 方面各 10 分 • 描述的数据质量是否合理 • 基于数据分析给出证据
4. 对未来的看法	10%	合理并符合实际情况 有前瞻性

https://blog.csdn.net/weixin_43531163

5 答案提交

参赛选手需要提交CSV格式预测结果文件，文件模板可在下载的数据集中获得。

参赛选手将文件名称命名为“sub_file.csv”进行提交，结果文件示例如下：

Id	size1	size2	size3
1	sp11	sp12	sp13
2	sp21	sp22	sp23
3	sp31	sp32	sp33
.....
n	sp41	sp42	sp43

https://blog.csdn.net/weixin_43531163

1. 初赛阶段，选手每周日全天（0时至24时）可提交3次答案。
2. 决赛阶段，选手在决赛的每个比赛日均可以提交2次答案。