

2018 智能制造邀请赛-工业大数据赛题描述

本题目的主题是半导体刻蚀机的异常检测。

【刻蚀背景介绍】

半导体制造素来以复杂、更迭快速而著称。随着传感器与自动化技术的进步与普及，半导体制造已经变成了名副其实的大数据环境。获得更多数据的好处是，复杂的制造过程透明化增强，使利用数据驱动的方法对制造过程优劣进行预测成为可能。

本题目所关注的是半导体制造过程中晶圆制造的关键环节之一，刻蚀过程。本赛题涉及的刻蚀过程是金属铝的刻蚀，所采用的机器是 LAM 9600 等离子刻蚀机。该刻蚀过程的目标是利用 BCl_3/Cl_2 电感耦合等离子体刻蚀 $\text{TiN}/\text{Al}-0.5\% \text{ Cu}/\text{TiN}/\text{氧化层}$ 。此过程主要分为六个步骤：前两个步骤是气体流量与压力的稳定；第三步是等离子点火过程；第四步是主要的金属铝层刻蚀；第五步针对 TiN 与氧化层进行刻蚀；第六步是排气过程^[1]。

为了研究刻蚀过程数据驱动的异常检测方法，在实验性地刻蚀晶圆的过程中，引入了多种不同的故障，包括气体流量、卡盘压力等的人为改变等，一共 20 类。本题目的是，用机器健康状态运行的数据做训练，然后在健康与故障数据混合的测试数据集上做测试，判断生产某个晶圆对应的过程是否为异常。



图 1. 刻蚀机样例图（资料来源：<http://www.photonicmicrodevices.com/images/DSC01953.JPG>）

【数据描述】

本次题目所提供的数据中包含 19 个采自传感器的变量，1 个时间变量，2 个流程性变量，

共 22 个变量。每个变量为一列，构成了训练与测试数据。每个变量的具体名称及其对应的含义如下表所示。

表格 1. 数据表头描述

序号	列名	数据含义
1	Run ID	晶圆流水号
2	Time	时间戳
3	Step Number	制造工艺步骤
4	BCI3 Flow	过程物理量
5	Cl2 Flow	
6	RF Btm Pwr	
7	RF Btm Rfl Pwr	
8	Endpt A	
9	He Press	
10	Pressure	
11	RF Tuner	
12	RF Load	
13	RF Phase Err	
14	RF Pwr	
15	RF Impedance	
16	TCP Tuner	
17	TCP Phase Err	
18	TCP Impedance	
19	TCP Top Pwr	
20	TCP Rfl Pwr	
21	TCP Load	
22	Vat Valve	

注：同样的晶圆流水号对应一个晶圆，且每个晶圆的晶圆流水号是唯一的。

【问题陈述】

正如在题目背景中介绍的，刻蚀机异常检测的目标是只利用健康状态的数据训练模型，然后利用该模型在含有故障的数据上做判断，判断生产哪一片晶圆时机器的状态是健康的，哪一片是故障的：

1. 训练数据集：etch_train.csv
2. 测试数据集：etch_test.csv

题目要求：

1. 以晶圆为单位，对每个晶圆的数据进行特征提取。
2. 利用健康状态训练数据建立基线模型，将该模型用于测试数据，输出对每个晶圆的异常检测结果。

【结果提交格式】

提交文件格式：CSV 文件

文件名: 如果有参赛组序号的话,“参赛组序号_组员 1 姓名_组员 2 姓名_组员 3 姓名_刻蚀机异常检测”;如果没有参赛组序号的话,“组员 1 姓名_组员 2 姓名_组员 3 姓名_刻蚀机异常检测”。

内容: 一列,表头为“Label”,数据取值为“0”或者“1”,每一行对应预测某个晶圆加工过程是否存在故障的标签,“0”代表无故障,“1”代表有故障。预测标签对应的晶圆流水号的排列顺序应该与测试数据“etch_test.csv”文件中的“Run ID”顺序保持一致。

【评分标准】

总分评判以及排名按照以下规定执行:

1. 根据选手提供的预测结果和真实结果计算 F1 score 值,作为评分结果。F1 score 的计算方法参见: https://en.wikipedia.org/wiki/Precision_and_recall#F-measure。该项总分为 100 分。排名前列的小组进入第二天的答辩环节。
2. 基于给定的 PPT 模版提交评审答辩材料,包括题目与背景理解、分析思路总览,和总结(建模结果、模型优劣评价、收获等)三部分内容。此项为加分项,总分 20 分。该项主要考察整体思路的逻辑严密性,以及分析方法的有效性和创新性,最终经过评审专家问答给出得分结果。

【参考资料】

[1] B. M. Wise, N. B. Gallagher, S. W. Butler, D. White, and G. G. Barna, “Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: impact of measurement selection and data treatment on sensitivity,” in *IFAC SAFEPROCESS*, 1997, vol. 97, pp. 35–42.