MATLAB EXPO

面向时间序列的异常检测

袁航，MathWorks

MathWorks

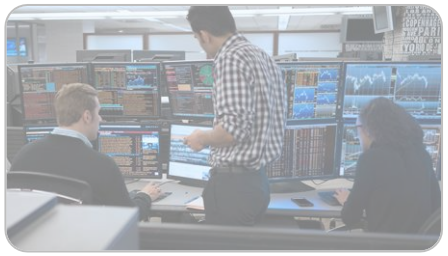
时间序列异常检测应用场景

医疗设备和数字健康

- **需求：** 床旁患者监测系统，可穿戴健康管理设备
- **挑战：** 临床数据质量、因果关系复杂

工业设备运行状态监控

- **需求：** 全生命周期预测性维护服务、提高设备可靠性
- **挑战：** 缺乏运营数据，故障机理不明或不可测定



金融交易和网络数据

- **需求：** 金融欺诈、网络入侵检测
- **挑战：** 种类众多或不可知

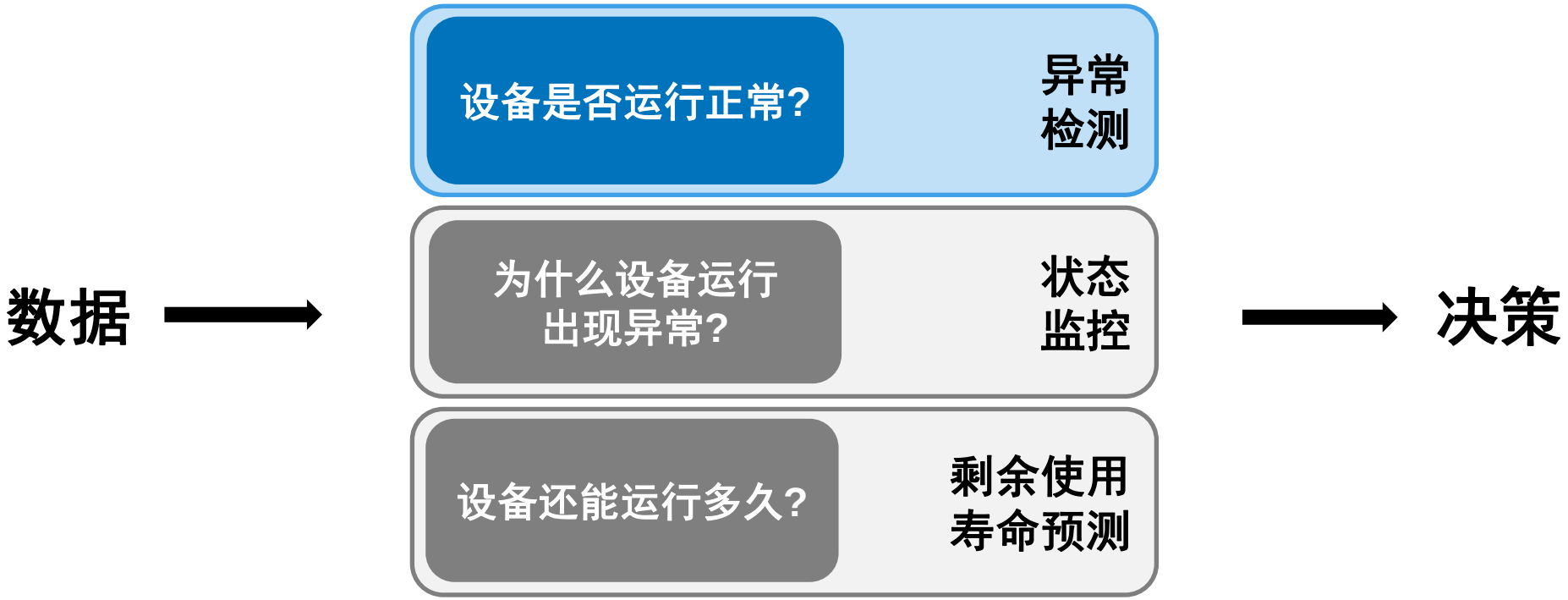
内容概要

- 什么是异常检测
- 时间序列的异常检测问题
- 异常检测算法开发流程

- 有监督、无监督和半监督异常检测方法
- 多元时间序列的异常检测

工业设备的预测性维护算法

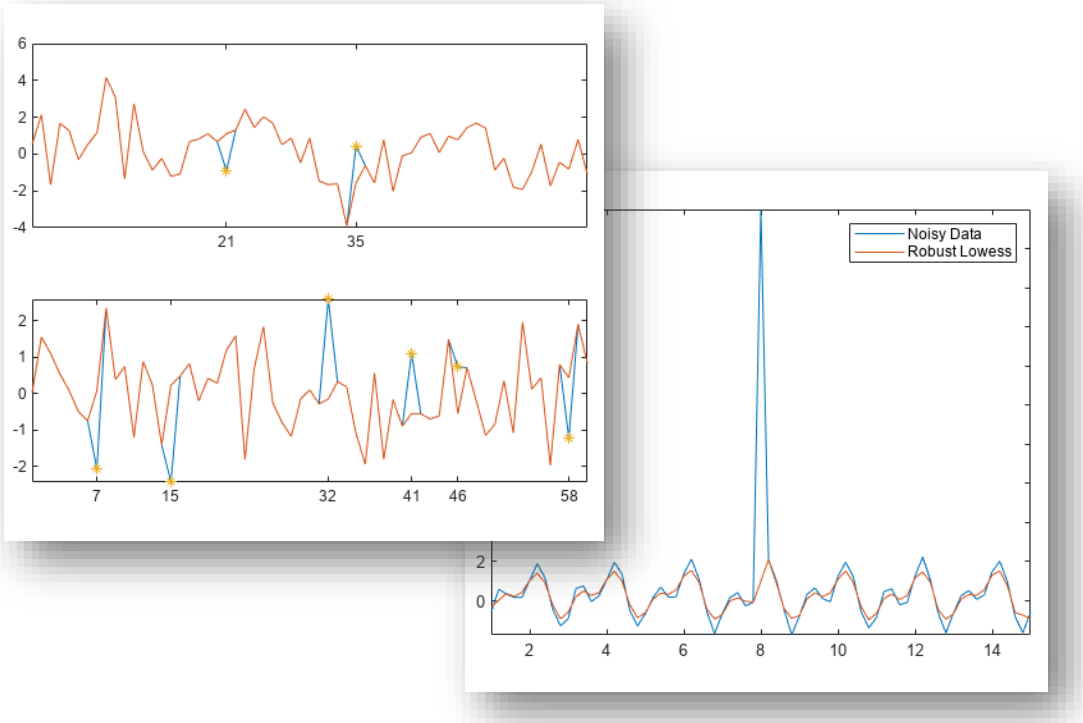
根据大量复杂数据辅助选择维护策略



异常：偏离预期的行为

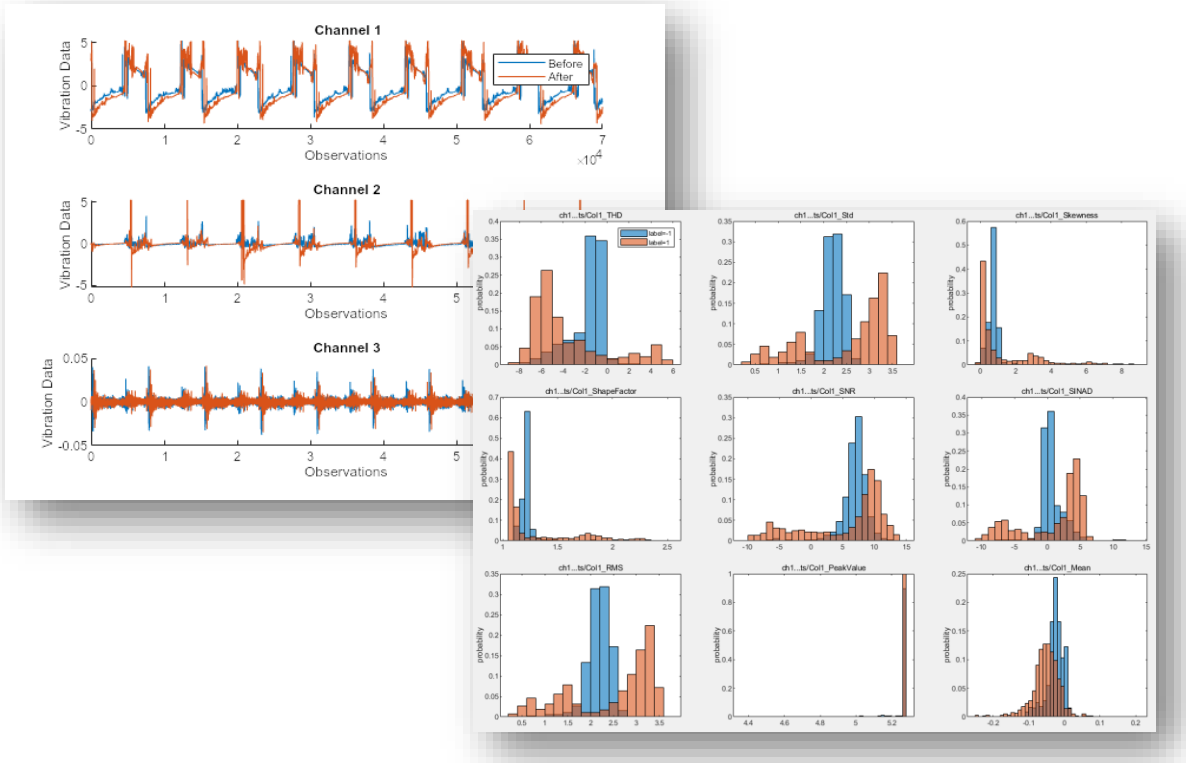
点异常 (point anomaly)

可直接通过统计分析、离群值检测判定，但易于和噪声干扰混淆



组合异常 (collective anomaly)

需要综合考虑多个特征的统计分布情况进行判定

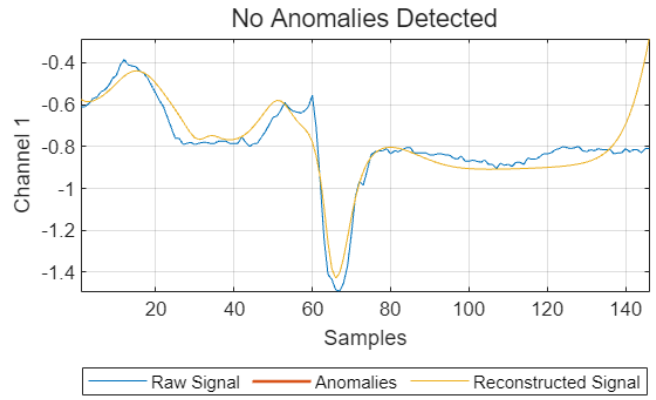


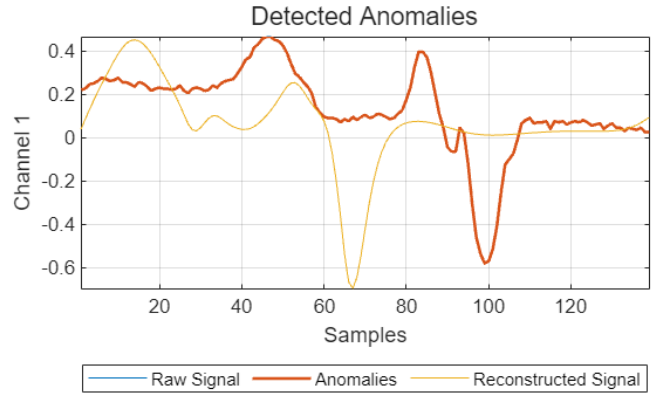
异常检测问题

- 异常检测：
  - 奇异值检测
  - 离群值检测

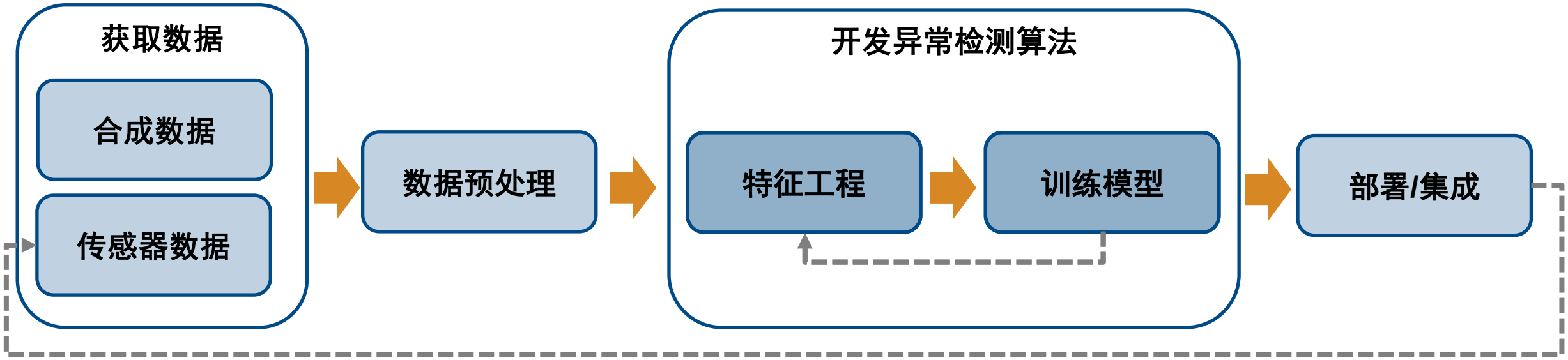
- 异常是一类模式，而不是一个离群点
- 离群点通常为噪声干扰

- 通过规则进行异常检测往往比较困难：
  - 通常，正常和异常模式无法预先定义
  - 噪声水平会影响可检出的异常
  - 领域知识和经验，对于识别异常非常重要
- 正常行为也可能伴随着老化和磨损等退化特性

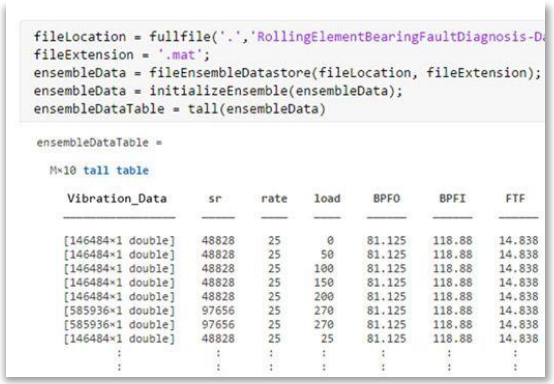




数据驱动的异常检测算法开发流程

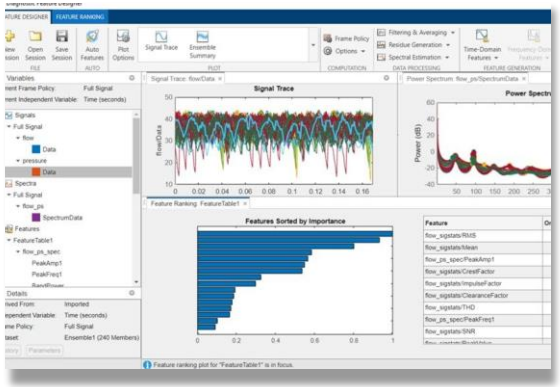


重点介绍：

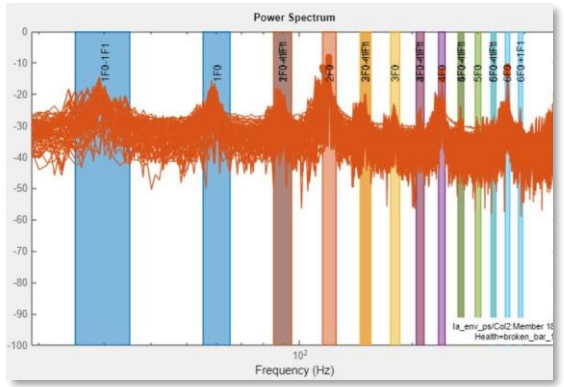


```
fileLocation = fullfile('.','RollingElementBearingFaultDiagnosis-D
fileExtension = '.mat';
ensembleData = fileEnsembleDatastore(fileLocation, fileExtension);
ensembleData = initializeEnsemble(ensembleData);
ensembleDataTable = tall(ensembleData)

ensembleDataTable =

  M×10 tall table

    Vibration_Data      sr      rate    load    BPFO      BPFI      FTF
    ______________    _____    ____    ____    _____    ______    ______

    [146484×1 double]    48828    25       0    81.125    118.88    14.838
    [146484×1 double]    48828    25      50    81.125    118.88    14.838
    [146484×1 double]    48828    25     100    81.125    118.88    14.838
    [146484×1 double]    48828    25     150    81.125    118.88    14.838
    [146484×1 double]    48828    25     200    81.125    118.88    14.838
    [585936×1 double]    97656    25     270    81.125    118.88    14.838
    [585936×1 double]    97656    25     270    81.125    118.88    14.838
    [146484×1 double]    48828    25      25    81.125    118.88    14.838
          :                :       :       :       :         :         :
          :                :       :       :       :         :         :
```

大规模数据管理



时间序列数据的特征工程



数据驱动的异常检测方法

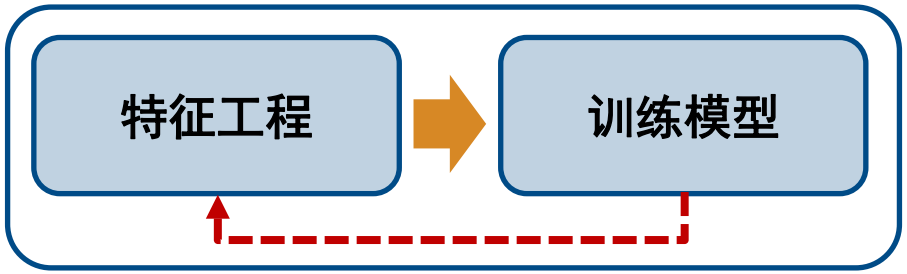
使用FileEnsembleDatastore管理大规模数据集
文件集合+单一文件

```
fensemble = fileEnsembleDatastore(location,extension);
```

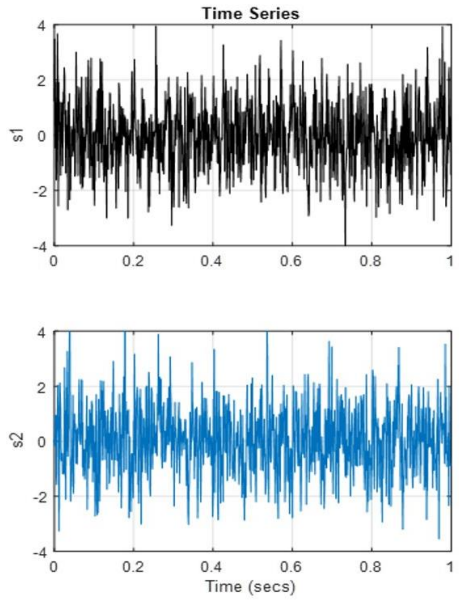
Date	Vibration	Tacho	EnvPower	SigMean	SigMedian	SigApproxEntropy	SensorDrift	ShaftWare	ToothFault	FaultCode
"13-Nov-2017 09:17:01"	[20071×1 timetable]	[40272×1 timetable]	0.00022907	-0.96896	-0.98752	0.020944	true	false	false	1
"13-Nov-2017 09:17:02"	[30132×1 timetable]	[40300×1 timetable]	9.1568e-08	-0.97537	-0.98958	0.03786	true	false	false	1
"13-Nov-2017 09:17:03"	[30118×1 timetable]	[40282×1 timetable]	3.1343e-07	1.0502	1.0267	0.031586	true	true	false	3
"13-Nov-2017 09:17:04"	[30127×1 timetable]	[40291×1 timetable]	2.5787e-07	1.0227	1.0045	0.032109	true	true	false	3
"13-Nov-2017 09:17:05"	[30142×1 timetable]	[40308×1 timetable]	2.2397e-07	1.0123	1.0024	0.032891	true	true	false	3
"13-Nov-2017 09:17:05"	[30124×1 timetable]	[40288×1 timetable]	1.9224e-07	1.0275	1.0102	0.033449	true	true	false	3
"13-Nov-2017 09:17:06"	[30160×1 timetable]	[40326×1 timetable]	1.6263e-07	1.0464	1.0275	0.034182	true	true	false	3
"13-Nov-2017 09:17:07"	[30125×1 timetable]	[40290×1 timetable]	1.2807e-07	1.0459	1.0257	0.035323	true	true	false	3
:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:

- 文件集合 (表格 ~ 所有文件)
- 成员 (行 ~ 单一文件)
- 独立变量 Independent variables
- 数据变量 Data variables
- 状态变量 Condition variables

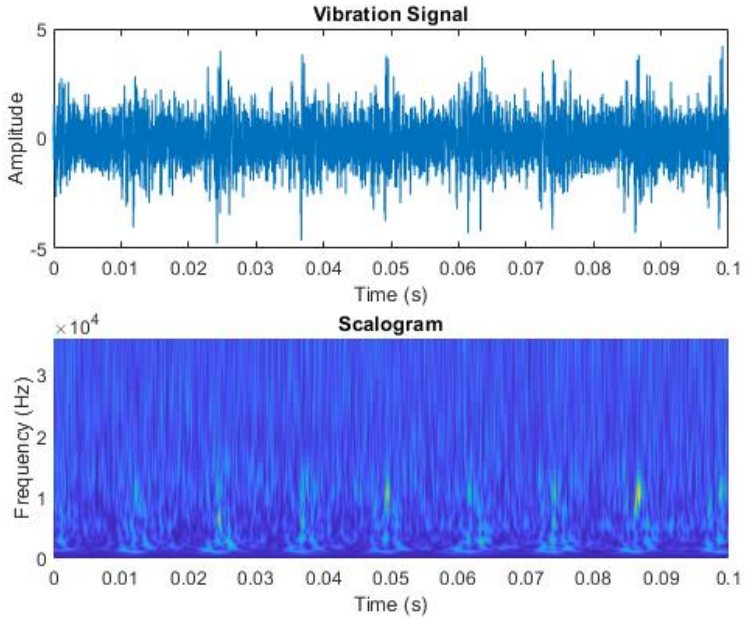
时间序列数据的特征工程



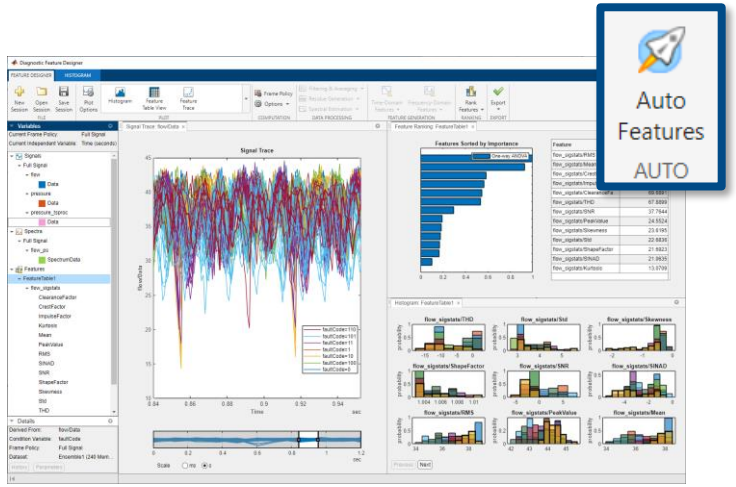
- 直接处理时序数据



- 时频变换
  将信号数据转换为图像数据

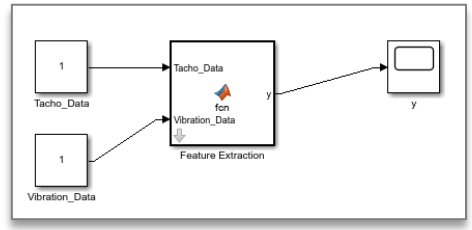


- 按时间窗口划分并提取特征值
  将信号数据转换为表格数据



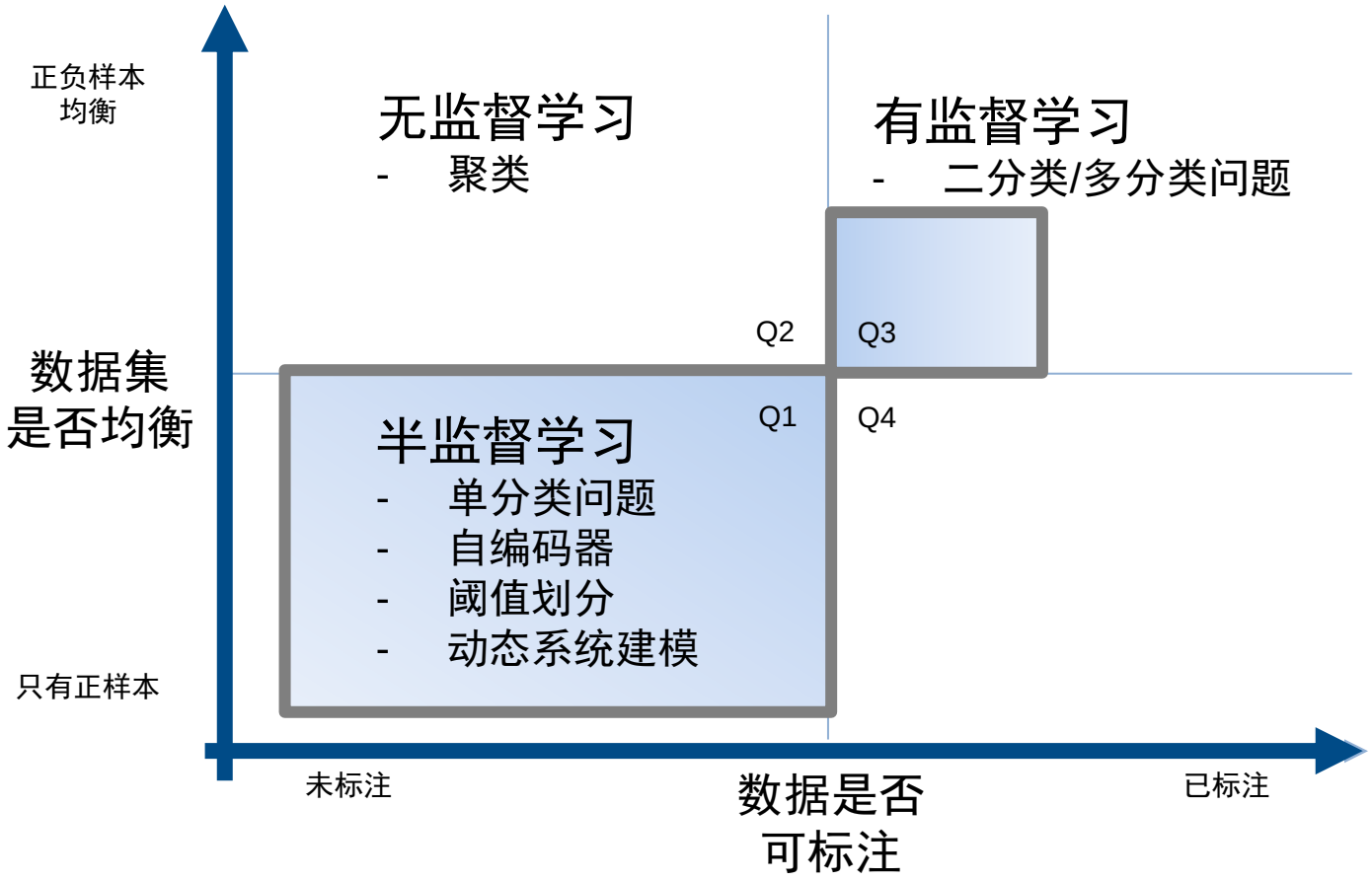
gearMetrics1=*10×9 table*

RMS	Kurtosis	CrestFactor	FM4	M6A	M8A	FM0	EnergyRatio
5.1119	2.074	2.4377	2.4633	9.0009	42.31	1.5499	0.060057
5.1272	2.087	2.4819	1.9331	4.9869	15.634	1.5785	0.10044
5.1526	2.102	2.4744	1.7084	3.6211	8.8635	1.5881	0.14423
5.1877	2.1264	2.5443					
5.2385	2.1566	2.5985					
5.2953	2.1879	2.605					
5.365	2.2277	2.6551					
5.4425	2.2574	2.6428					
5.5269	2.2891	2.7112					
5.6219	2.3214	2.6979					



异常检测算法开发

- **基于信号处理和统计分析的方法**
  变化点检测，控制图、3σ准则、箱型图、Grubbs检验
- **有监督方法**
  正常数据和异常数据样本均衡
- **无监督方法**
  聚类分析
- **半监督方法**
  仅根据正常数据，进行奇异值检测



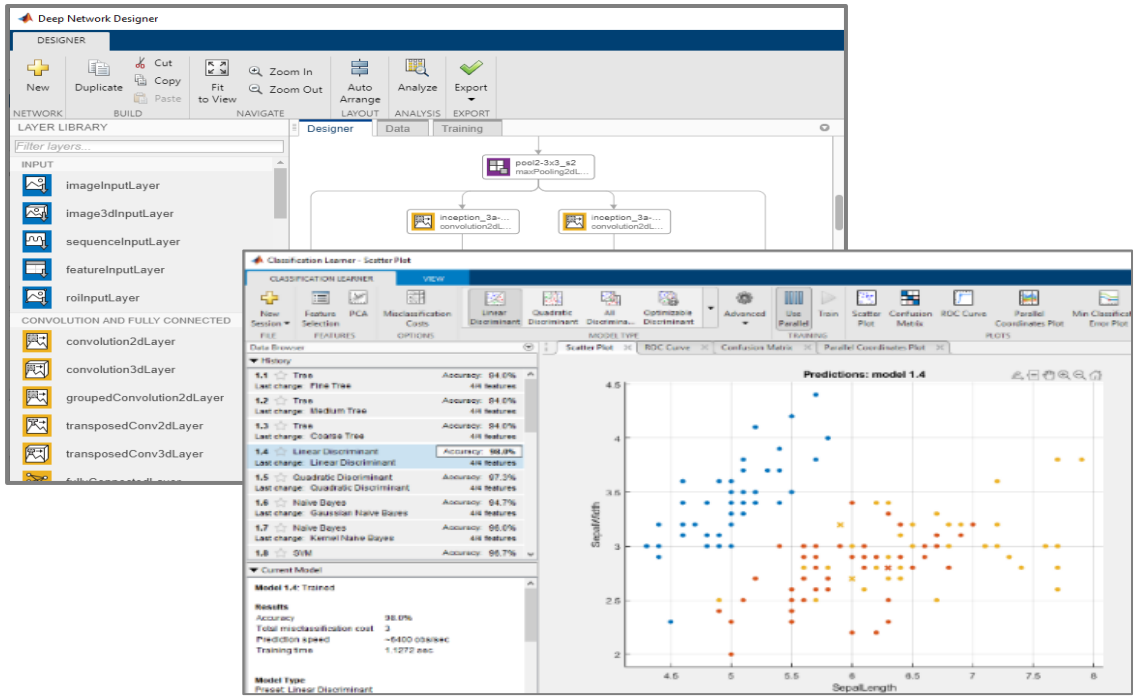
内容概要

- 什么是异常检测
- 时间序列的异常检测问题
- 异常检测算法开发流程

- 有监督、无监督和半监督异常检测方法
- 多元时间序列的异常检测

无监督学习 - 聚类	有监督学习
半监督学习 - 单分类问题 - 自编码器	

有监督异常检测方法



低代码开发 - 交互式应用程序

Classification Learner, Regression Learner, Deep Network Designer

```
imageInputLayer([2 spf 1], 'N
'Input Layer')

convolution2dLayer(filterSize
'Name', 'CNN1')

batchNormalizationLayer('Name
reluLayer('Name', 'ReLU1')
maxPooling2dLayer(poolSize, '
'MaxPool1')
```

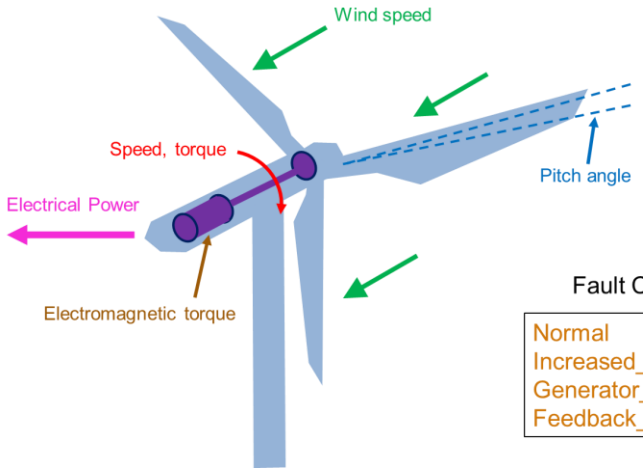
代码实现方式

数据不均衡？尝试用机理模型合成训练数据

Measurements from IoT Devices or SCADA

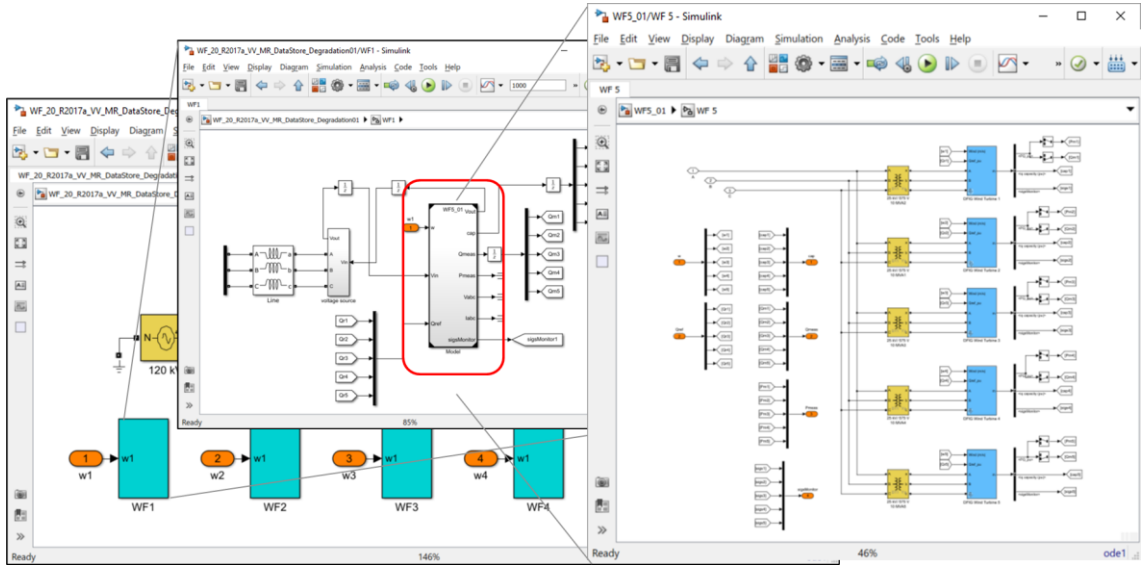
wind_speed
pitch_angle
turbine_speed
turbine_torque
shaft_speed
shaft_torque
electromagnetic_torque
P_gen
Q_gen
P_out
Q_out
V_DC
Qref

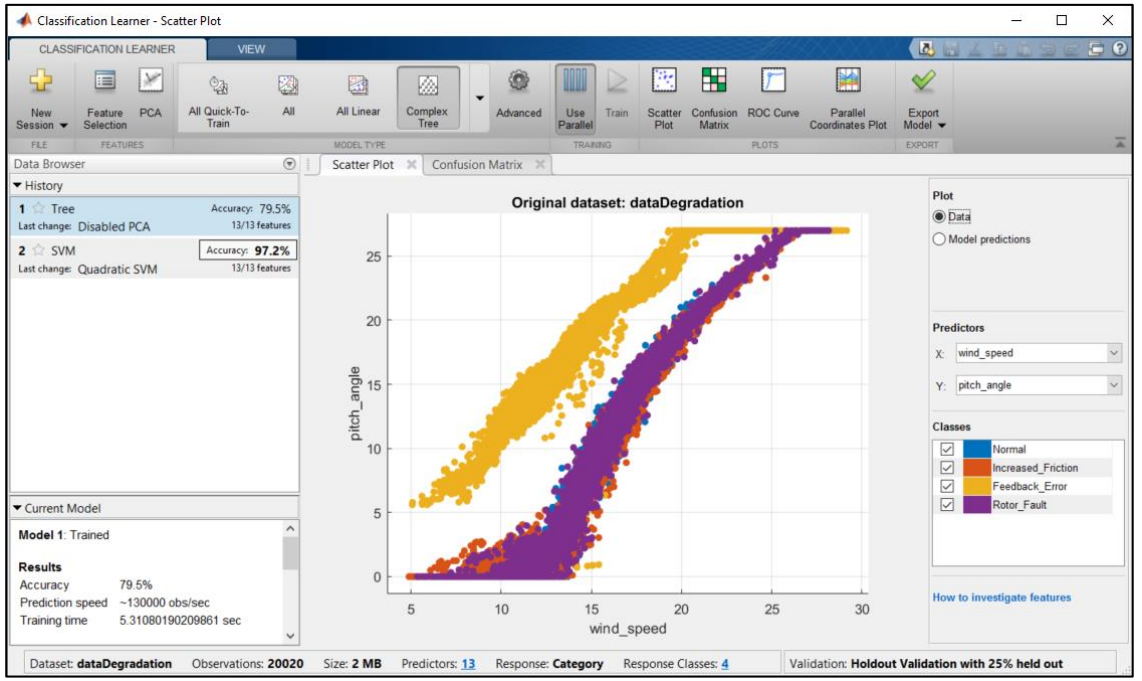
X 20

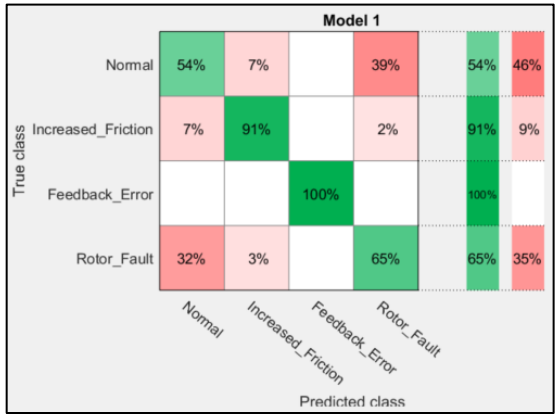


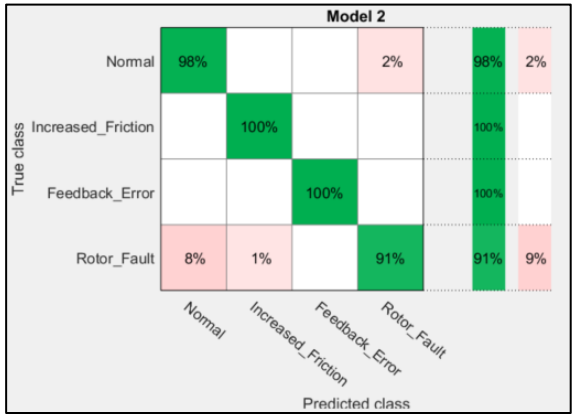
Fault Categories

Normal
Increased_Friction
Generator_Rotor_Fault
Feedback_Error

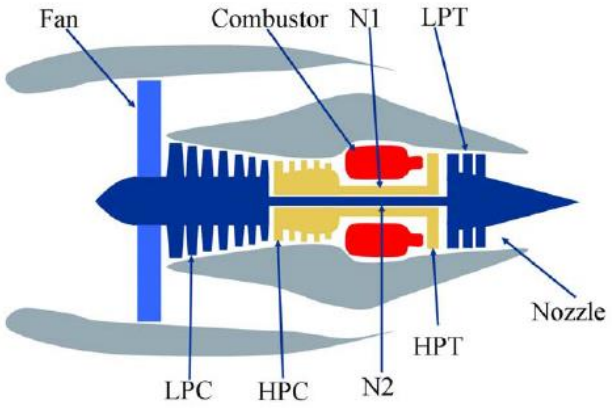




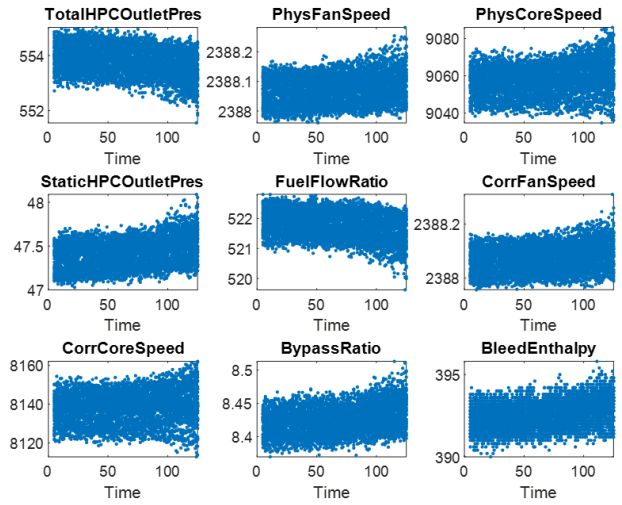




无监督异常检测方法

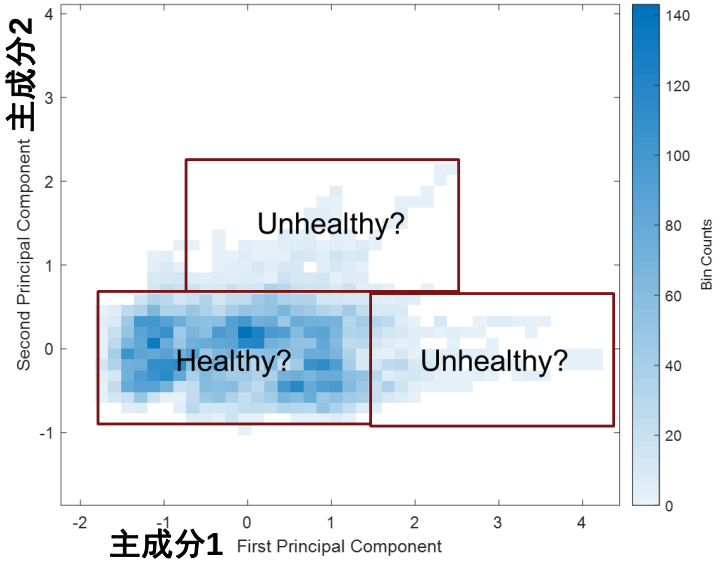


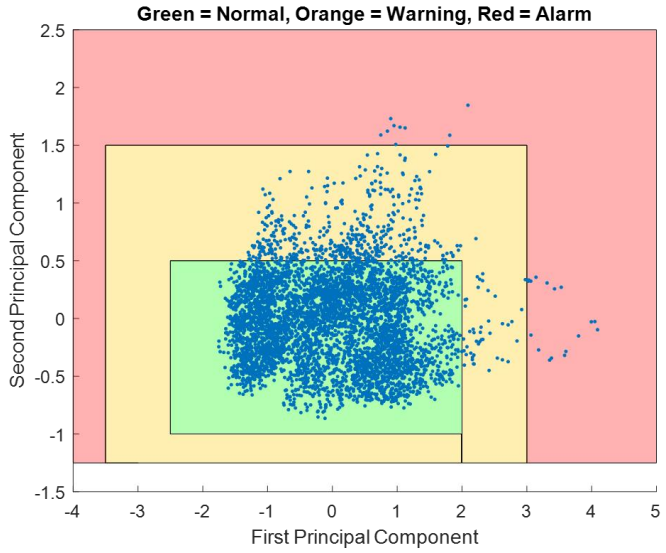
- 安全性、性能要求高，系统失效率极低
- 没有故障数据或故障模式不明确



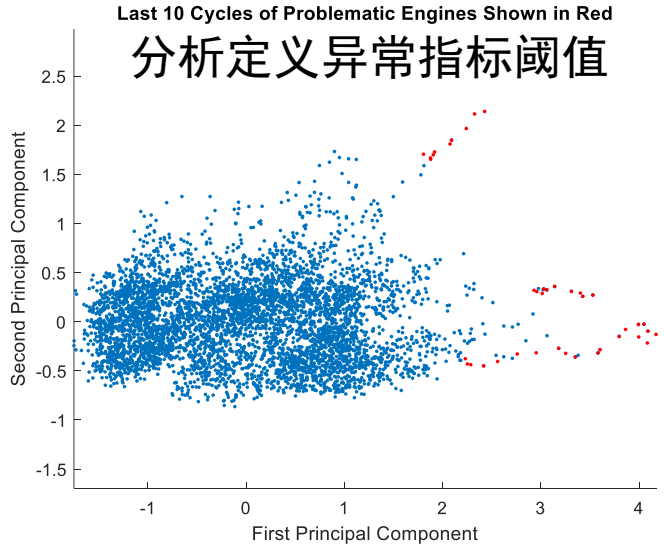
难以从9+个传感器观测信号中识别趋势

PCA
降维



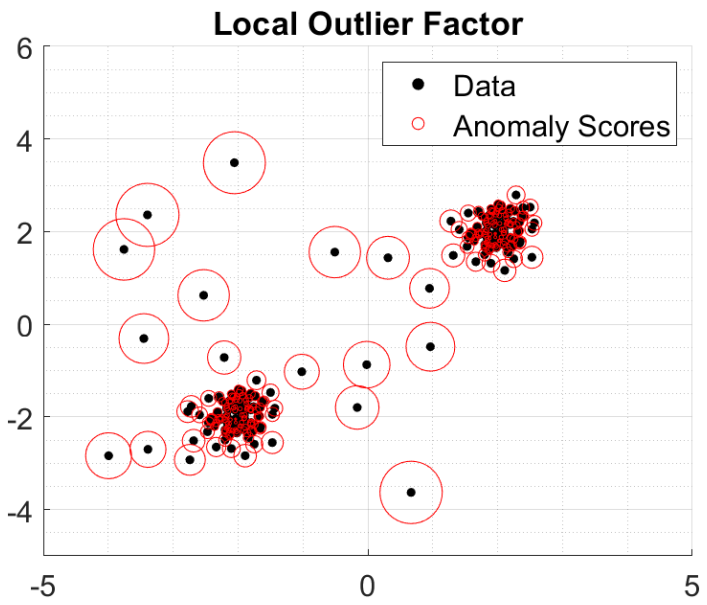


阈值划分

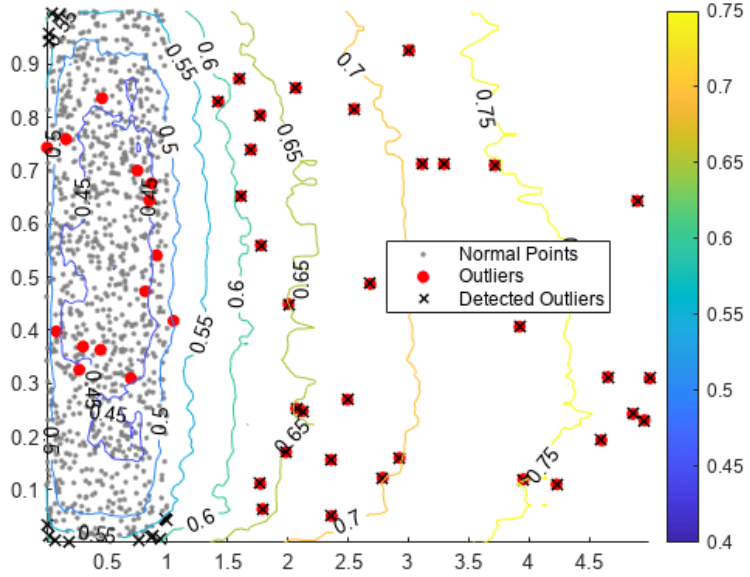


半监督方法 - 基于正常数据的奇异值检测

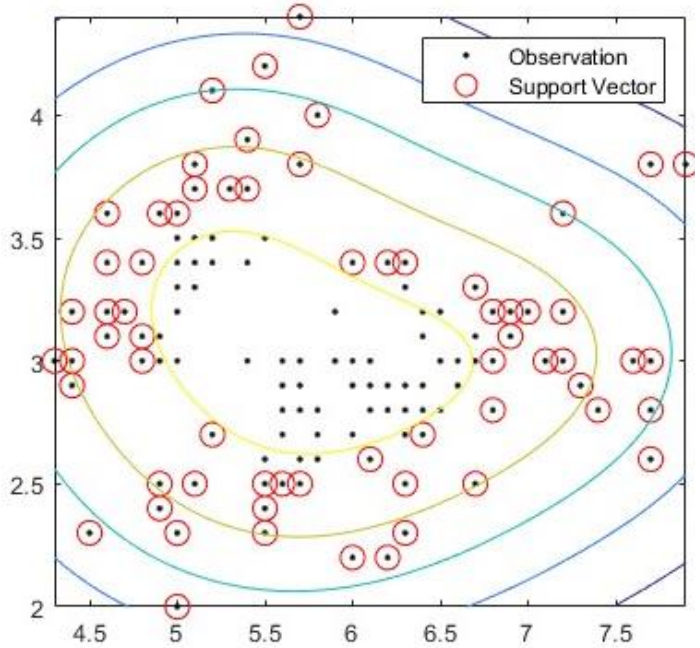
局部离群因子
Local Outlier Factor



孤立森林
Isolation Forest



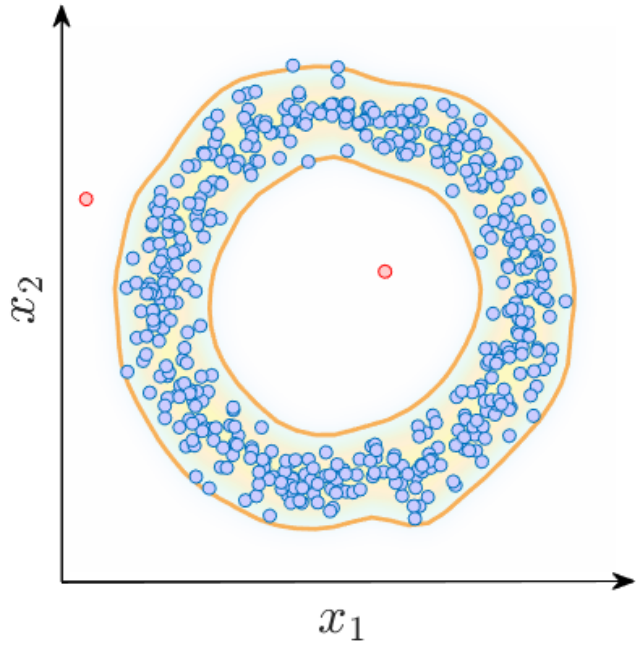
单类支持向量机
One-class SVM (ocsvm)



1. 将任务定义有监督或无监督的分类问题 – 结果不理想
2. 单分类问题 – 常用机器学习算法多用于解决二分类/多分类问题

基于正常数据的奇异值检测 – 单类支持向量机 OCSVM

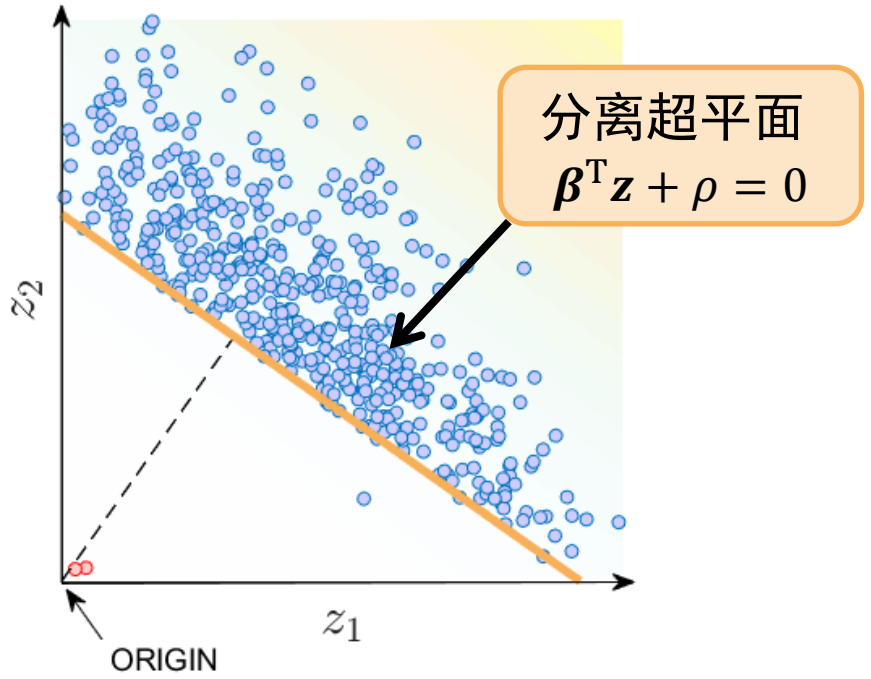
- 单类支持向量机 – 单分类器



$$\boldsymbol{\phi}: \boldsymbol{x} \rightarrow \boldsymbol{z}$$

$$\boldsymbol{z} \in \mathbb{R}^{\infty}$$

升维变换



$$\textit{Anomaly score} = \sum_{i}^{N} \alpha_i \boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x}) + \rho$$

核技巧

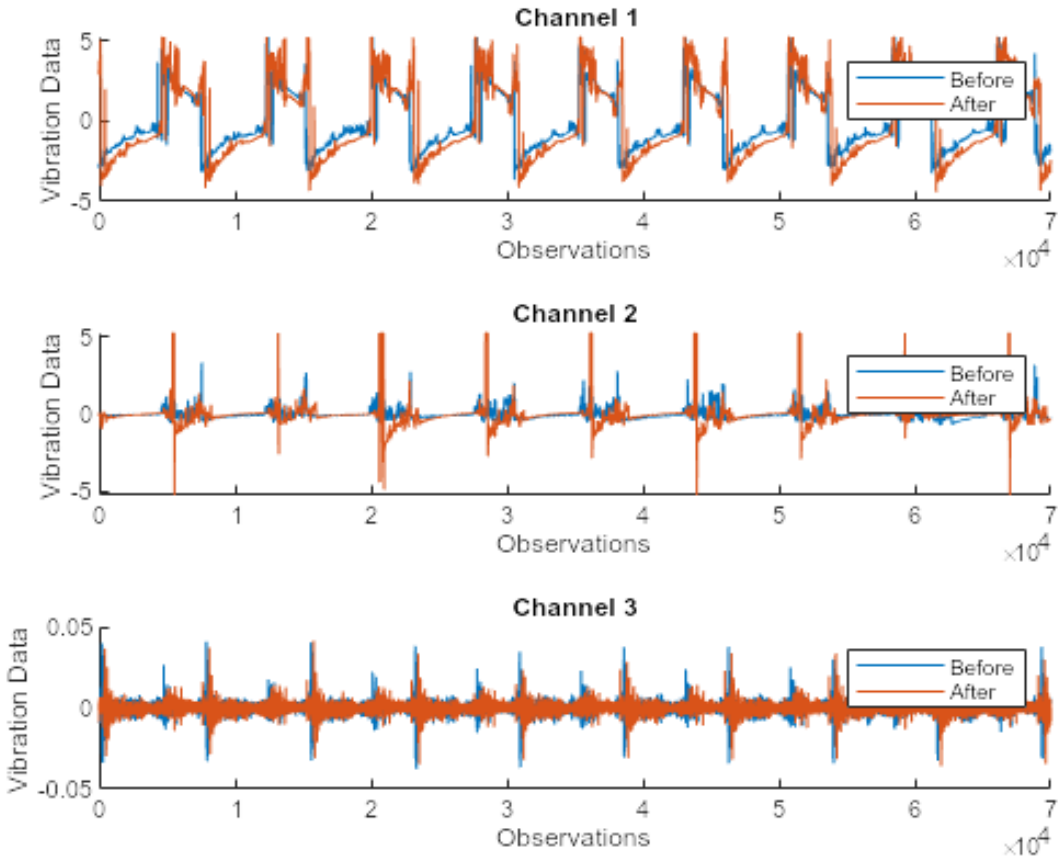
$$\textit{Anomaly score} = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{z} + \rho$$

* 奇异值检测算法输出 – 量化得分而非标签

利用自编码器检测异常

问题设定

[示例链接]



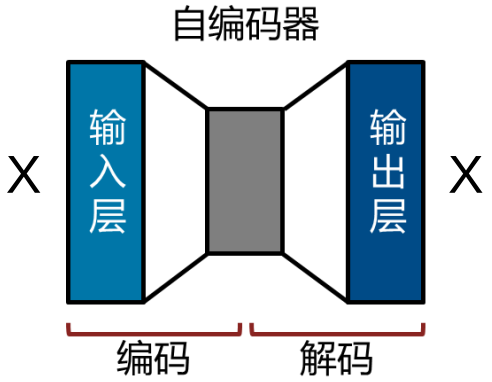
工业设备运维方式

- 每月定期维护
- 故障发生次数较少
- 故障发生时后果严重 …

- 标签信息：“维护前”、“维护后”
- 默认“**维护后**”数据为正常数据

利用自编码器检测异常

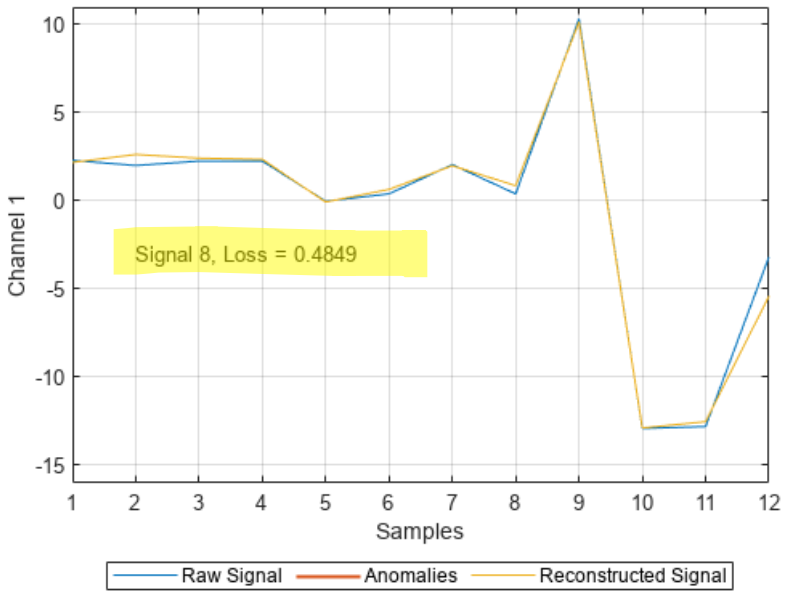
结果



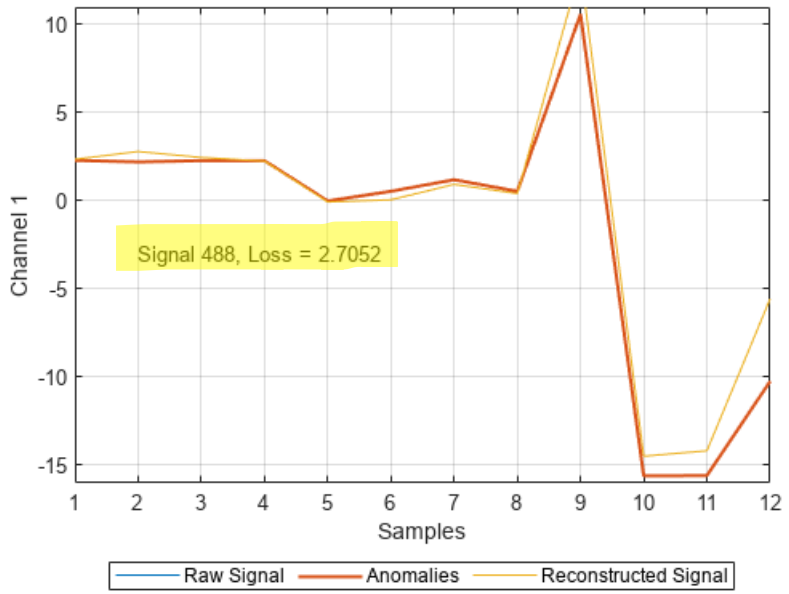
- 从信号数据中提取12个特征
- 仅使用维护后数据训练模型
- 根据各个特征重建损失判定是否异常

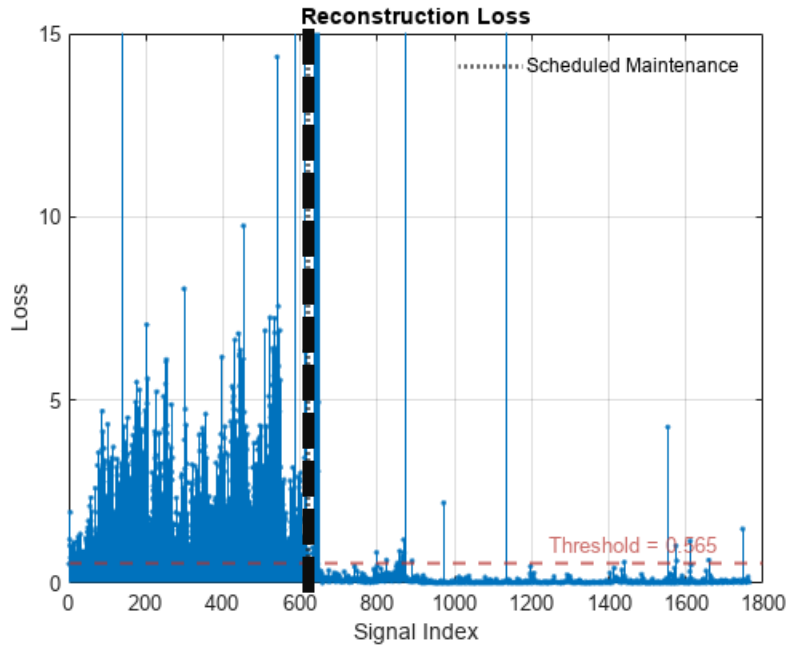
```
» detector = deepSignalAnomalyDetector(1,"lstm",...
    EncoderHiddenUnits=[16 32], ...
    DecoderHiddenUnits=16, ...
    WindowLength="fullSignal", ...
    ThresholdMethod="mean", ...
    ThresholdParameter=0.8);
```

未检测出异常



检测出异常

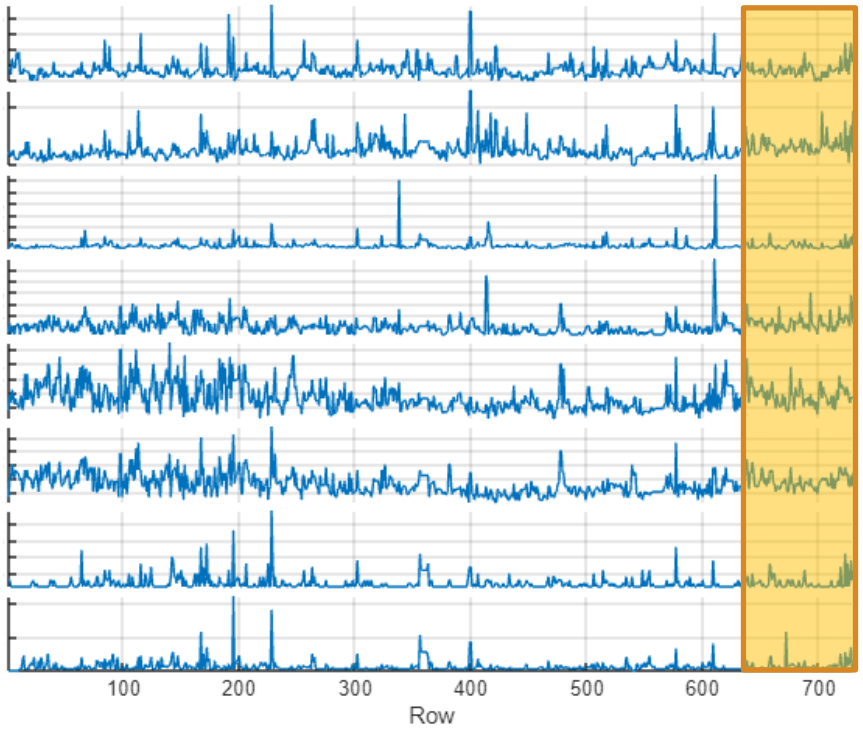




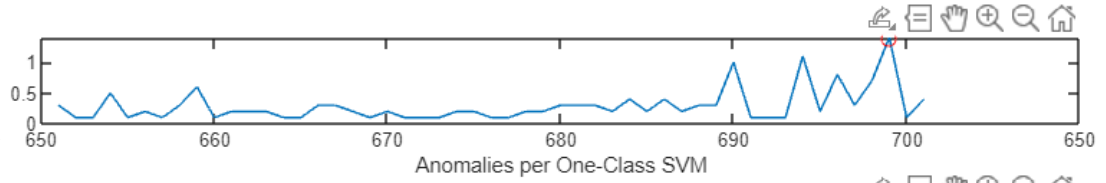
特征序号

信号序号

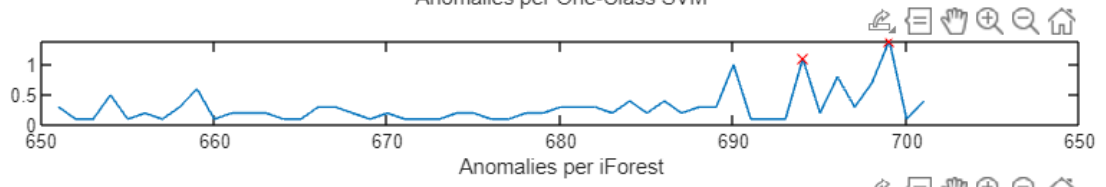
多元时间序列的异常检测问题



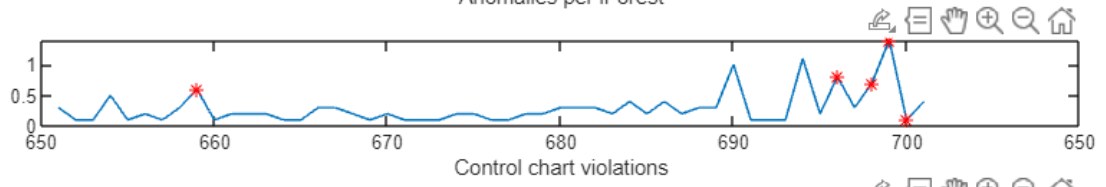
单类支持向量机



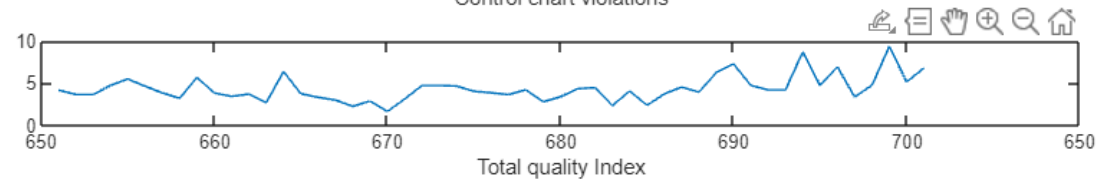
孤立森林



统计过程控制



质量指标

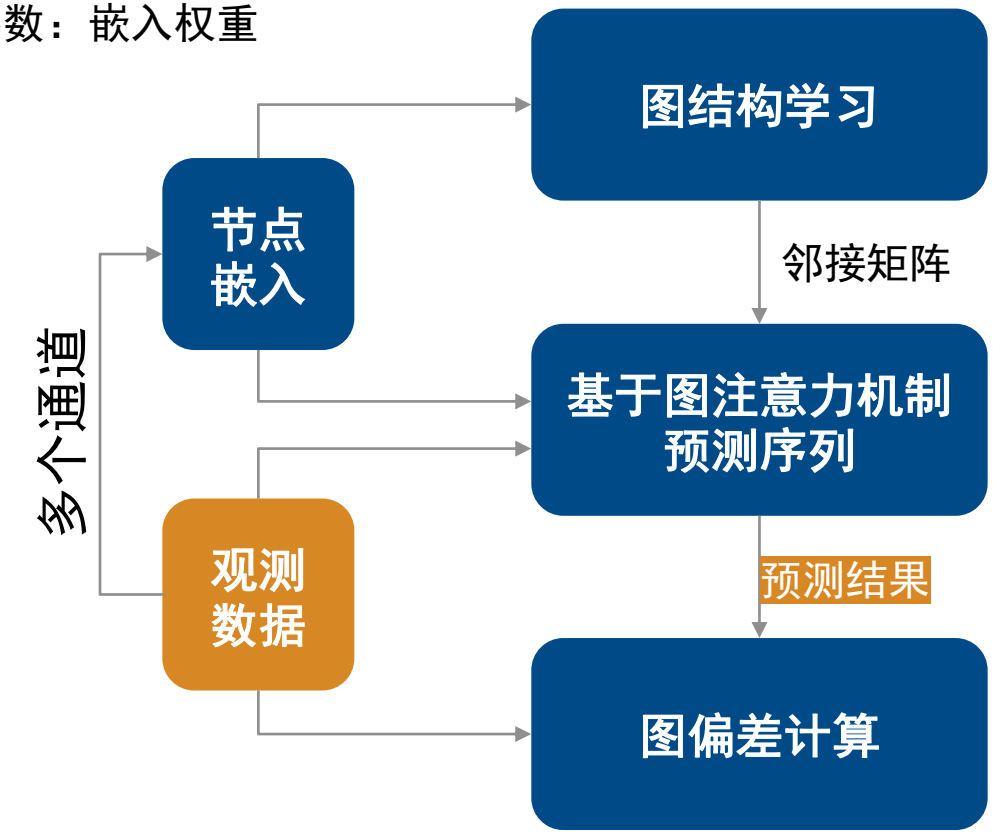


多变量异常检测问题和单变量问题一样吗？
变量间的相关性对检测结果是否会有影响？

示例：基于图偏差神经网络的多元时间序列异常检测

Demo: Multivariate Time Series Anomaly Detection Using Graph Neural Network

1. 将时间序列处理为图数据
   为每个通道生成嵌入向量以表征信号特性
   可学习参数：嵌入权重



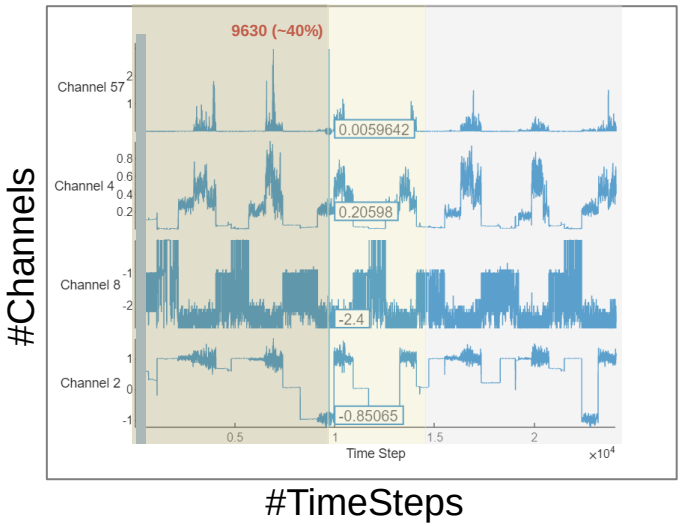
2. 结构未知的图结构学习
   生成邻接矩阵表示通道间的相关性
   将图作为输入

3. 引入自注意力机制提取特征，用于预测时序

4. 比较预测值和观测值得出异常判定指标

图偏差神经网络训练过程和结果

数据集准备

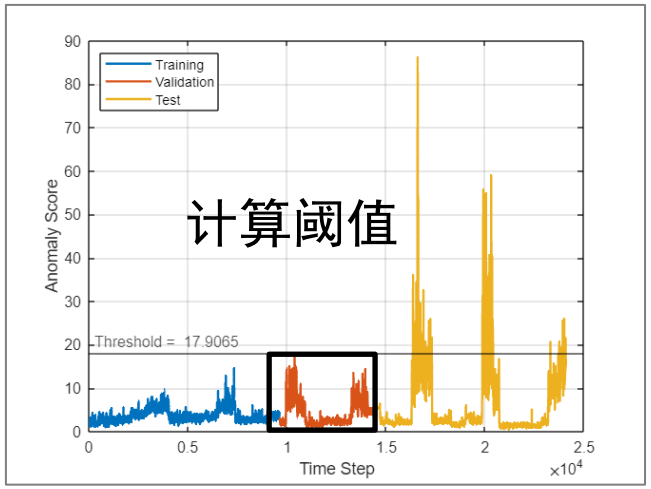


训练模型 – 前40%观测样本
- 输入X：(t-w,t-1)短序列（通过滑窗掩码处理）
- 输出T：t 时刻的观测量

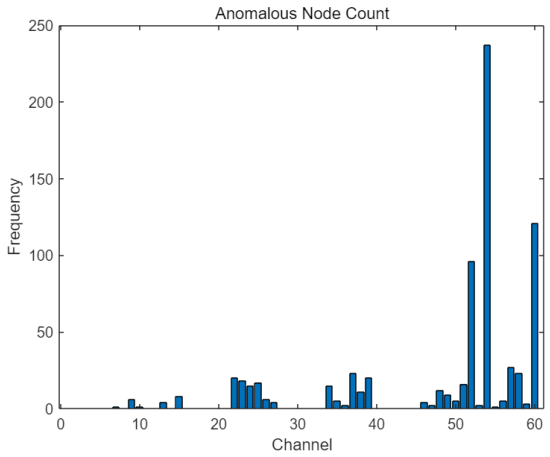
验证模型 – 接下来的20%观测样本
测试模型 – 剩余40%观测样本

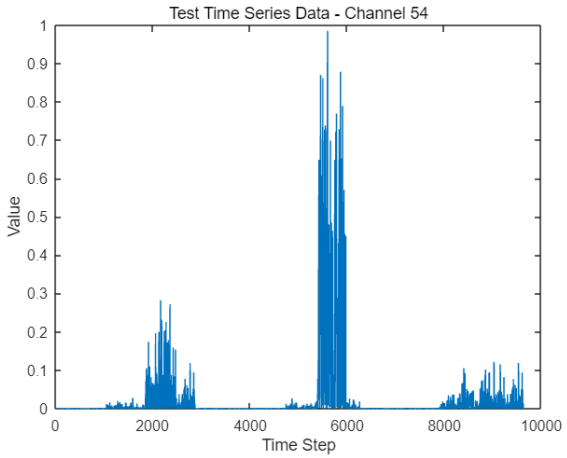
结果

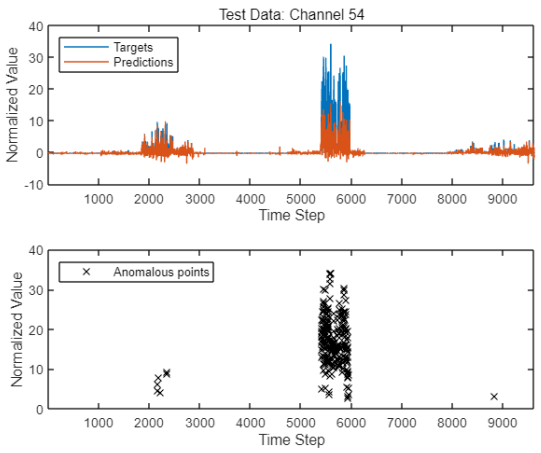
每个通道的异常得分（标准化处理）
整体异常得分 （最大值）



定位异常

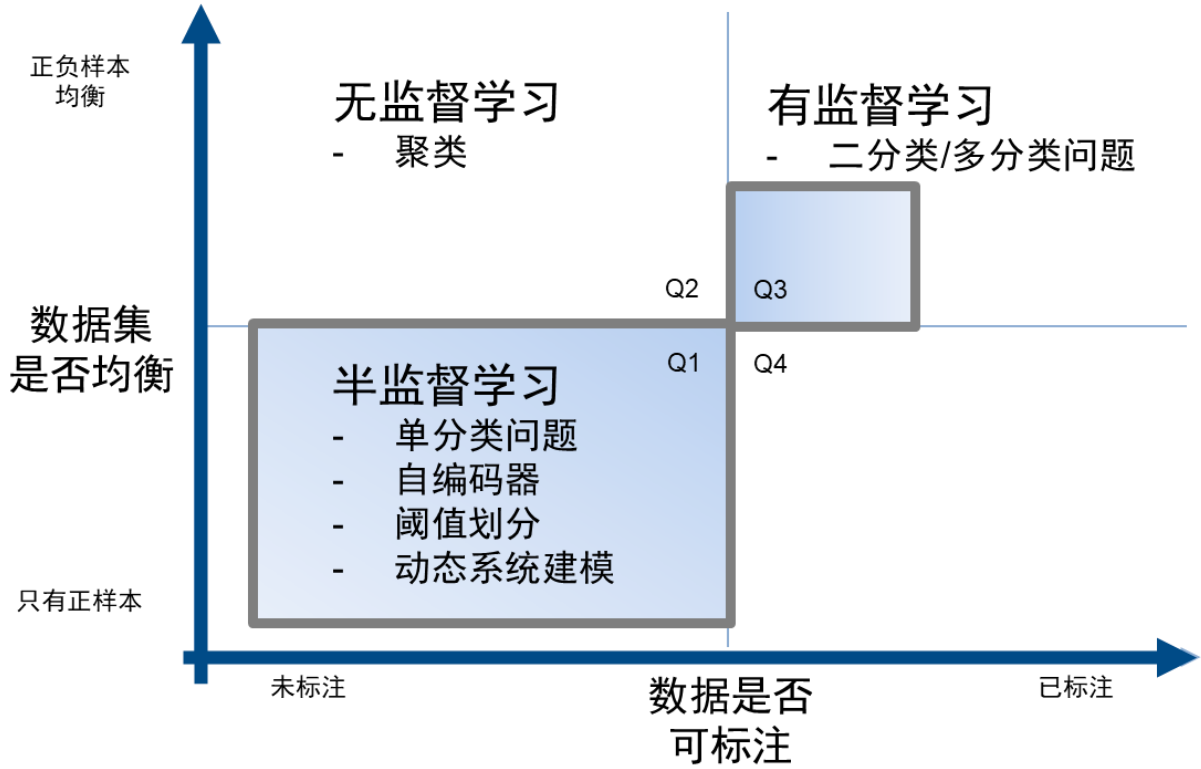




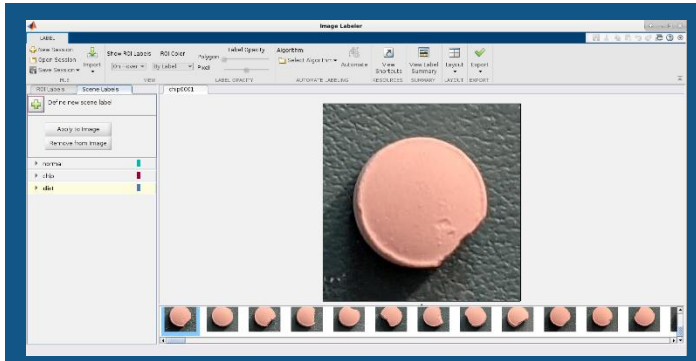


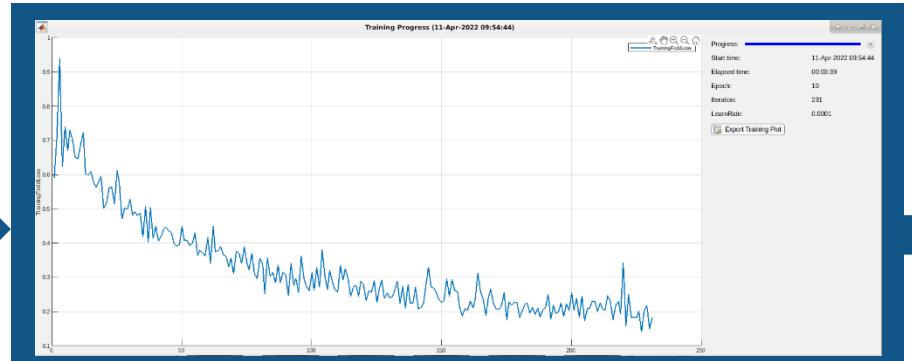
总结

- 传感器测量长时间序列
- 数据噪声问题：时间戳缺失、测量噪声、时间扭曲和移位
- 多数情况下无法准确标注
- 即使数据集有标签，正负样本数量可能不均衡，异常数据占比低
- 采集的数据来自多种不同工况，随时间推进存在老化、磨损等退化现象
- 多元时间序列的异常检测需要考虑变量之间的关联以确定系统级异常

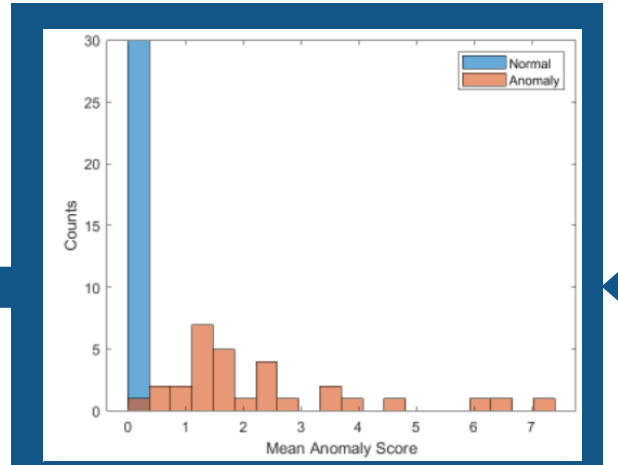


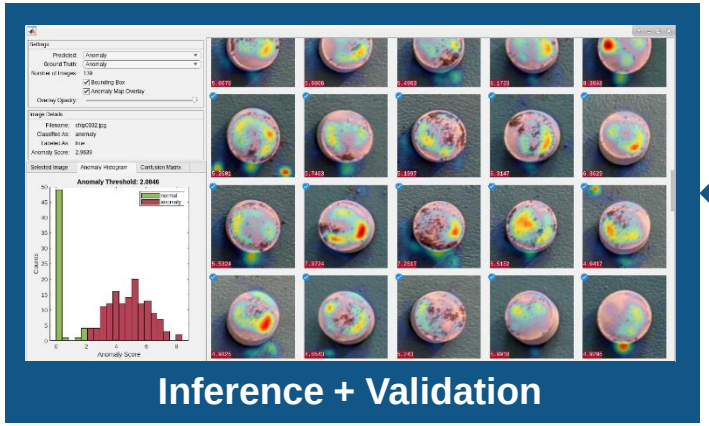
了解更多：图像数据异常检测





- FCDD Anomaly Detector
- FastFlow Anomaly Detector
- PatchCore Anomaly Detector





- MaxFalsePositiveRate
- MaxFalseNegativeRate
- Max F1 score

Image Processing Toolbox
Computer Vision Toolbox
Automated Visual Inspection Library

MATLAB EXPO

谢谢！ 问题？