

《交叉项目训练—卫星数据分析和健康监测》

期末大作业考核方式

一、（必做题）卫星数据故障检测与故障诊断（90 分）

1. 数据集说明

课程提供 3 个卫星分系统（姿轨控、供配电与激光载荷）的训练集数据，以 csv 格式存储，保存在 train 文件夹下，文件组织方式如下图。



姿轨控数据共 40329 条，包含 32896 条正常数据与 7433 条故障数据。数据共有 95 个特征维度，故障类别共 10 种，类型与个数如下表所示。

分系统	故障类型	样本数
姿轨控	陀螺 1 输出常值故障	724
	动量轮 2 输出力矩下降故障	689
	磁力矩器 X 输出磁矩下降故障	1253
	太阳敏感器 1 故障	600
	负 Y 翼 A 轴故障	763
	陀螺 2 输出常零故障	600
	星敏感器 1 输出噪声变大故障	600
	动量轮 1 输出力矩下降故障	600
	动量轮 3 输出力矩下降故障	804

	动量轮 4 输出力矩下降故	800
	总计	7433

供配电数据共 22784 条，包含 16654 条正常数据与 6130 条故障数据。数据共有 55 个特征维度，故障类别共 15 种，类型与个数如下表所示。

分系统	故障类型	样本数
供配电	单体电池开路	46
	蓄电池组加热带误通	659
	单串电池片短路	517
	单体电池短路	565
	蓄电池组加热带误断	681
	第 17 级 S3R 模块只能分流故障	366
	单串电池片开路	105
	五串电池片功率损失故障	498
	五串电池开路故障	321
	第 17 级 S3R 模块只能供电故障	349
	第 13 级 S3R 模块只能分流故障	207
	第 13 级 S3R 模块只能供电故障	331
	五串电池短路故障	684
	五串电池性能严重衰减	96
	下位机通信异常或死机	705
	总计	6130

激光载荷数据共 42964 条，包含 37864 条正常数据与 5100 条故障数据。数据共有 132 个特征维度，故障类别共 14 种，类型与个数如下表所示。

分系统	故障类型	样本数
激光载荷	转台方位轴电机故障	6
	转台俯仰轴电机故障	6
	转台方位轴码盘故障	6
	章动跟踪快反镜执行体故障	620
	转台方位轴精度下降	2018

	转台俯仰轴驱动器故障	6
	转台方位轴驱动器故障	6
	EDFA 光输出中断	34
	信号激光器故障	18
	转台俯仰轴精度下降	2018
	SWIR 图像探测器有坏点	48
	俯仰轴解锁异常	240
	转台俯仰轴码盘故障	6
	精跟踪快反镜 1 执行体故障	68
	总计	5100

同时助教保留了同一分布下相同数量的 3 个分系统数据作为测试集数据，以评价故障检测算法与故障诊断算法的优劣。测试集数据保存在 `test` 文件夹下。文件组织形式与 `train` 文件夹相同。

2. 成果说明

成果一：故障检测算法模型。对于有监督或无监督的机器学习算法，有监督或无监督的神经网络模型，选择其一。可以只使用训练集中的正常数据，也可以额外用到训练集中的故障数据，完成算法学习或模型训练。算法模型能准确地将正常数据检测为正常，将异常或故障数据检测为异常。

成果二：故障诊断算法模型。利用监督式的机器学习算法（例如，贝叶斯决策、Logistic 回归、SVM、神经网络模型），对故障数据进行分类。使用到有标签的故障数据，完成算法学习或模型训练。算法模型能准确地将故障数据归类到其对应的故障类型。

3. 成果形式

- 1) 自行编写数据处理、算法模型学习训练等的代码。编程语言建议使用 Python，可以使用 Python 脚本或 Jupyter Notebook。可以参考平时作业。注意代码的可读性与可维护性。将算法封装成易调用的函数或类，以便助教使用测试集数据进行测试。代码文件夹下需要有 README 文件介

绍各个函数与类的作用以及具体的调用方法。

- 2) 可以保留计算的中间结果。
- 3) 撰写简要的实验报告，WORD 或 PDF 均可。实验报告中需要报告模型自测的结果（参见第 5 部分“考核方式”与第 6 部分“成果考核表”）。
- 4) 进行课堂成果分享的 PPT 演示文稿。
- 5) 将源代码、中间结果（如有）、实验报告与 PPT 在规定时间内打包上传至网络学堂。不接受任何形式的补交。逾时未交，大作业一项成绩记 0 分。

4. 考核指标

1) 故障检测

对于成果一：故障检测算法模型，使用虚警率与漏警率评价性能指标。

预警 \ 故障	故障	无故障
报警	TP	FP
未报警	FN	TN

表中 TP 表示模型给出报警且实际预发生了故障；FP 表示模型给出报警但实际无故障；FN 表示模型未给出报警但实际发生了故障；TN 表示模型未给出报警且实际也无故障。

虚警率是在未发生故障的样本中检测算法模型报出故障的比例。

虚警率(False Alarm Rate)计算式为： $FAR = \frac{FP}{FP+TN}$

漏警率是在发生故障样本中检测算法模型未报出故障的比例。

漏警率(Miss Alarm Rate)计算式为： $MAR = \frac{FN}{FN+TP}$

2) 故障诊断

对于成果二：故障诊断算法模型，使用诊断准确率进行评价。

诊断准确率计算公式： $Acc = \frac{n}{m}$

其中， Acc 表示诊断准确率， n 表示诊断模型正确诊断故障的次数， m 表示诊断模型诊断故障的次数。故障准确率表示为诊断模型正确诊断故障的次数占总的故障诊断次数的比例。

5. 考核方式

对算法模型进行性能测试。

- 1) 模型自测。可以在训练集数据当中根据需要自行划分验证集。计算模型的检测虚警率、检测漏警率与诊断准确率，输出成果考核表中的各项结果。推荐将成果考核表中的各项结果保存为 EXCEL 文件。Python 读写 EXCEL 方法参考 Pandas 文档 (<https://pandas.pydata.org/docs/index.html>)。
- 2) 助教测评。助教将 train 文件夹替换为 test 文件夹，同时根据 README 文件修改代码中的输入/输出部分。要求程序能正常运行并输出成果考核表中的各项结果。推荐将成果考核表中的各项结果保存为 EXCEL 文件。

此外，考核也会参考代码的可读性与可维护性以及实验报告和 PPT 的完成程度。

6. 成果考核表

1) 故障检测：

分系统	名称	检测次数	报警次数	虚警率/漏警率
姿轨控	正常数据	32896		
	故障数据	7433		
供配电	正常数据	16654		
	故障数据	6130		
激光载荷	正常数据	37864		
	故障数据	5100		

2) 故障诊断：

分系统	故障类型	诊断次数	诊断正确次数	准确率
姿轨控	陀螺 1 输出常值故障	724		
	动量轮 2 输出力矩下降故障	689		
	磁力矩器 X 输出磁矩下降故障	1253		

	太阳敏感器 1 故障	600		
	负 Y 翼 A 轴故障	763		
	陀螺 2 输出常零故障	600		
	星敏感器 1 输出噪声变大故障	600		
	动量轮 1 输出力矩下降故障	600		
	动量轮 3 输出力矩下降故障	804		
	动量轮 4 输出力矩下降故障	800		
	总计	7433		
供配电	单体电池开路	46		
	蓄电池组加热带误通	659		
	单串电池片短路	517		
	单体电池短路	565		
	蓄电池组加热带误断	681		
	第 17 级 S3R 模块只能分流故障	366		
	单串电池片开路	105		
	五串电池片功率损失故障	498		
	五串电池开路故障	321		
	第 17 级 S3R 模块只能供电故障	349		
	第 13 级 S3R 模块只能分流故障	207		
	第 13 级 S3R 模块只能供电故障	331		
	五串电池短路故障	684		
	五串电池性能严重衰降	96		
	下位机通信异常或死机	705		
	总计	6130		
激光载荷	转台方位轴电机故障	6		
	转台俯仰轴电机故障	6		
	转台方位轴码盘故障	6		
	章动跟踪快反镜执行体故障	620		
	转台方位轴精度下降	2018		

	转台俯仰轴驱动器故障	6		
	转台方位轴驱动器故障	6		
	EDFA 光输出中断	34		
	信号激光器故障	18		
	转台俯仰轴精度下降	2018		
	SWIR 图像探测器有坏点	48		
	俯仰轴解锁异常	240		
	转台俯仰轴码盘故障	6		
	精跟踪快反镜 1 执行体故障	68		
	总计	5100		

二、（选做题）电源分系统剩余寿命预测（10 分）

1. 数据集说明

数据集包含 124 个标称容量为 1.1 Ah、标称电压为 3.3 V 的锂离子电池在各种充放电情况下的测量结果。完整的数据集可查阅 <https://data.mtr.io/1> 或 <https://cloud.tsinghua.edu.cn/d/8a6d27ce37d549278fb8/>。详细说明可查阅 Severson, K.A., Attia, P.M., Jin, N. et al. "Data-driven prediction of battery cycle life before capacity degradation." Nat Energy 4, 383–391 (2019). <https://doi.org/10.1038/s41560-019-0356-8>。

训练集数据包含 41 个电池测量数据，验证集数据包含 43 个电池测量数据，测试集数据包含 40 个电池测量数据。数据以 MATLAB 文件格式存储，其中包括以下信息：

- 1) 描述性数据： 电池条形码、充电策略、循环寿命；
- 2) 每周周期汇总数据： 周期数、放电容量、内阻、充电时间；
- 3) 周期内收集的数据： 时间、温度、样条插值放电容量、样条插值电压。

2. 成果说明

利用课堂上讲授的方法，实现一个根据电池早期循环数据（例如前 100 次循环）对电池循环寿命进行预测的模型。可以使用特征工程，也可以训练一个端到

端的神经网络。要求能够在只使用电池前几次（例如前 100 次）循环的充放电数据的前提下对电池剩余循环寿命进行准确预测。

3. 成果形式

- 1) 自行编写数据处理、算法模型学习训练等的代码。编程语言建议使用 Python，可以使用 Python 脚本或 Jupyter Notebook。可以在助教提供的代码(<https://cloud.tsinghua.edu.cn/d/f33c2ff7ae5b4a8fa9f6>)基础上加以改进。注意代码的可读性与可维护性。代码文件夹下需要有 README 文件介绍各个函数与类的作用以及具体的调用方法。
- 2) 可以保留计算的中间结果。
- 3) 撰写简要的实验报告，WORD 或 PDF 均可。实验报告中需要报告模型在训练集、验证集与测试集上的性能指标。
- 4) 如果选做此题，将选做部分的源代码、中间结果（如有）和实验报告与必做部分一同打包（选做部分与必做部分分成两个文件夹），并在规定时间内上传至网络学堂。

4. 考核指标

使用均方误差 (Mean Square Error, MSE) 与平均百分比误差 (Mean Percentage Error, MPE) 进行评价。假设真实的电池剩余寿命为 $\{y_i\}_{i=1}^n$ ，利用模型预测的电池剩余寿命为 $\{\hat{y}_i\}_{i=1}^n$ ，均方误差的计算公式为

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

平均百分比误差的计算公式为

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|}$$

5. 考核方式

根据实验报告和源代码的完成程度进行评分，主要考察方法是否具有创新性、模型的性能指标是否达到要求，代码是否有可读性与可维护性等。