

天津大学

数据挖掘综合实践课程设计任务书



题目：面向工业时间序列的异常检测

2025 年 12 月

一 任务概述

完成数据预处理，实现一个有监督异常检测方法与两个无监督异常检测方法，分析比较不同方法的异常检测表现。数据具有异常数据与正常数据不平衡的特点，在课程设计中需要重点分析。

二 任务内容

2.1 数据预处理

- (1) 数据归一化：数据为原始数据
- (2) 数据标注：数据无标签
- (3) 训练集与测试集划分：是否进行信号筛选，是否使用全部数据。

可扩展数据去噪等。

2.2 算法实践与比较

- (1) 数据可视化分析
- (2) 1个有监督算法实践，针对异常建立二分类方法。
- (3) 2个无监督算法实现，建立异常评分标准。
- (4) 计算异常检测准确率、Precision、Recall
- (5) 实验结果分析
- (6) 实验结果深入分析：长尾数据对异常检测的影响

可扩展：故障诊断分类，根据工业异常检测难点进行实验结果深入分析

三 数据集

以下两个数据集，任选其一。

3.1 Cranfield Multiphase Flow Facility 数据集

场景：多相流（油、气、水）实验设施。

数据特点：压力、温度、密度、差压信号。

注意选择多相流数据集不可利用数据集中提供的 CVA 方法进行异常检测

3.2 Secure Water Treatment (SWaT) 数据集

场景：现代化的六阶段水处理工厂。

数据特点：包含 51 个传感器（水位、流量等）和执行器（泵、阀门）的连续数据。

四 成果提交要求

1. 课程设计报告或论文，提交 pdf 版本。
2. 完整的源代码工程文件。
3. 将课程设计报告、代码等源文件整理成 zip 格式的压缩包，以“数据挖掘综合实践-学号姓名-学号姓名-...-学号姓名”的方式命名，包含小组内所有同学的“学号和姓名”，提交至智慧树。
4. **提交截止时间：2025 年 12 月 28 日晚 23:59。迟交 1 周以内扣分 10%，以此类推。**

五 组队要求

每个小组不超过 5 人，在报告中注明小组成员姓名、学号、具体分工。

六 评分标准

6.1 基本评分

- 数据预处理（15%），
- 1 个有监督算法实践（20%），
- 2 个无监督算法实现（40%），
- 实验分析（20%），
- 报告写作（5%）。

6.2 额外加分

- 数据去噪、特征工程等数据预处理、模型训练扩展内容（可额外加 3 分）
- 多分类故障诊断（可额外加 5 分）
- 根据工业异常检测难点进行实验结果深入分析（可额外加 5 分）

额外加分不超过 5 分，总分不超过 100 分