

# 题目：多变量时间序列异常检测

**简介：**多变量时间序列由不同变量随时间变化的序列组成，这些时间序列在实际应用中通常来自不同的传感器或数据源。多变量时间序列异常检测是指在多个相互关联的时间序列中识别出不符合正常模式的异常点。异常检测的应用场景包括 IT 运维监控、交通流量管理、环境检测、能源管理等。例如在空气质量检测数据中，通过监测 PM2.5、二氧化碳等数据的变化，可以及时发现污染事件，采取相应的环境保护措施。

**任务描述：**给定一段时期内的服务器性能监测数据，且是随时间记录等间隔的观测数据。

- 多变量时间序列异常检测的输入表示为  $x \in R^{n \times k}$ ，其中  $n$  表示时间戳的最大长度， $k$  表示输入特征数量。
- 多变量时间序列异常检测任务的输出向量是  $y \in R^n$ ，其中  $y_i \in \{0,1\}$  表示第  $i$  个时间戳是否为异常，0 为正常，1 为异常。

**数据：**本数据集包含 25 个服务器在不同时间点收集的指标数据。每个指标表示服务器的某种性能度量，例如 CPU 使用率、内存使用率等。数据采样间隔为 5 分钟，每条时间序列代表一个指标。

## 文件说明：

train.csv: 包含训练数据，数据形状 [时间步数, 特征数]，每个时间步表示一个采样时间点，每个特征表示一个服务器指标。文件包含 26 列，第 1 列 timestamp\_(min) 表示时间戳，其余 25 列 feature\_X 表示不同服务器指标数据。训练数据集中不包含异常标签。

test.csv: 包含测试数据，数据形状 [时间步数, 特征数]。其余含义同 train.csv 文件说明。

## 提交文件：命名为 detection.npy

应包含模型对测试数据的检测结果，数据形状为 [时间步数, 1]。值为 1 表示检测为异常，0 表示检测为正常。确保检测结果的时间步顺序与测试数据一致。例如：

	A	B	C
1	timestamp_(min)	label	
2	132480	0	
3	132481	0	
4	132482	0	
5	132483	0	
6	132484	0	
7	132485	0	
8	132486	0	
9	132487	0	
10	132488	0	
11	132489	0	
12	132490	0	
13	132491	0	
14	132492	0	
15	132493	0	
16	132494	0	
17	132495	0	

**评估指标：**

同时考虑三个评估指标：精确率（Precision）检测到的异常点中实际为异常的比例，召回率（Recall）实际异常点中被正确检测出来的比例，F1-score 精确率和召回率的调和平均数，综合评估模型的检测性能。

**提示：**

可参考以下论文及相应代码

- Title: Outlier detection for multidimensional time series using deep neural networks  
Github: [https://github.com/matanle51/LSTM\\_AutoEncoder](https://github.com/matanle51/LSTM_AutoEncoder)
- Title: DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection  
Github: <https://github.com/DAMO-DI-ML/KDD2023-DCdetector>
- Title: Timesnet: Temporal 2d-variation modeling for general time series analysis  
Github:  
(1) <https://github.com/thuml/TimesNet>  
(2) <https://github.com/thuml/Time-Series-Library>

**作业提交形式：**

1. 检测结果 detection.npy 文件。
2. 一页 A4 纸的报告。描述如何实现（比如基于上述某篇文章），以及对已有方法的可能改进方向的思考。