

基于深度学习的光电倍增管暗噪声 计数率实时异常检测系统

吴寅慧

实验粒子物理计算研讨会

西昌, 四川

2025.07.17

主要内容

01

江门中微子实验 Jiangmen Underground Neutrino Observatory, JUNO

实验介绍 探测器系统

02

JUNO探测器异常检测

需求分析 技术选型

03

基于深度学习的光电信增管暗噪声计数率异常检测

整体架构 模型训练 异常检测 模型评估 模型优化 后续计划 总结

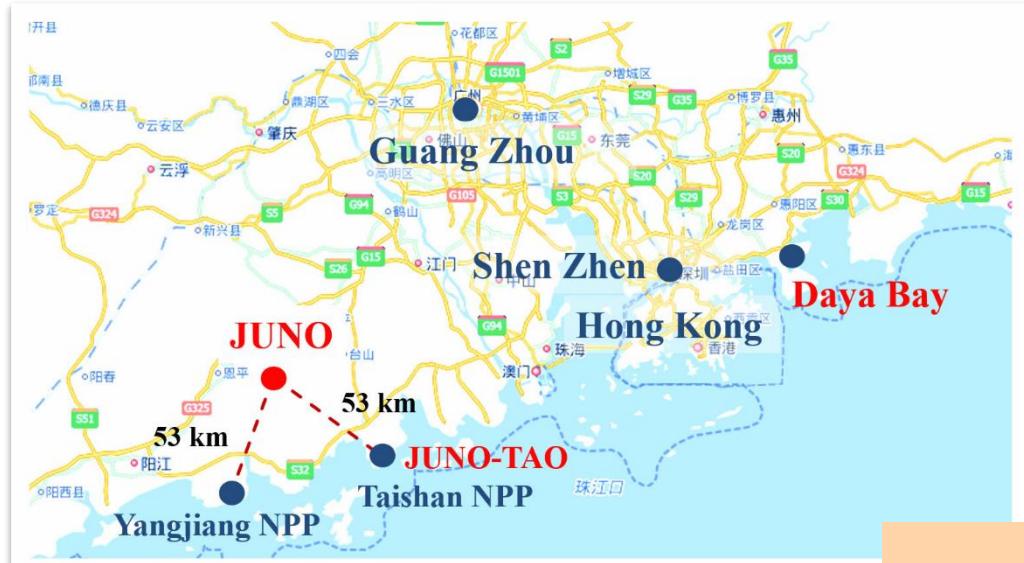
01江门中微子实验

实验介绍

江门中微子实验 Jiangmen Underground Neutrino Observatory, JUNO

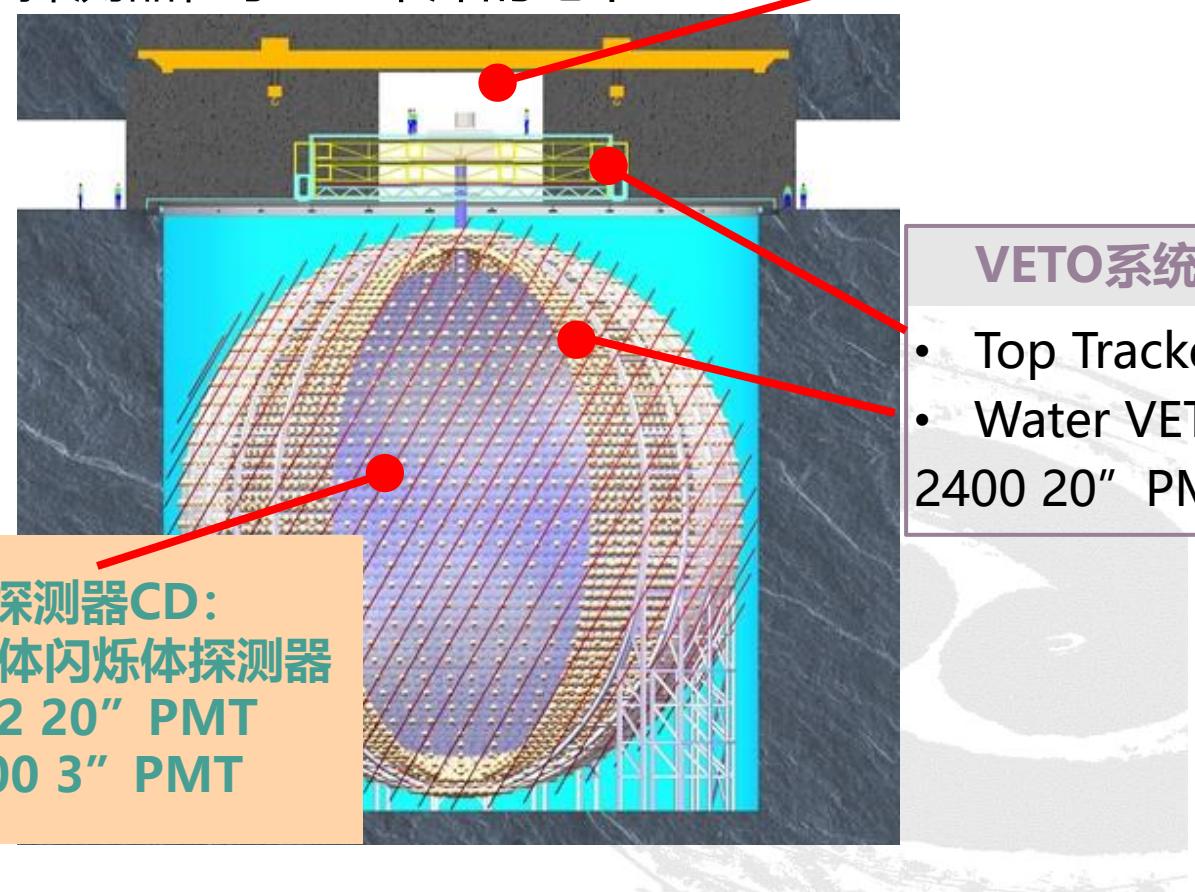
新一代中微子观测站，建设在中国广东江门市，探测器位于700米深的地下

刻度系统



目标：明确不同种类中微子质量顺序
精确测量中微子振荡参数
探测超新星爆发产生的中微子

中心探测器CD：
两万吨液体闪烁体探测器
17612 20" PMT
25600 3" PMT



VETO系统

- Top Tracker
- Water VETO

2400 20" PMT

JUNO探测器系统

探测器是JUNO实验的“眼睛”

组成：中心探测器、水切伦科夫探测器、顶部径迹探测器

任务：采用光电倍增管photomultiplier tube, PMT阵列进行中微子信号探测

探测器通过三大核心设计，成为破解中微子质量顺序的关键工具

- ◆ 超高能量分辨率
- ◆ 超低本底环境
- ◆ 大规模光电探测阵列

超高的能量分辨率依赖于大规模的光电探测阵列的正常工作

- ◆ 正常工作状态下，PMT的暗噪声计数率dark count rate, DCR保持相对稳定
- ◆ 当PMT出现异常发光现象时，会诱发周边PMT产生响应，导致虚假信号（本底噪声）的生成

探测器的稳定关系到实验数据的物理准确性，因此对PMT的健康监测变得尤为重要！



02 探测器通道级异常检测

需求分析

提出对PMT的暗噪声计数率DCR进行异常监测

■ **核心目标：**确保PMT的稳定工作，从而保障整个实验数据的可靠性

由于PMT的制造差异，环境不同等因素，其DCR的特性也各不相同

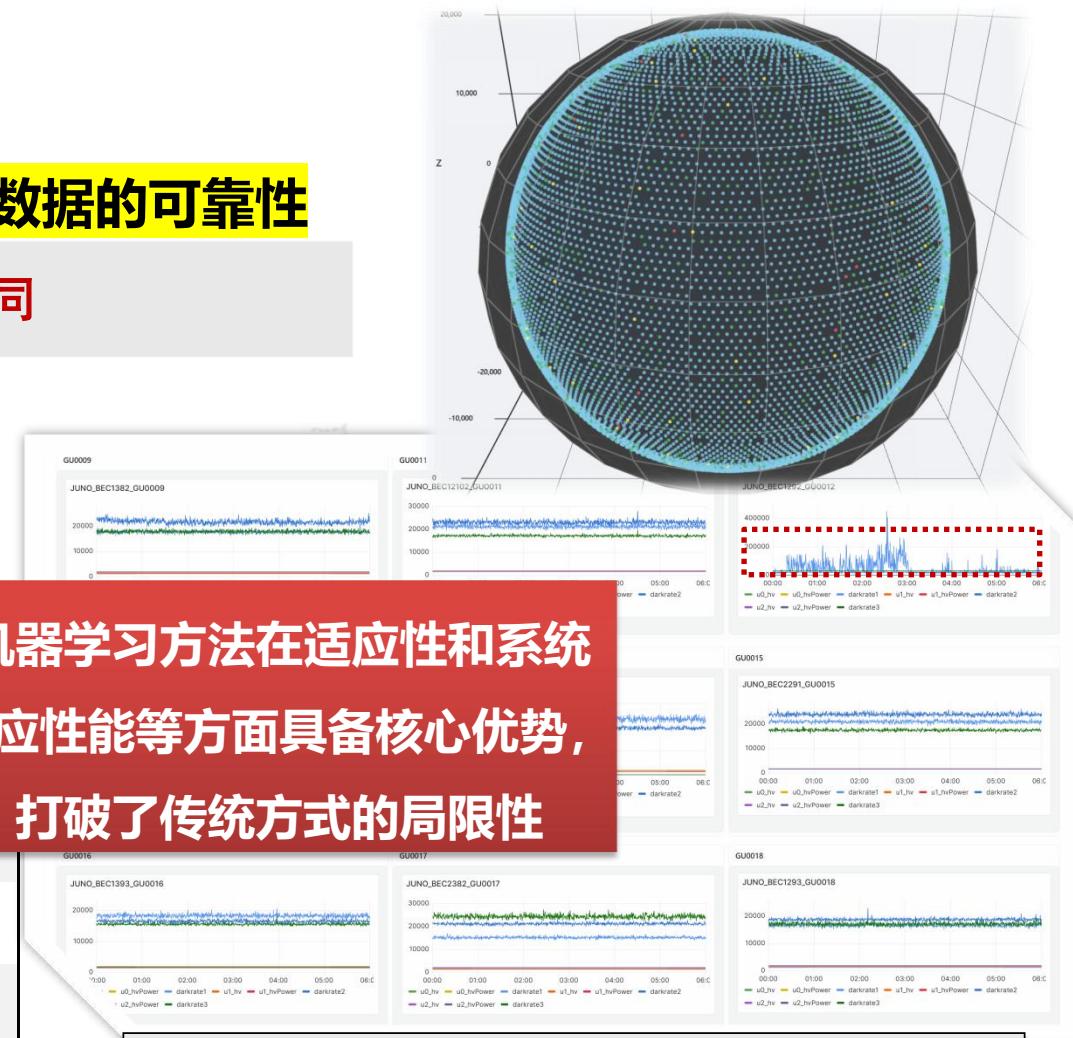
■ **传统异常检测方法**

单一的阈值检测：完全依赖人工经验，无法动态适应

基于规则引擎：仅能处理预定义的简单空间模式

	传统方法	机器学习方法
环境适应性	固定阈值	动态百分位阈值
通道差异处理	全局统一阈值	通道独立阈值
时空关联性	无法量化级联响应	实现复杂时空建模
响应性能	分钟延迟	秒量级响应

机器学习方法在适应性和系统
响应性能等方面具备核心优势，
打破了传统方式的局限性

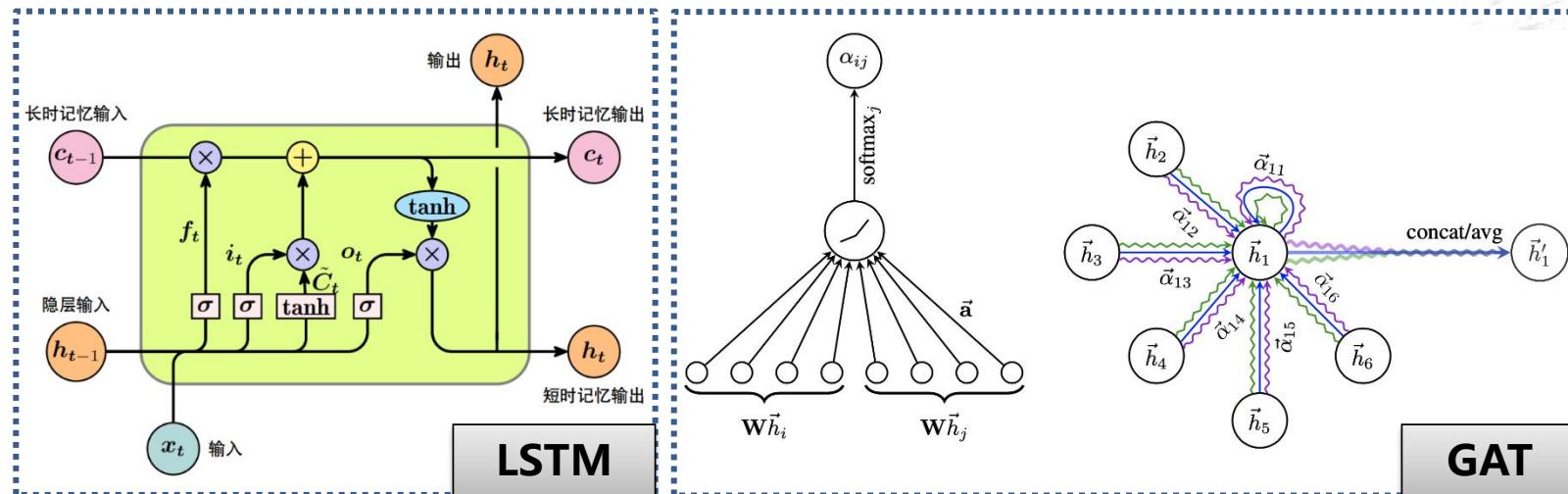


根据PMT的DCR特性，选择适合的机器学习模型

核心思想：建立数据驱动的物理规律学习机

需要考虑PMT特性及相互之间的影响和依赖关系

- ◆ 探测器有自己的生命周期，老化会发生缓慢漂移
- ◆ 每个PMT不能被视为一个独立单元，临近的PMT之间存在依赖



模型选择

- 综合PMT时间和空间特征

LSTM模型

捕获复杂时间依赖具有不可替代性

GAT模型

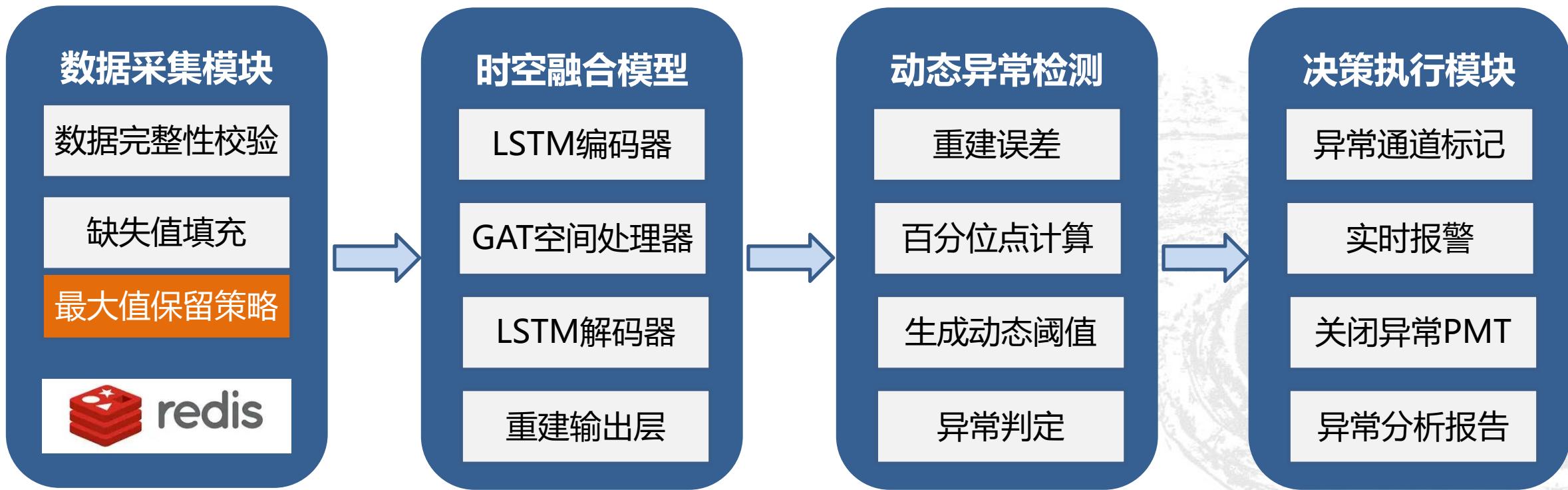
具有注意力机制
灵活学习不同邻居的重要性权重

03基于深度学习的PMT DCR实时异常检测系统

整体架构

为了更好的完成对PMT的健康监测，保障实验可靠性

- 提出对将近20000个通道进行实时异常检测的架构方案
- 采用四层分级架构



03基于深度学习的PMT DCR实时异常检测系统

模型训练

搭建并训练时空融合特征模型

训练样本选取

2025.06.17~06.19的数据

17134个PMT 45233760条样本

训练数据集预处理

- 数据合并和清洗
- 线性插值填补DCR缺失值
- 基于时间戳的数据重采样
- **通道独立标准化处理**
- **构建空间关系图**
- GAT编码器具有**图注意力机制**，学习不同PMT的相关性，检测异常按照距离的传播

LSTM编码器

时间特征向量

GAT空间处理器

空间特征处理

LSTM解码器

特征映射回原始维度

全连接层

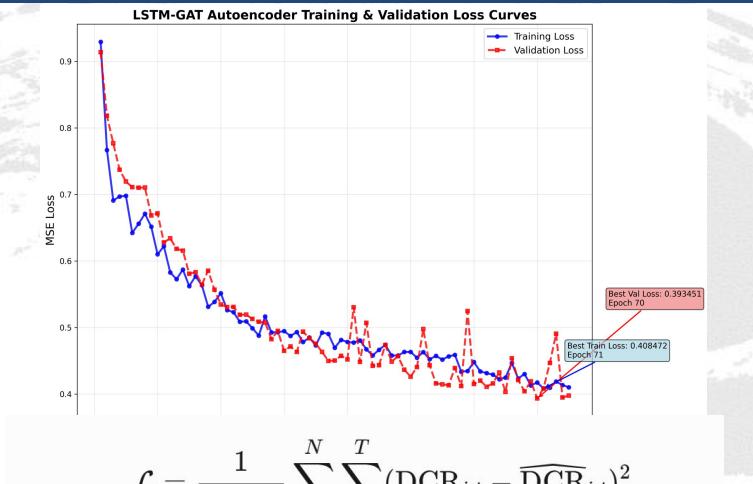
特征映射回原始维度

核心思想：学习正常的DCR模式

训练目标：最小化重建误差

- 动态调整学习率
- 早停机制
- 批处理优化
- 梯度裁剪

损失函数

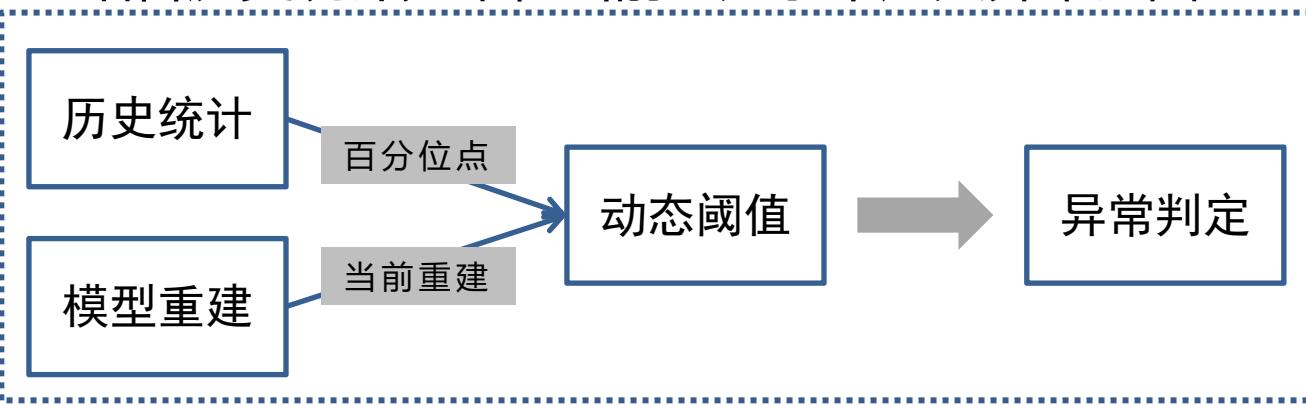


03基于深度学习的PMT DCR实时异常检测系统

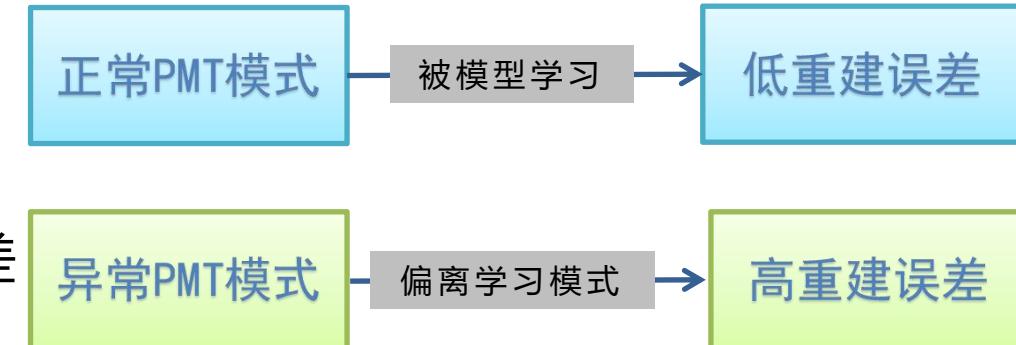
异常检测

核心：动态阈值异常检测算法

- 目的：区分正常波动和真实异常，避免误判
- 对每个通道的当前数据，计算重建序列的均值和标准差
- 结合历史统计分布和当前重建水平定义异常范围



- 历史百分位：建立基线，捕捉历史行为
- 重建均值：反应当前通道状态
- 通道自适应：每个PMT独立计算阈值，考虑个体差异
- 环境自适应：随运行条件变化自动调整



- ◆ 正常数据 → 低重建误差
- ◆ 异常数据 → 高重建误差 → 触发告警

$$\begin{cases} \text{下界} = Q_{1\%}(\text{历史}) - \mu_{\text{重建}} \\ \text{上界} = Q_{99\%}(\text{历史}) + \mu_{\text{重建}} \end{cases}$$

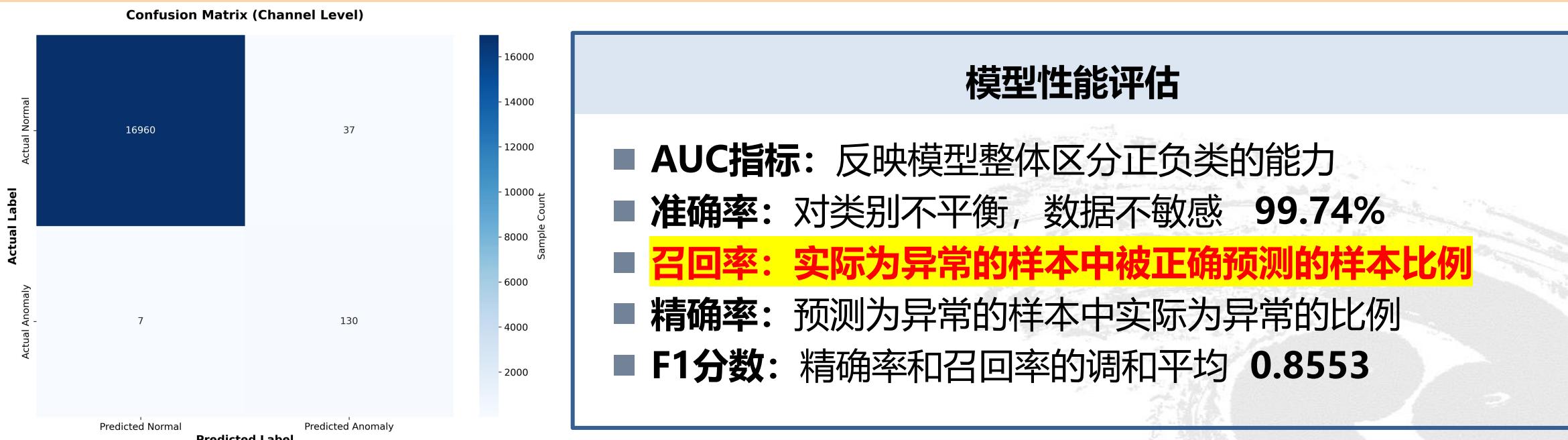
动态融合历史统计和模型重建信息，实现高召回，低漏报的实时异常检测

03基于深度学习的PMT DCR实时异常检测系统

模型评估

检测样本选取：2025.06.20 六个小时的数据 17134个PMT 6168240条样本

成功加载 17134 条通道信息，其中 137 个被标记为异常，共涉及 17134 个独立通道进行评估



◆ 精确率 = $TP/(TP+FP)$ 避免误报 **77.84%**

◆ 召回率 = $TP/(TP+FN)$ 避免漏报 **94.89%**

精确率和召回率提供了模型性能的互补视角

从多维度探索异常检测方法的优化空间

- 提升异常模式的分辨率
- 增强空间关联建模能力
- 聚焦关键异常信号
- 平衡短期波动和长期趋势的捕获
- 防止对噪声信号的过拟合现象

模型架构

LSTM/GAT隐藏层维度
LSTM/GAT层数
重建层的激活函数类型

训练过程

学习率调整
训练轮数

数据处理

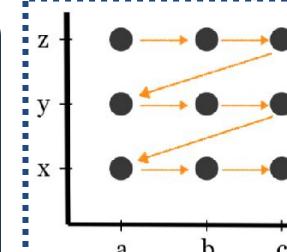
训练批次大小
batch_size

异常检测

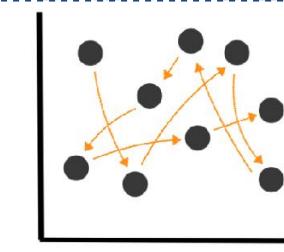
上下界计算系数
百分位点选择

参数调优策略

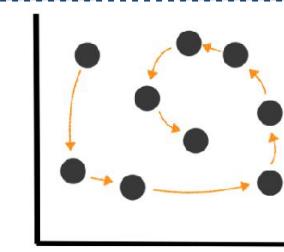
- 随机搜索方法能够在超参数空间中随机采样组合，效率较高，能发现意外好解。
- 网格搜索方法是一种超参数优化方案，通过遍历预定义的参数组合空间，寻找最佳模型配置，适合参数空间较小。



网格搜索



随机搜索

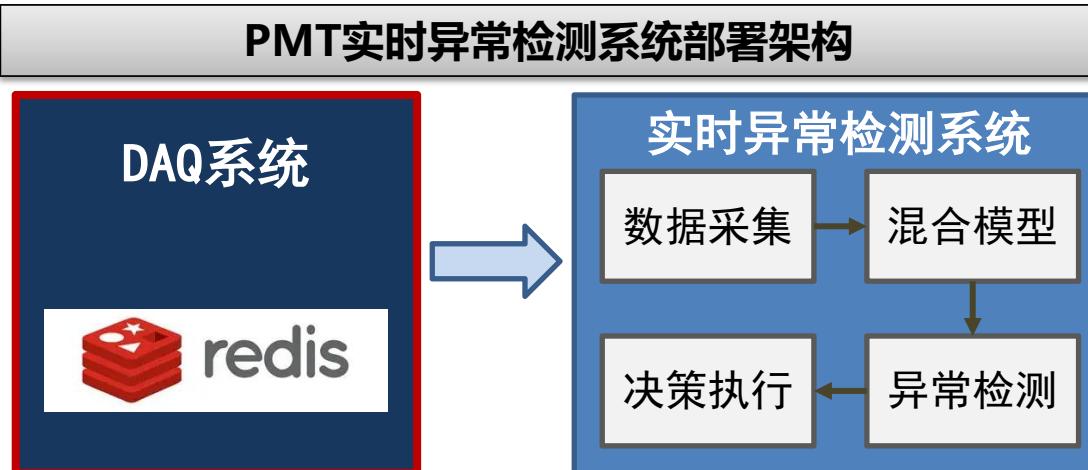


贝叶斯搜索 CSDN @Big波斯

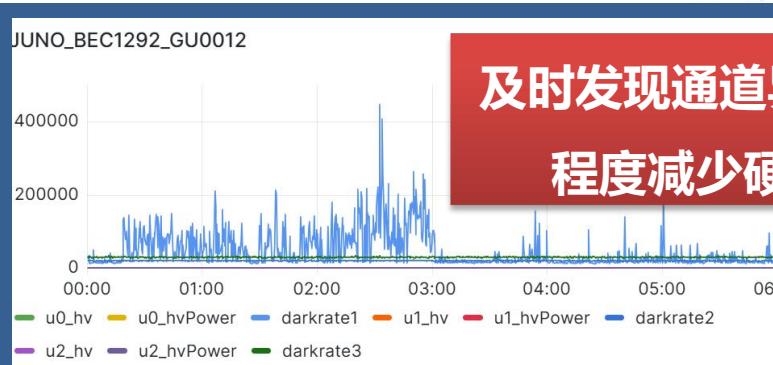
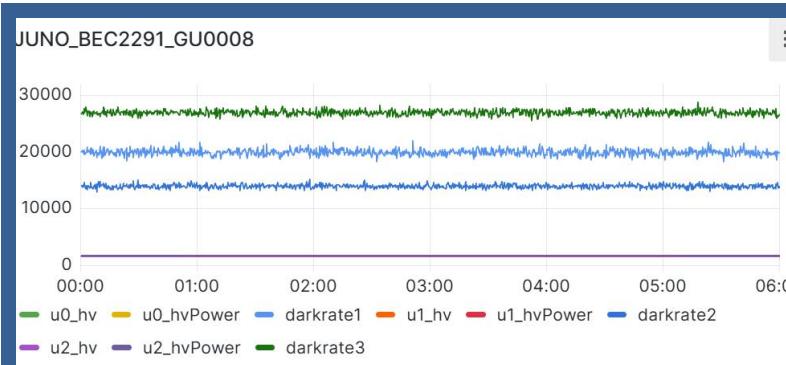
03基于深度学习的PMT DCR实时异常检测系统

后续计划

实时异常检测在JUNO实验上的应用与部署



- ◆ 模型和方法的进一步完善和优化
- ◆ 与JUNO数据获取系统DAQ进行对接
- ◆ 基于redis缓存数据库获取实时DCR
- ◆ 实时检测模块采用双缓存流式处理机制，设计动态数据接口支持分钟级延迟处理，实时获取并处理DCR数据



正常PMT的DCR波动曲线

异常PMT的DCR波动曲线

受临近异常通道影响的通道

03基于深度学习的PMT DCR实时异常检测系统

总结

PMT的稳定工作关系到超高的能量分辨率，对其暗噪声计数率的实时异常检测很有必要

传统方法

依赖人工经验，难以动态适应
人力检测存在局限，延时高
后续维护成本过高

PMT异常检测

- 设备保护：避免硬件的损坏
- 数据层面：提升物理数据的准确性
- 科学发现：保障物理研究的可靠性

- ◆ **结合PMT特性：**基线不同，受临近影响
- ◆ **设计LSTM-GAT混合自动编码器模型**
 - LSTM层提取时间依赖
 - GAT模型建模PMT阵列的空间拓扑关联
- ◆ 引入**通道自适应阈值机制**，结合历史百分位分布动态调整判定阈值
- ◆ 目前检测系统仍在测试阶段，后续会作为JUNO监测子系统投入试运行

该研究提出的时空特征融合框架和轻量化实时处理算法，为大型粒子探测器监测提供了新的解决思路

PMT的稳定工作关系到超高的能量分辨率，对其暗噪声计数率的实时异常检测很有必要

传统方法

依赖人工经验，难以动态适应
人力检测存在局限，延时高
后续维护成本过高

PMT异常检测

- 设备保护：避免硬件的损坏
- 数据层面：提升物理数据的准确性
- 科学发现：保障物理研究的可靠性

- ◆ **结合PMT特性：**基线不同，受临近影响
- ◆ **设计LSTM-GAT混合自动编码器模型**
 - LSTM层提取时间依赖
 - GAT模型建模PMT阵列的空间拓扑关联
- ◆ 引入**通道自适应阈值机制**，结合历史百分位分布动态调整判定阈值
- ◆ 目前检测系统仍在测试阶段，后续会作为JUNO监测子系统投入试运行

该研究提出的时空特征融合框架和轻量化实时处理算法，
为大型粒子探测器监测提供了新的解决思路

谢谢大家！