

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

SPMT直方图异常检测

TDAQ组

张水涵，导师：季筱璐

2024/10/16

汇报人：陈博平

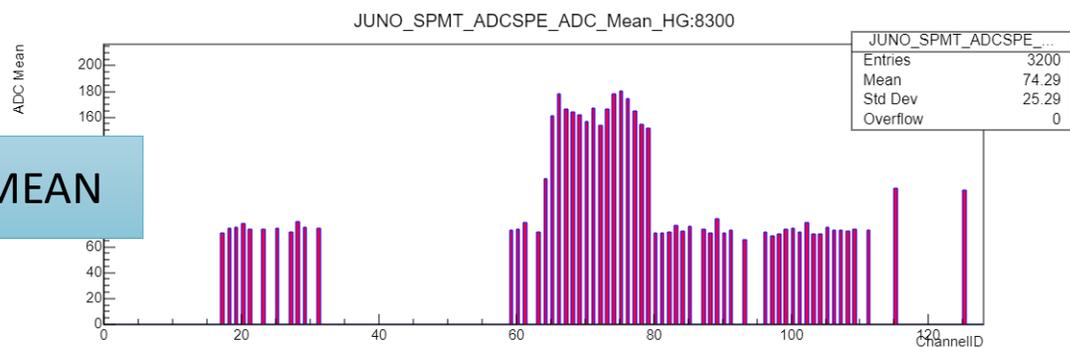
介绍

以JUNO-SPMT异常检测为例（现有方式）

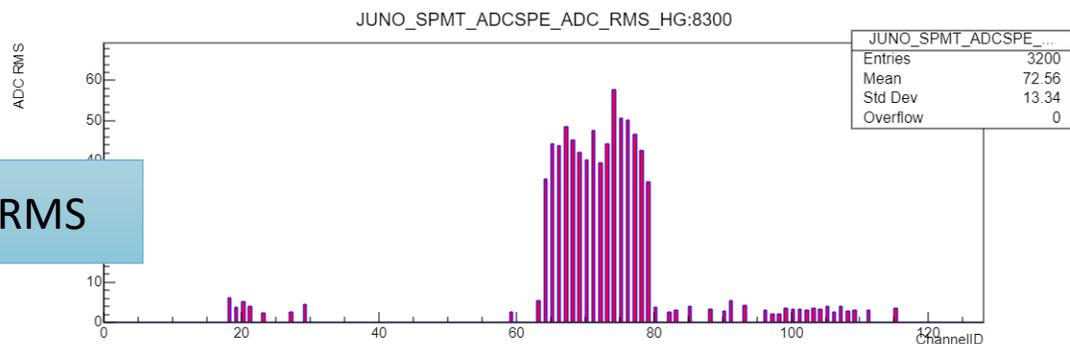
- 统计电荷能谱，均值和方差
- 可视化查看，在线绘制直方图，人工识别异常
- 存在问题：目前关灯测试已经超过~12000个通道，十分耗时

依赖人工智能和机器学习方式实现自动异常检测

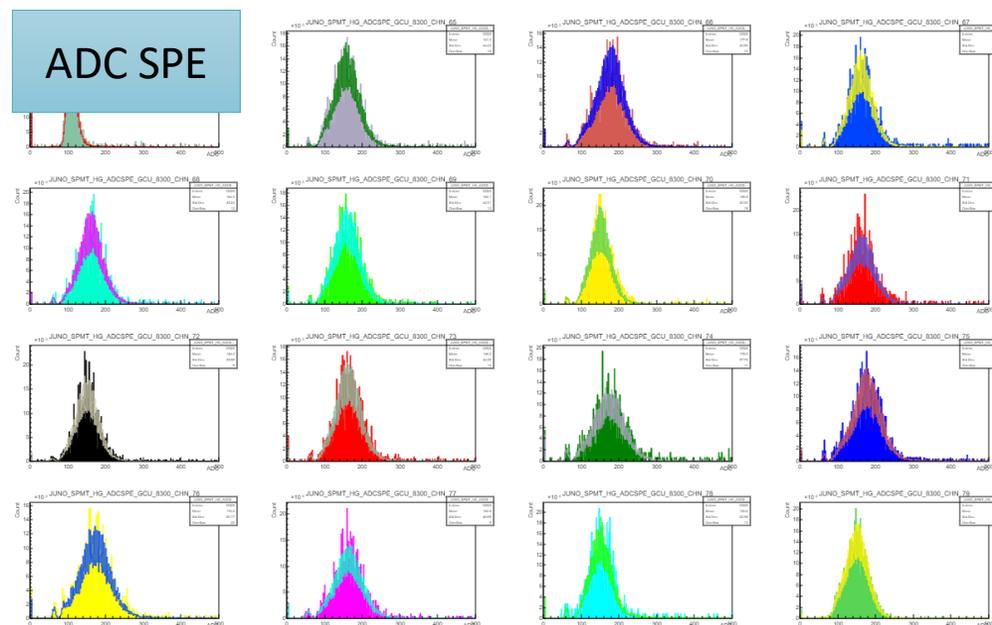
ADC MEAN



ADC RMS

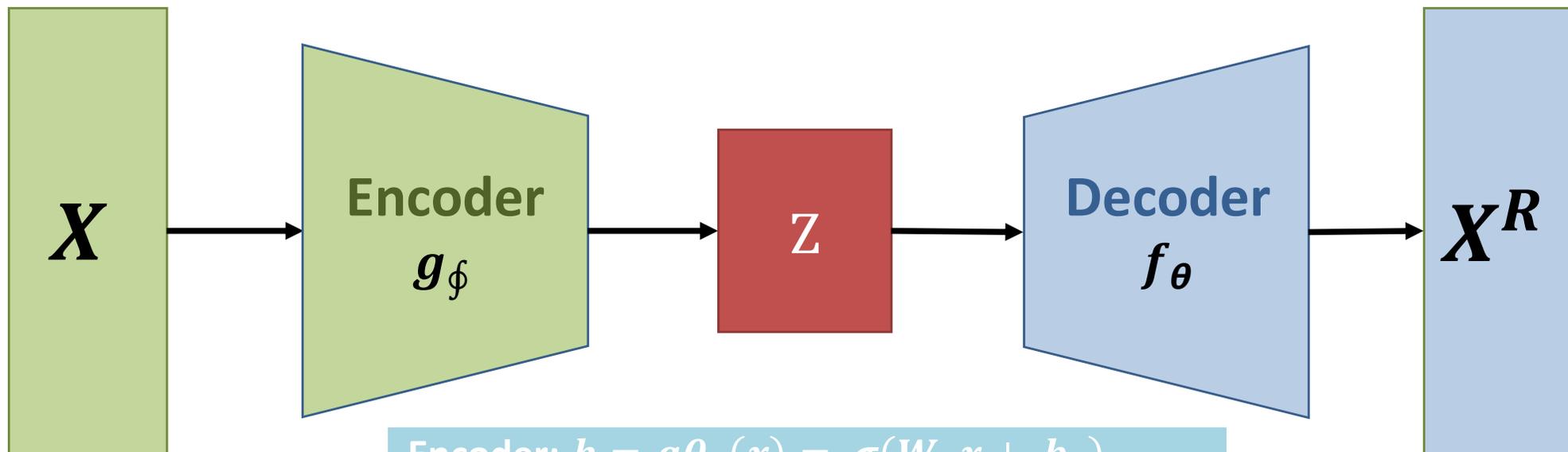


ADC SPE

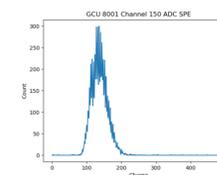
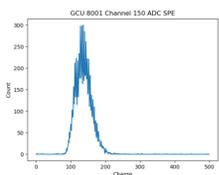


SPMT直方图异常检测

- 输入：SPMT直方图
- 损失函数：衡量原始数据与重构数据之间的差异
 - 均方误差（MSE）或交叉熵损失（BCE）
- 训练过程：MLP，反向传播算法

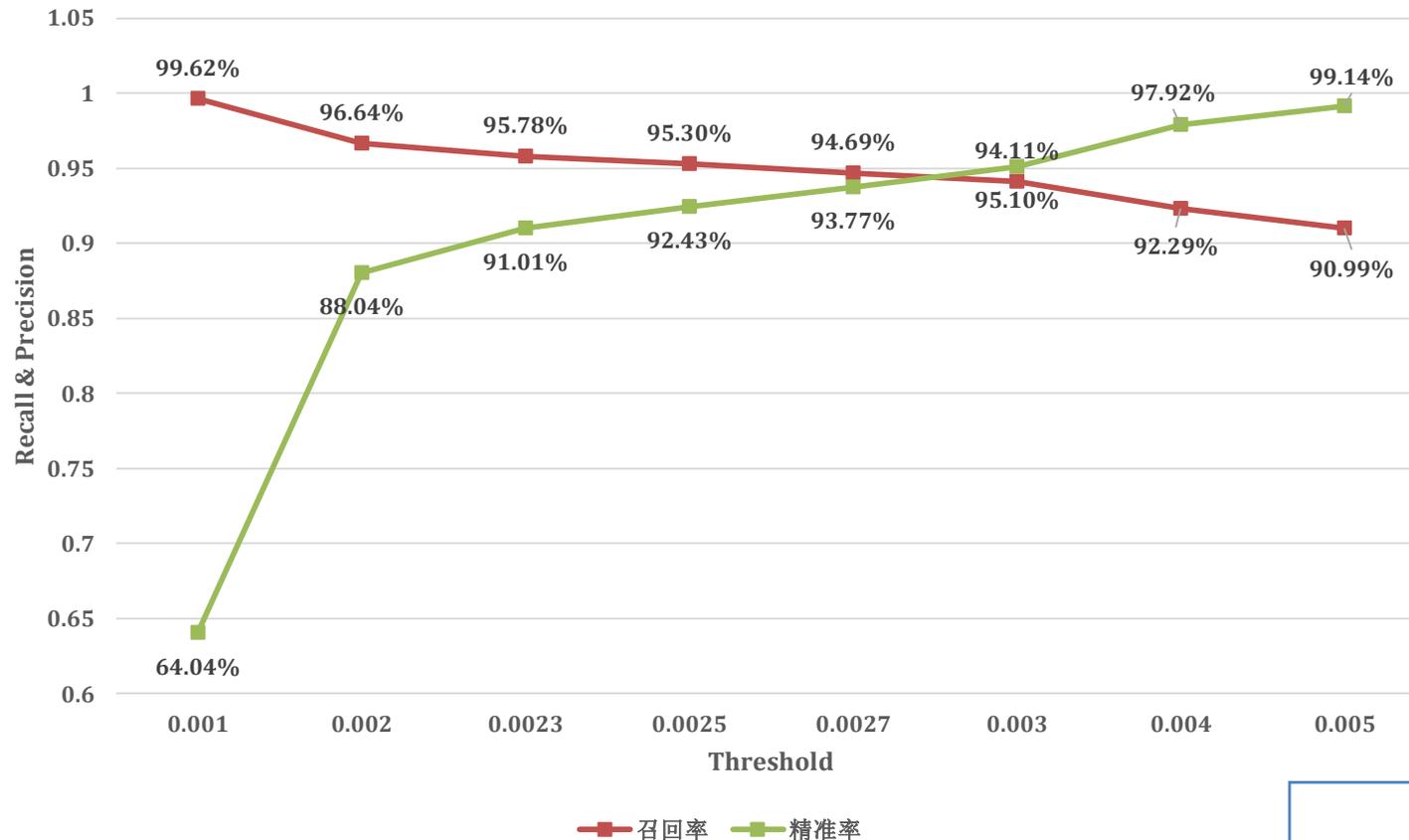


Encoder: $h = g\theta_1(x) = \sigma(W_1x + b_1)$
Decoder: $\hat{x} = g\theta_2(h) = \sigma(W_2x + b_2)$
Loss: $dist(X, X^R)$



SPMT直方图异常检测——模型效果

阈值影响评估



使用模型探测异常时间 (CPU) :

- 加载模型: 0.4348s
- 测试时间: 0.7963s
- 一共检测直方图数量: 62,276条数据样本

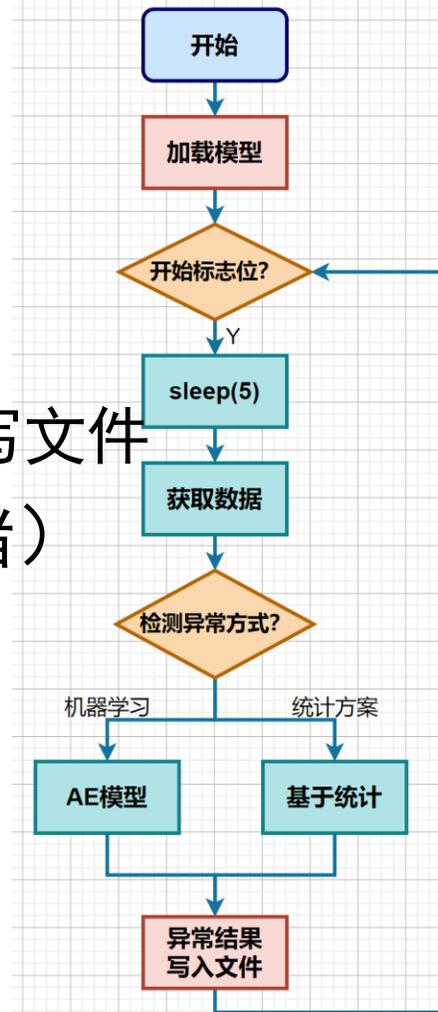
$$\text{召回率} = \frac{\text{预测正确的异常直方图个数}}{\text{实际异常直方图的个数}}$$

$$\text{精准率} = \frac{\text{预测正确的异常直方图个数}}{\text{预测出的异常直方图总数 (预测正确+错误)}}$$

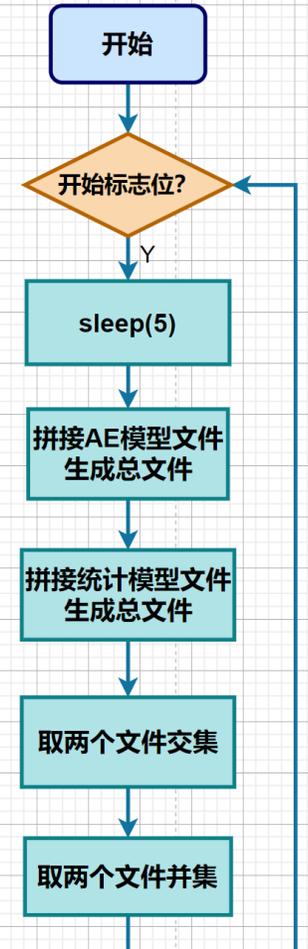
关灯测试下实际应用场景

- 使用python实现与数据流接口对接
- 复用数据流和ROBOT系统对接的接口格式
- **全流程跑通 (python)**
 - 加载模型, Redis中拿数据, 过模型, 得结果写文件
- 目前得到的文件格式 (以RUN号为单位存储)
 - 异常GCUID CHNID ERROR_CNTS TIME LOSS
- **统计+机器学习结合 (未应用)**
 - 取交集 & 取并集
 - 最终生成两个总文件

实现流程

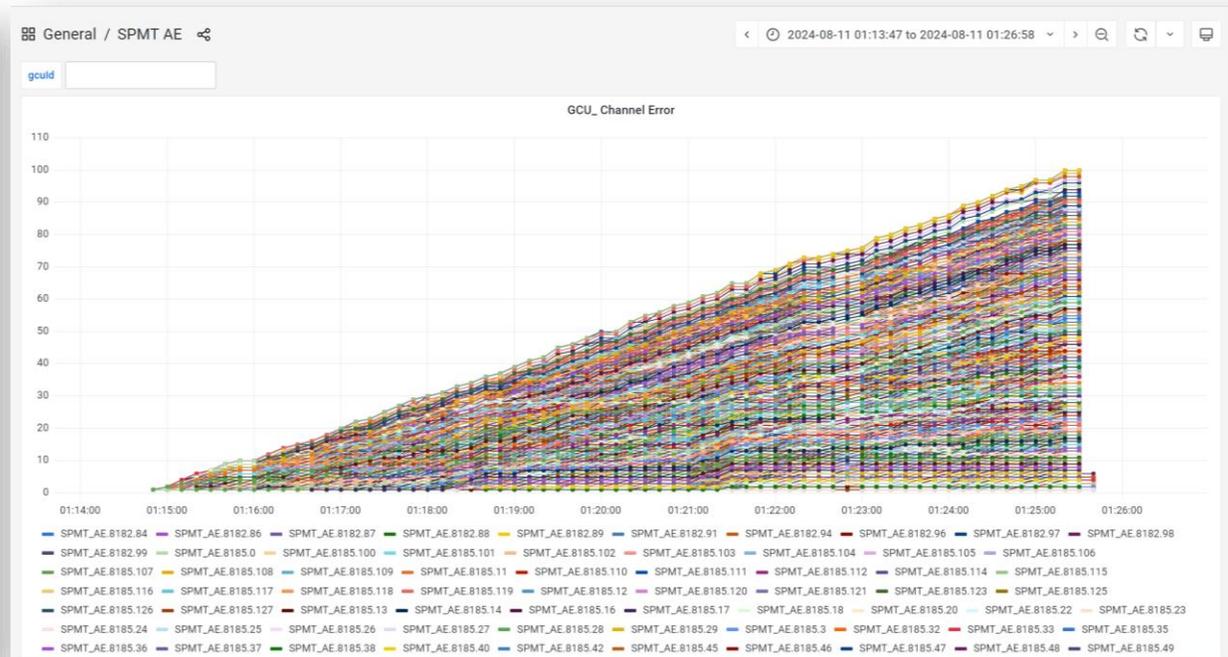
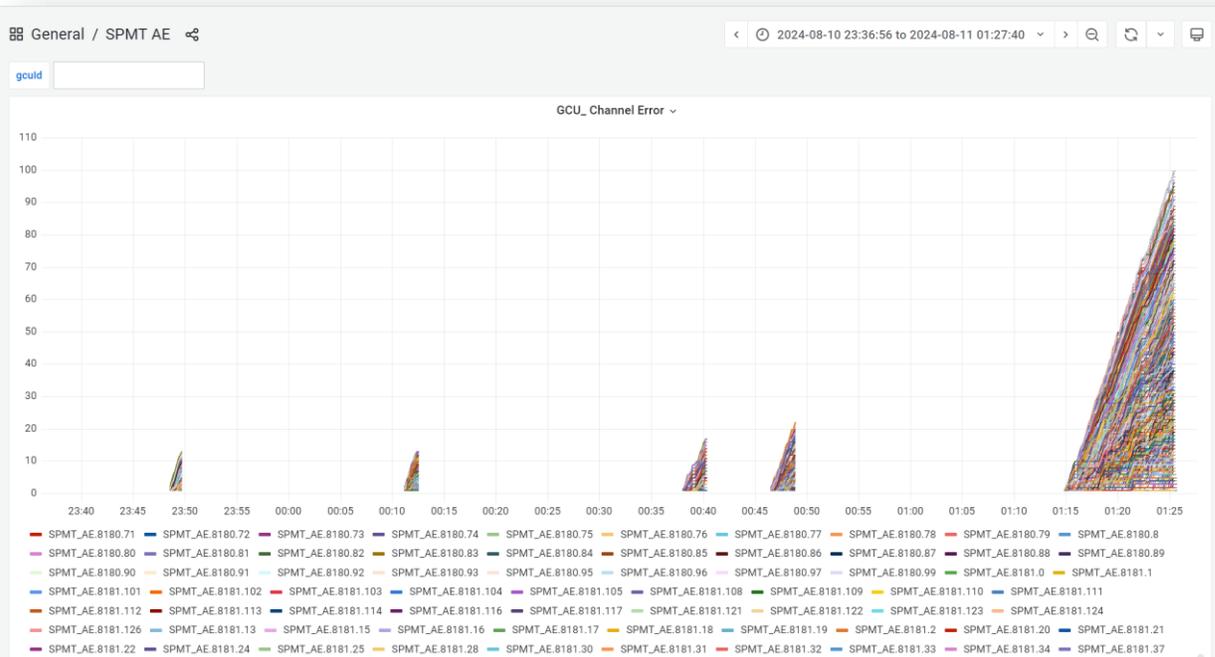


结合流程



第7次关灯测试下应用结果展示

□ Grafana实时展示异常总数（5s一次）



第7次关灯测试下应用结果展示

注：有些GCU未连PMT，人工检查未标记为异常，但统计时视为异常处理
另：人工检查异常总数=SPMT日志记载+裴文溪日志记载+未连PMT日志记载

	AE	人工
RUN562（取131套台阶）	只启了前2组，实时检验40个GCU，输出 $128*40=5120$ 个异常	取台阶，和训练数据不符，应该识别全部异常
RUN563（取两个GCU）	225个异常 (CHN58: 最大loss0.005440)	2个GCU均只连了cat3，记为224个异常
RUN564（取两个GCU）	224个异常	2个GCU均只连了cat3，记为224个异常
RUN570（新装的26个GCU）	1557个异常	790个异常（高压trip）
RUN571（同570，共3328个通道）	1010个异常	682个异常
RUN572（新装的19个GCU，2432个通道）	986个异常	577个异常
之后的RUN就是131套一起取（16768个通道）	9764个通道异常	人工未全部检查，只看了异常个例，太多了

结论：

- AE模型预测出异常总数：4002
- 人工检测出异常总数：1707
- 测试总通道数：9600

- 使用模型只需查看（4002/9600）~40%的通道
- 模型检测出来异常比例（1707/4002）~43%

THANKS

TDAQ组

张水涵，导师：季筱璐

2024/10/16

汇报人：陈博平