

# 钻井事故预测工作进展

2020-08-17

周壮

# 前期工作

- 标注数据分布对算法结果的影响分析 → **事故区间的标注并非越多越好**
- 样本不均衡下的评价指标 → **正确率、准确率并不适宜**
- 异常检测算法集成测试 → **基于局部特征进行模型选择的效果优于常规模型组合方法**
- 钻井应用场景的算法策略 → **✓ *Novelty Detection*  
? *Outlier Detection***

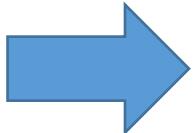
# 近期工作

- Novelty Detection策略的算法实验



**One classification method (半监督-正样本)**

OCSVM



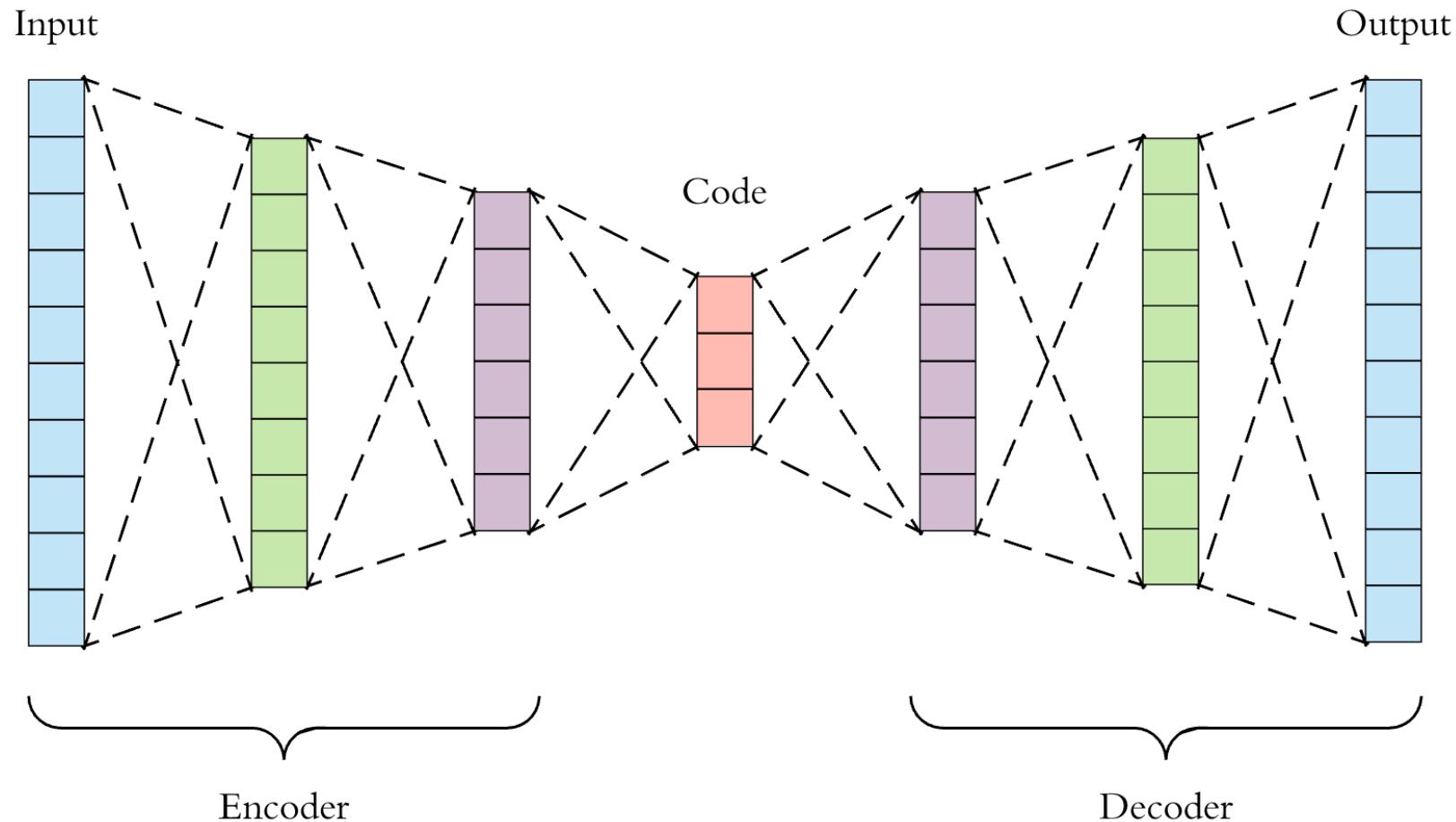
AutoEncoder

GAN, Generative Adversarial Networks

通过更多的非线性层级操作，  
学习并获取正常状态下录并  
检测参数更抽象、本质的特  
征，提高检测精度

# 近期工作

## 1. AutoEncoder

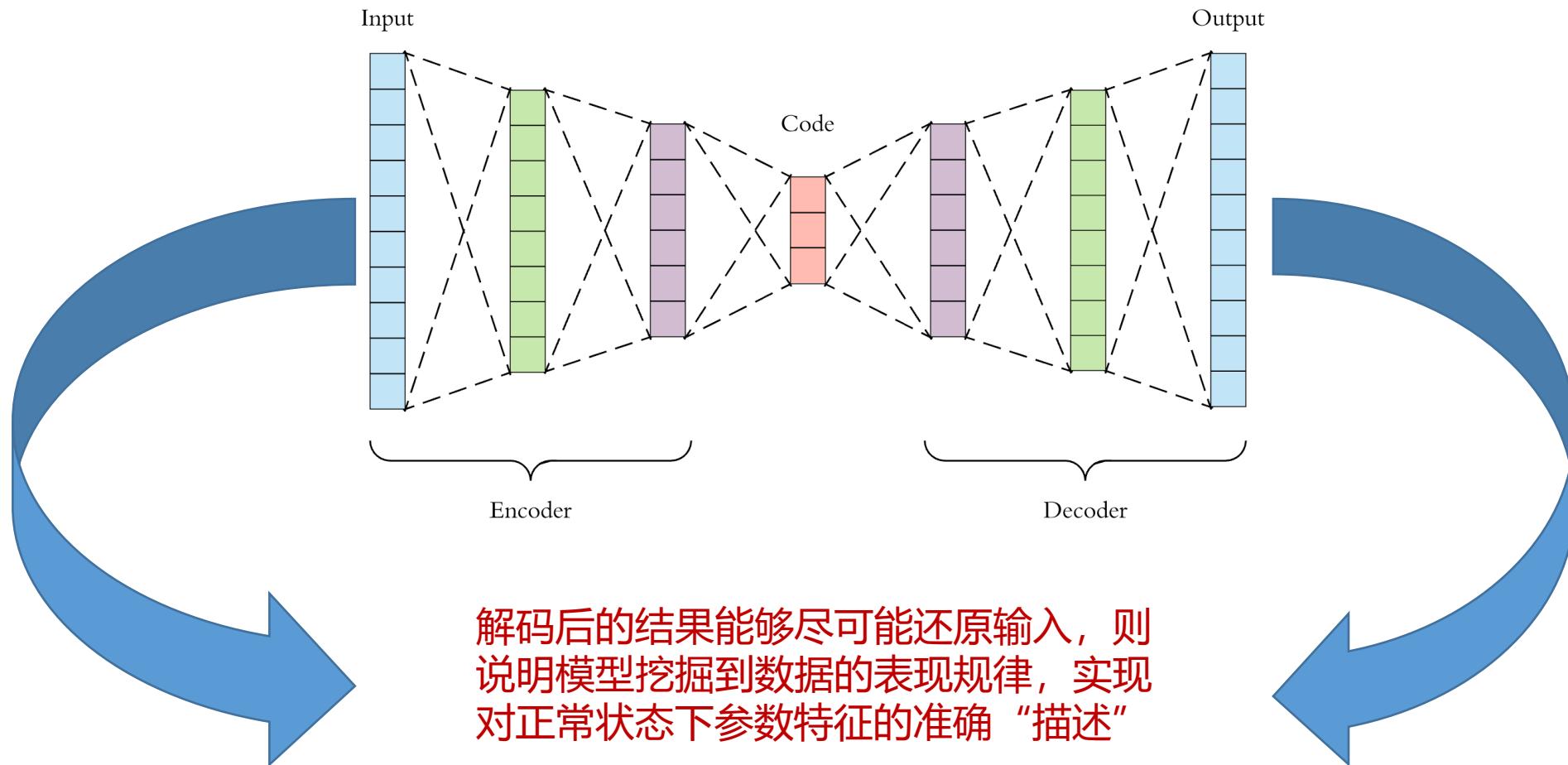


编码器：将输入参数压缩成潜在的空间表征

解码器：重构来自潜在空间表征的输入

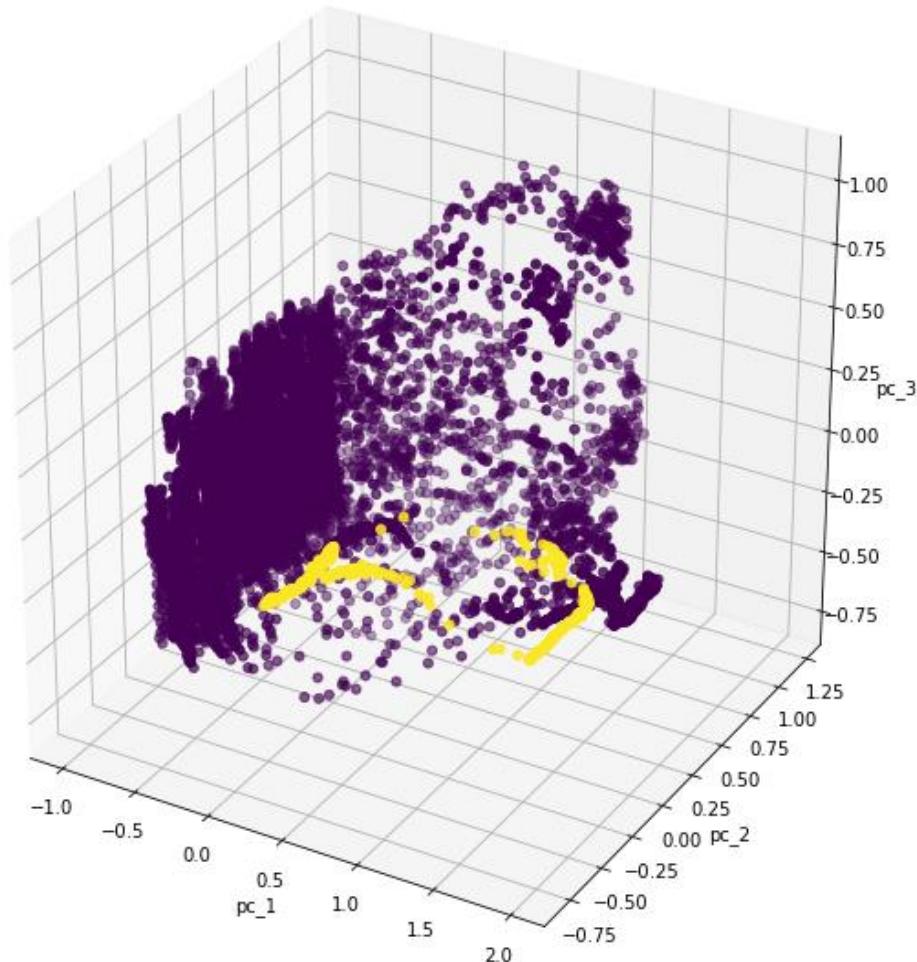
# 近期工作

## 1. AutoEncoder

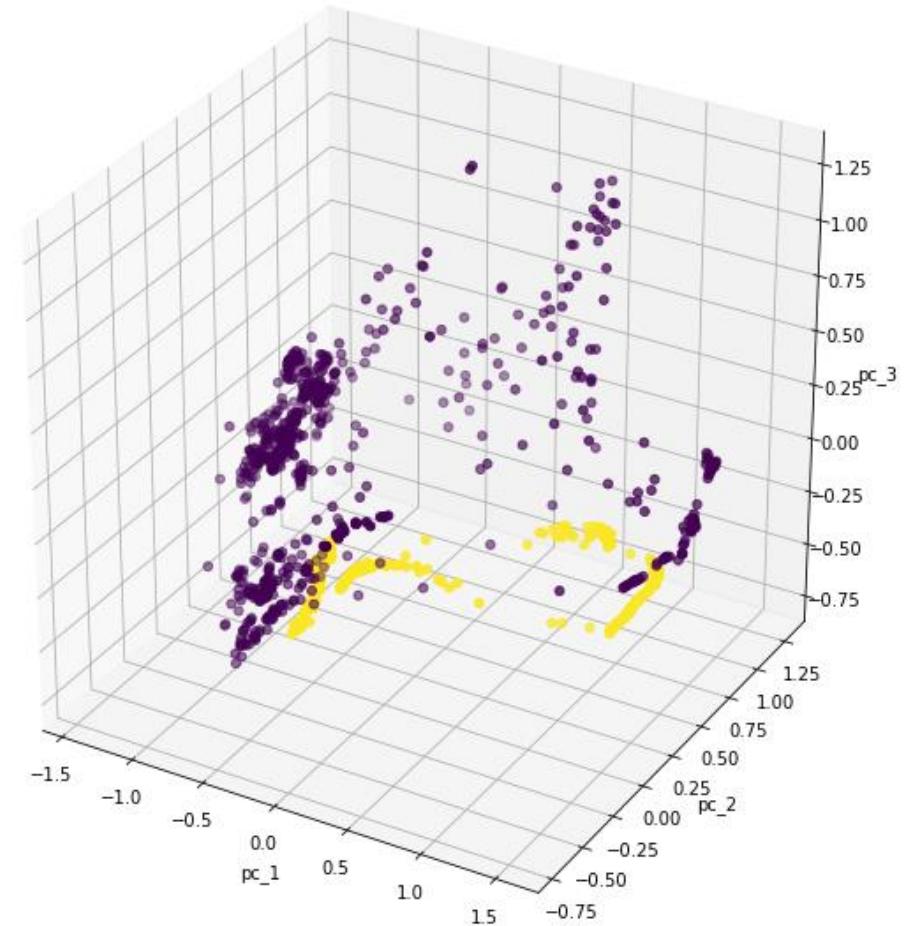
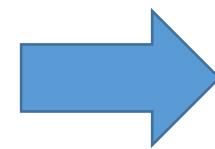


# 近期工作

## 1. AutoEncoder

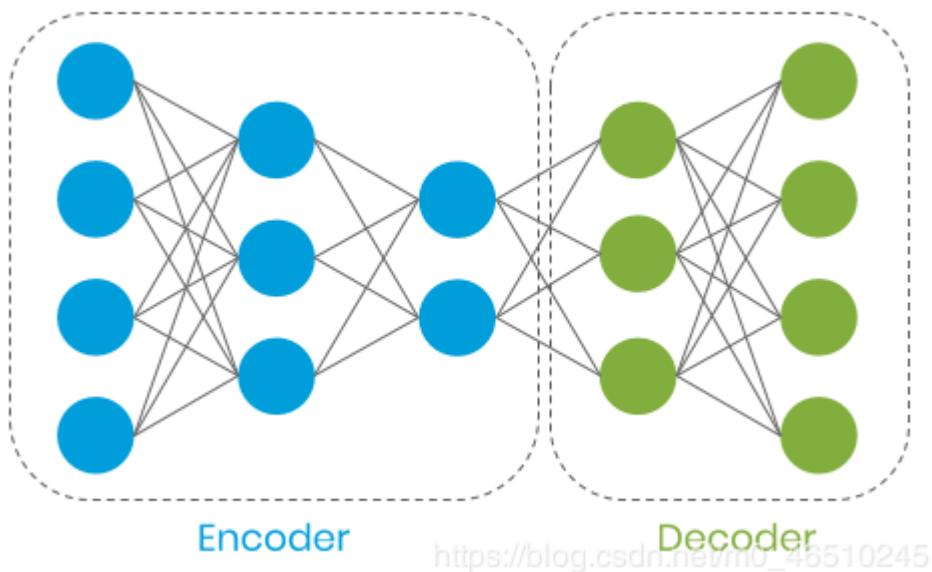


正常与事故状态下数据分布没有明显区分度!



# 近期工作

## 1. AutoEncoder

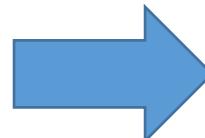


搭建网络：3层编码器和2层解码器  
输入：正常状态下19维录井参数

```
# model
encoder = models.Sequential(name='encoder')
encoder.add(layer=layers.Dense(units=20, activation=activations.relu, input_shape=[n_features]))
encoder.add(layers.Dropout(0.1))
encoder.add(layer=layers.Dense(units=10, activation=activations.relu))
encoder.add(layer=layers.Dense(units=5, activation=activations.relu))

decoder = models.Sequential(name='decoder')
decoder.add(layer=layers.Dense(units=10, activation=activations.relu, input_shape=[5]))
decoder.add(layer=layers.Dense(units=20, activation=activations.relu))
decoder.add(layers.Dropout(0.1))
decoder.add(layer=layers.Dense(units=n_features, activation=activations.sigmoid))

autoencoder = models.Sequential([encoder, decoder])
```

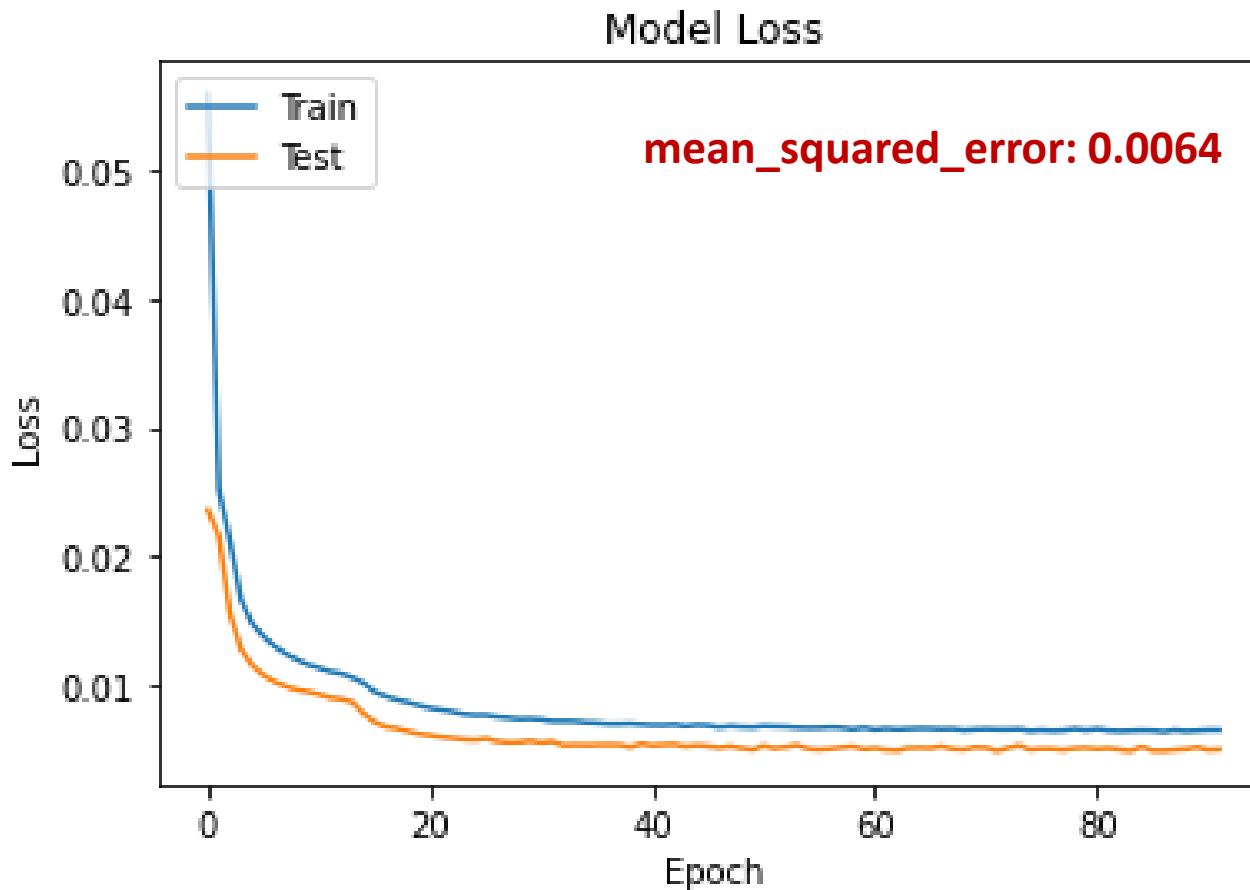


经过多次迭代训练后，模型能够还原输入参数

# 近期工作

## 1. AutoEncoder

模型评价：求解输入与输出的均方误差（MSE）

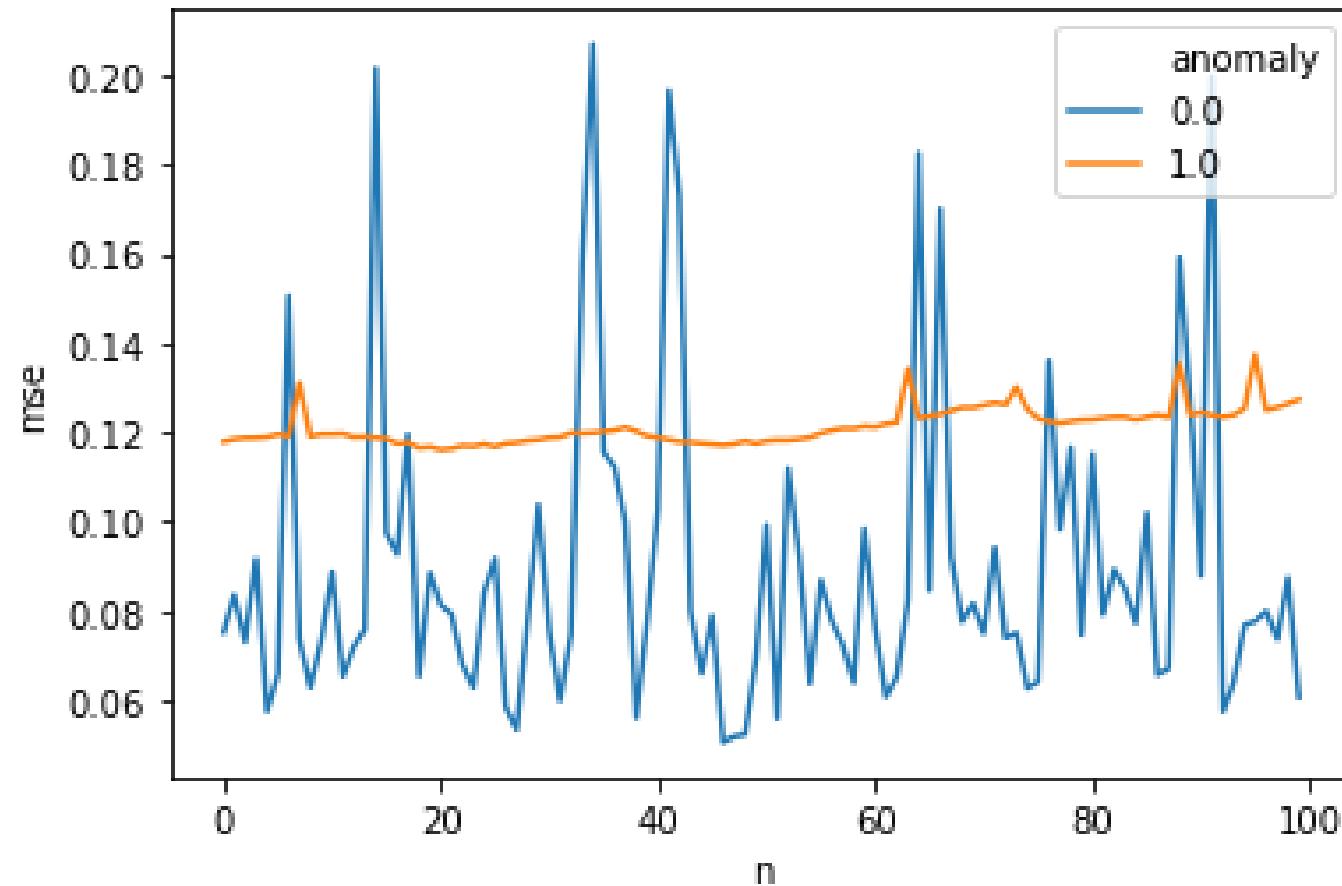


MSE越小，则输入数据和输出数据的相似度越高！

# 近期工作

## 1. AutoEncoder

抽样对比分析：

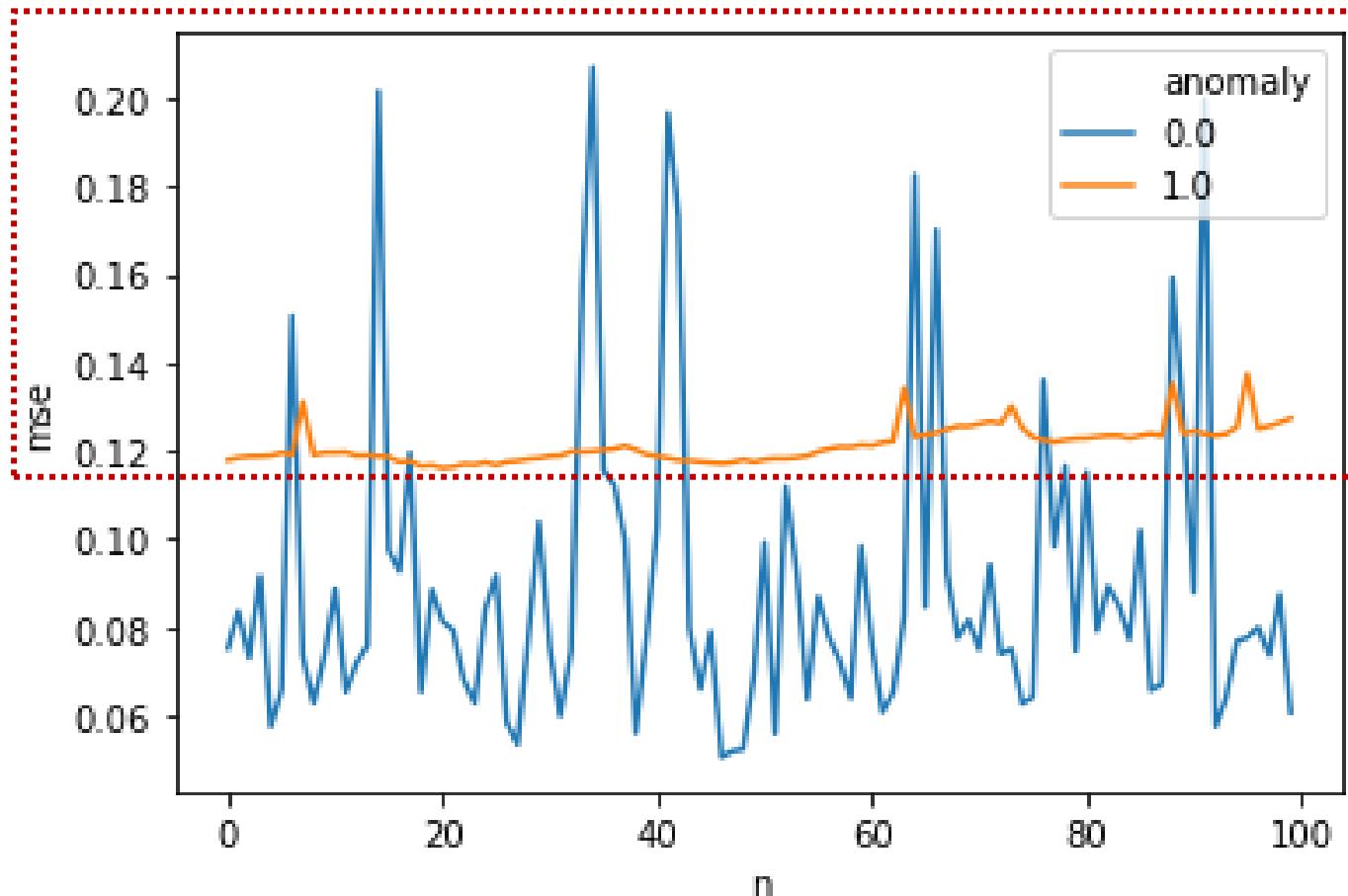


训练后的模型对多数正常状态数据还原度较高，异常数据的还原度较差，其MSE>0.12

# 近期工作

## 1. AutoEncoder

抽样对比分析：



训练后的模型对多数正常状态数据还原度较高，异常数据的还原度较差，其MSE>0.12

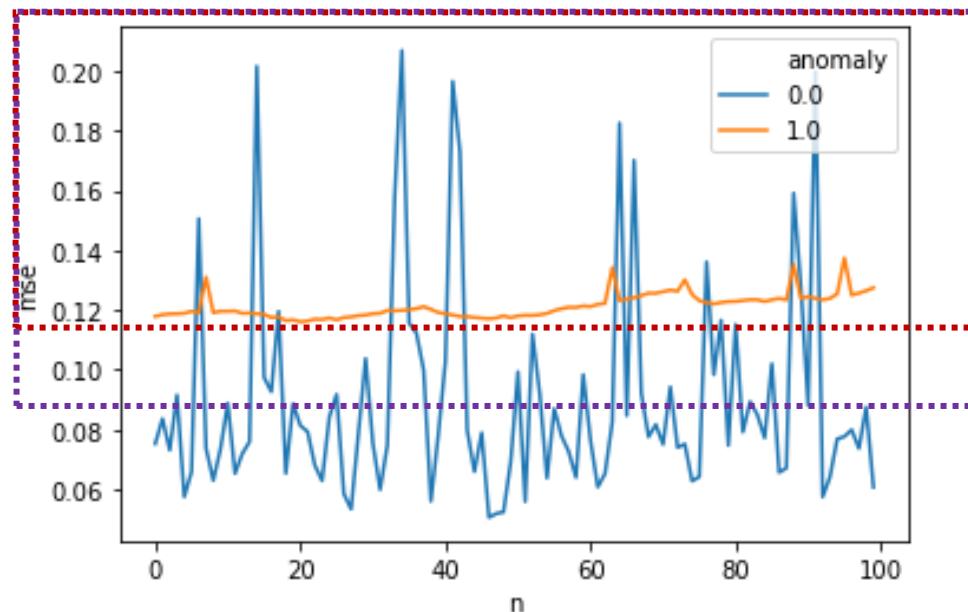


以MSE=0.12为阈值，可以过滤所有异常值及9%的正常值

# 近期工作

## 1. AutoEncoder

问题分析：



如何获取阈值？



1. 以事故下的MSE为阈值

最优&监督

2. 以正常状态下MSE均值为阈值

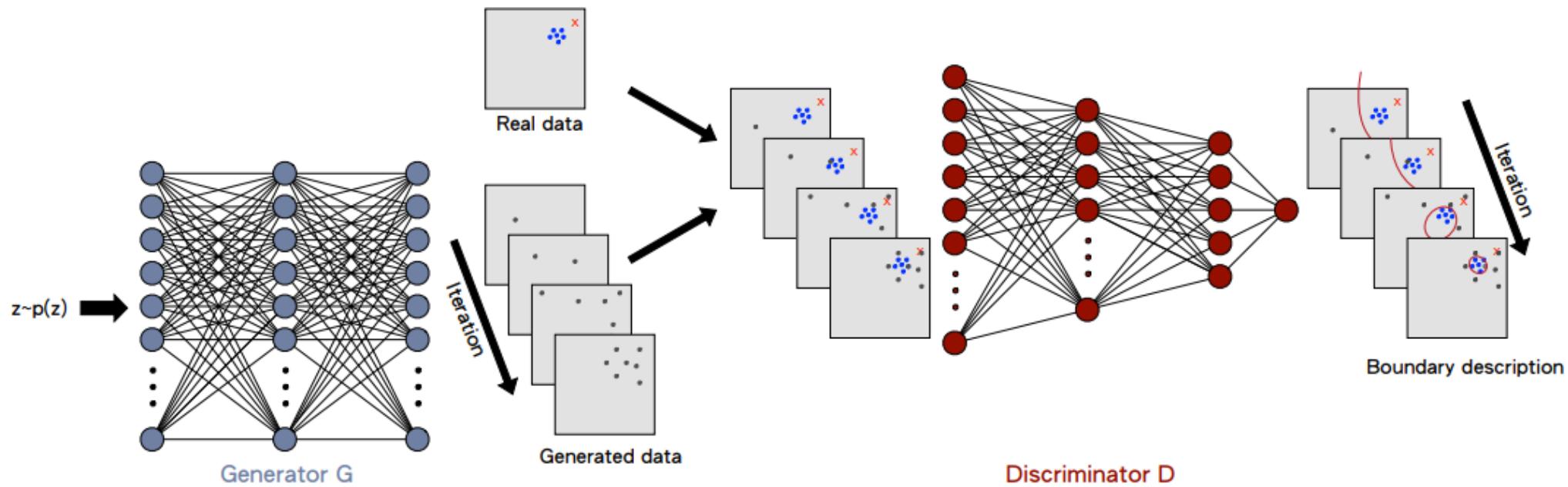
非监督&&低精度

Accuracy: 94% → 83%

Precision: 89% → 75%

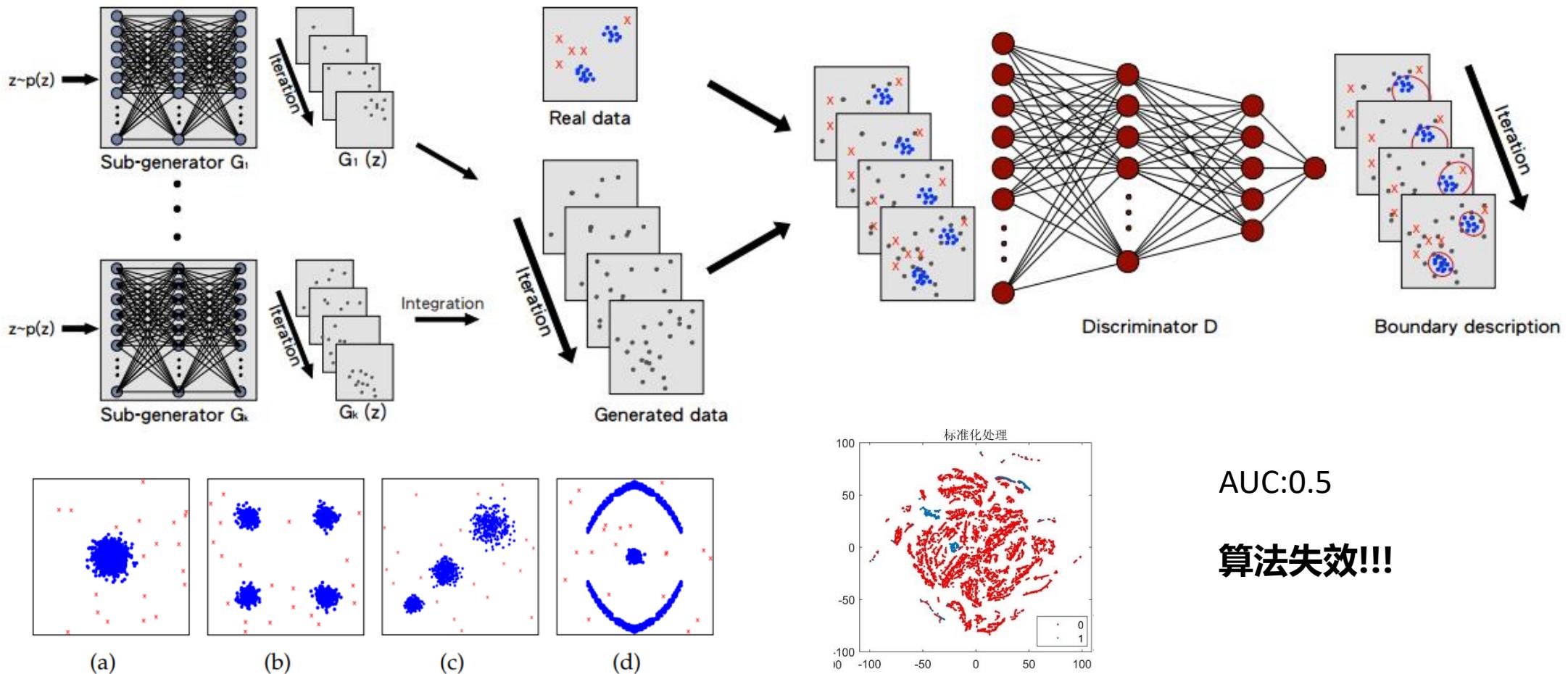
# 近期工作

## 2. Generative Adversarial Active Learning for Unsupervised Outlier Detection



# 近期工作

## 2. Generative Adversarial Active Learning for Unsupervised Outlier Detection



# 下一步工作

1. 完善Auto-encoder算法实验
2. 对比Auto-encoder与OCSVM
3. 调研GAN网络并开展实验



# 工作总结汇报

赵子飞

2020年8月18日



中国科学院空间应用工程与技术中心  
Technology And Engineering Center For Space Utilization  
Chinese Academy of Sciences



# 1. 钻井事故预警

## 1.1 HTM

### ➤ 技术路线

- ①数据集构建、②时间序列分解/征兆因子筛选、③预测模型优选、④异常检测、⑤事故预警

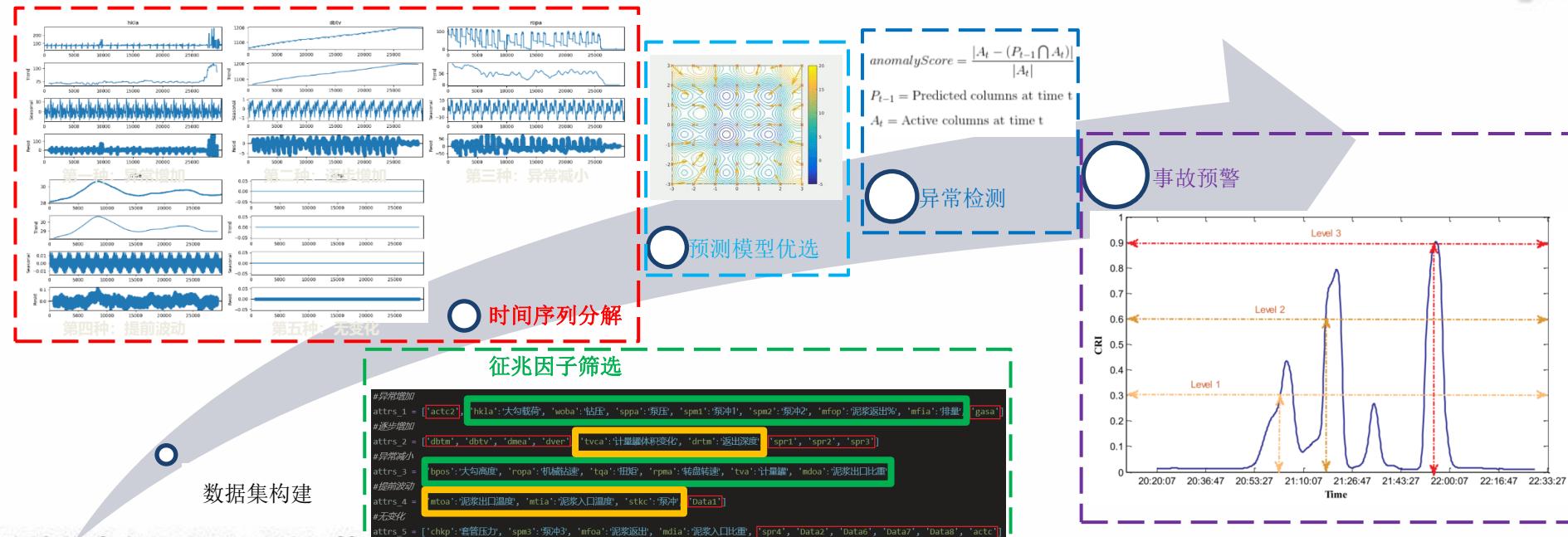


图1 技术路线



# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故预警示例

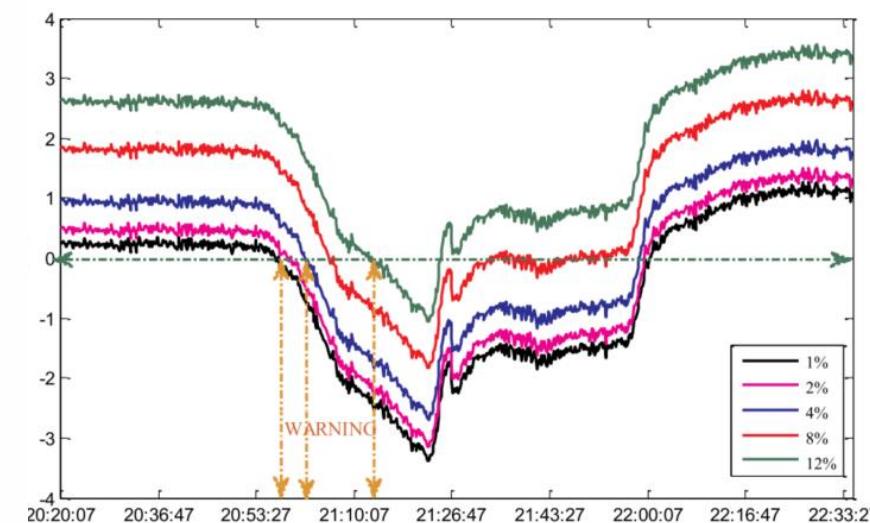


图1 单征兆因子的异常检测

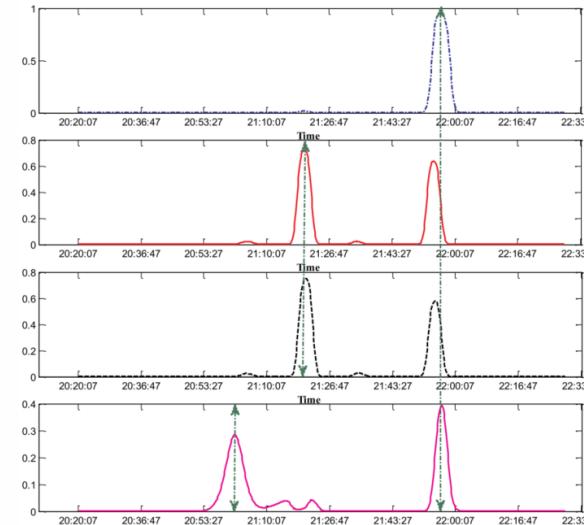


图2 各主要征兆因子的异常告警得分

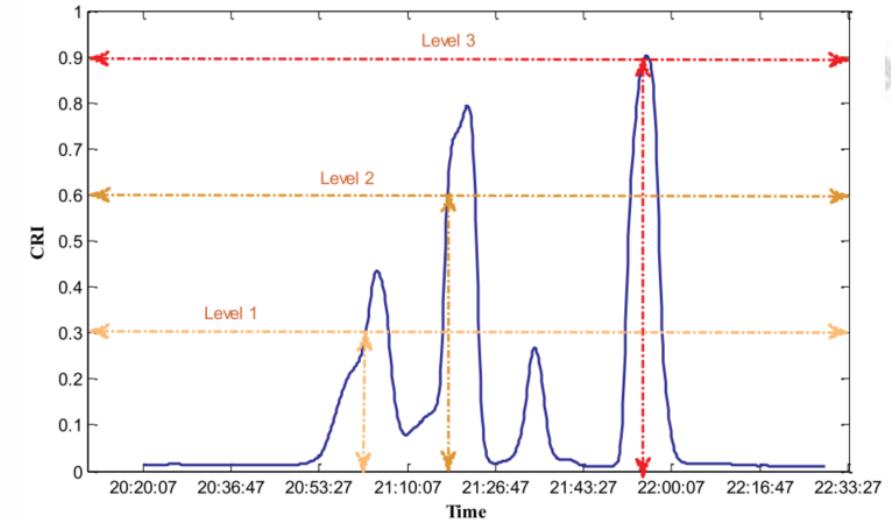


图3 基于CRI指数的事故风险预警





# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况统计

井名	工况	开始日期	开始时间	结束日期	结束时间	事故类型	事故时间	
WZ12-2-B32	湿起下(划眼)	2019/10/25	20:25:00	2019/10/25	21:46:00	★ 卡钻	2019/10/25	20:25:00
WZ12-2-B33	钻进	2019/12/2	0:00:00	2019/12/4	4:26:00	★ 卡钻	2019/12/4	4:26:00
WZ11-ZE-14dSa	湿起下	2019/2/15	21:08:00	2019/2/17	5:47:00	★ 下钻	2019/2/17	5:47:00
WZ11-2E-14dSa	湿起下	2019/2/18	0:32:00	2019/2/18	19:47:00	卡钻	2019/2/18	19:47:00
WC13-6-A18H	钻进	2019/12/4	23:46:00	2019/12/5	3:36:00	卡钻	2019/12/5	3:36:00
WC13-6-A18H	湿起下	2019/12/6	21:30:00	2019/12/7	5:00:00	卡钻	2019/12/7	5:00:00
WZ6-9-A19	干起下	2019/12/1	17:02:00	2019/12/1	22:10:00	卡钻	2019/12/1	22:10:00
DF13-2-A4H	湿起下(划眼)	2019/2/20	5:33:00	2019/2/20	5:45:00	卡钻	2019/2/20	5:45:00
DF1-1-P4H	钻进	2019/12/15	1:00:00	2019/12/16	2:46:00	★ 卡钻	2019/12/16	2:46:00
DF1-1-P10H	钻进	2019/12/22	4:00:00	2019/12/23	0:51:00	★ 卡钻	2019/12/23	0:51:00
DF1-1-P10H	湿起下(划眼)	2020/1/2	17:00:00	2020/1/3	3:20:00	卡钻	2020/1/3	3:20:00
DF1-1-P11H	钻进	2020/1/16	11:15:00	2020/1/18	5:46:00	★ 卡钻	2020/1/18	5:46:00
DF13-2-B9	湿起下(划眼)	2019/9/7	23:42:00	2019/9/7	23:55:00	卡钻	2019/9/7	23:55:00
DF13-2-B16H	湿起下	2019/10/30	0:00:00	2019/10/30	0:31:00	卡钻	2019/10/30	0:31:00
WZ12-2-B39H	下套管	2019/8/25	11:00:00	2019/8/25	20:30:00	卡钻	2019/8/25	20:30:00
WS23-5-3d	钻进	2020/1/25	15:20:00	2020/1/25	20:16:00	井漏	2020/1/25	20:16:00
WZ12-2-B32	湿起下	2019/10/29	6:50:00	2019/10/29	8:33:00	井漏	2019/10/29	8:33:00
DF13-2-B1H	钻进	2019/7/27	8:35:00	2019/7/27	10:07:00	井漏	2019/7/27	10:07:00
DF13-2-B1H	钻进	2019/7/29	20:58:00	2019/7/30	1:19:00	井漏	2019/7/30	1:19:00
YL8-3-1	钻进	2019/7/17	10:00:00	2019/7/17	16:18:00	井漏	2019/7/17	16:18:00
DF4-1-1	钻进	2019/3/29	0:00:00	2019/3/29	5:12:00	井漏	2019/3/29	5:12:00
DF4-1-1	钻进	2019/4/3	0:30:00	2019/4/3	5:06:00	井漏	2019/4/3	5:06:00
DF13-1-12	循环	2019/8/13	5:00:00	2019/8/13	10:23:00	井漏	2019/8/13	10:23:00
LD10-1-12	循环	2019/7/19	17:26:00	2019/7/20	0:03:00	井漏	2019/7/20	0:03:00
LD10-1-12	循环	2019/7/27	12:15:00	2019/7/27	13:18:00	井漏	2019/7/27	13:18:00
DF17-2-1	钻进	2020/1/9	0:00:00	2020/1/9	8:43:00	井漏	2020/1/9	8:43:00
DF17-2-1	循环	2020/1/12	20:27:00	2020/1/13	22:07:00	井漏	2020/1/13	22:07:00
DF17-2-1	干起下	2020/2/1	13:30:00	2020/2/1	13:45:00	井涌	2020/2/1	13:45:00

工况/事故	卡钻	井漏	井涌	合计
钻进	5	7	0	12
循环	0	4	0	4
下套管	1	0	0	1
干起下	1	0	1	2
湿起下	4	1	0	5
湿起下(划眼)	4	0	0	4
固井	0	0	0	0
电测	0	0	0	0
合计	15	12	1	28



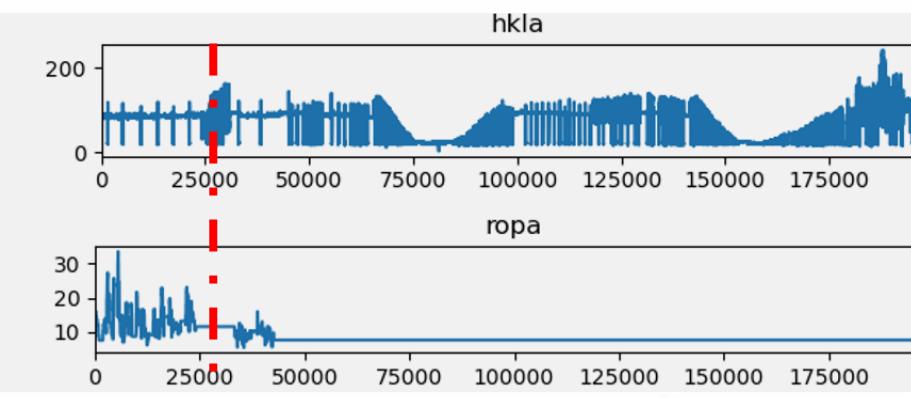
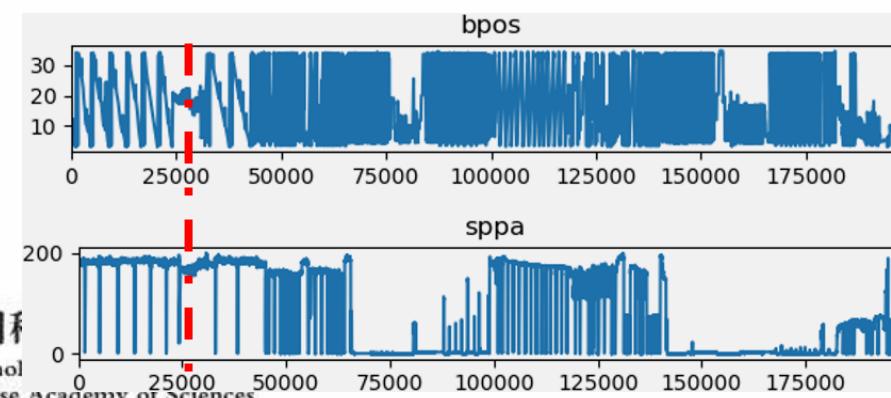


# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B33

```
#异常增加
attrs_1 = ['actc2', 'hkla': '大勾载荷', 'woba': '钻压', 'sppa': '泵压', 'spm1': '泵冲1', 'spm2': '泵冲2', 'mfpop': '泥浆返出%', 'mfia': '排量', 'gasa']
#逐步增加
attrs_2 = ['dbtm', 'dbtv', 'dmea', 'dver', 'tvca': '计量罐体积变化', 'drtm': '返出深度', 'spr1', 'spr2', 'spr3']
#异常减小
attrs_3 = ['bpos': '大勾高度', 'ropaa': '机械钻速', 'tqa': '扭矩', 'rpma': '转盘转速', 'tva': '计量罐', 'mdoa': '泥浆出口比重']
#提前波动
attrs_4 = ['mtoa': '泥浆出口温度', 'mtia': '泥浆入口温度', 'stkc': '泵冲', 'Data1']
#无变化
attrs_5 = ['chkp': '套管压力', 'spmb': '泵冲3', 'mfoa': '泥浆返出', 'mdia': '泥浆入口比重', 'spr4', 'Data2', 'Data6', 'Data7', 'Data8', 'actc']
```



中国科

技

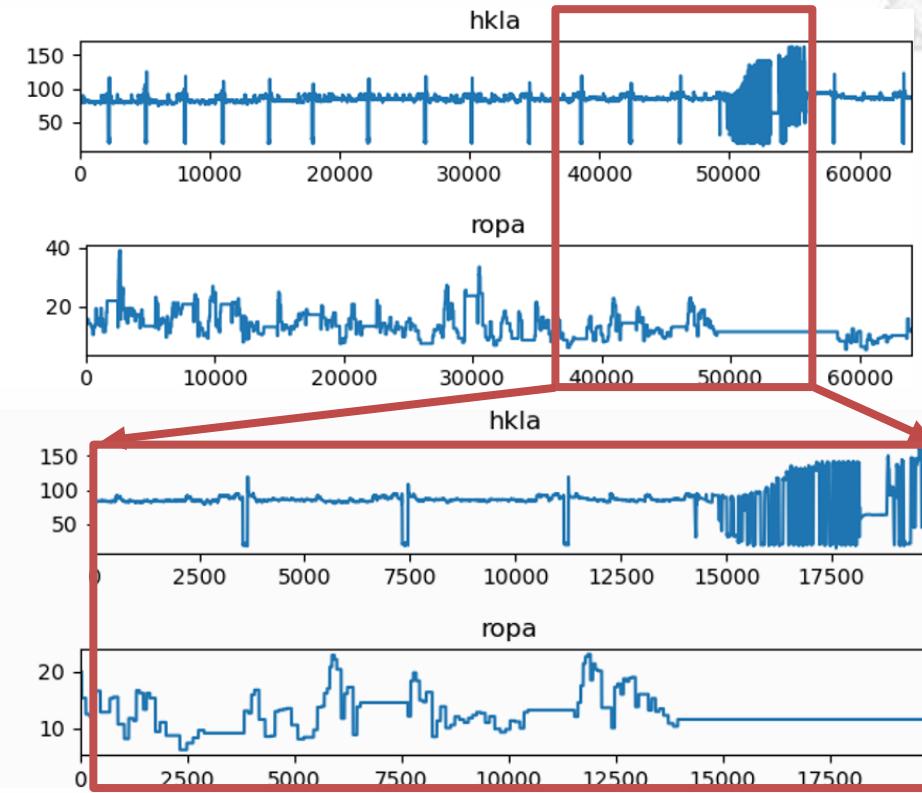
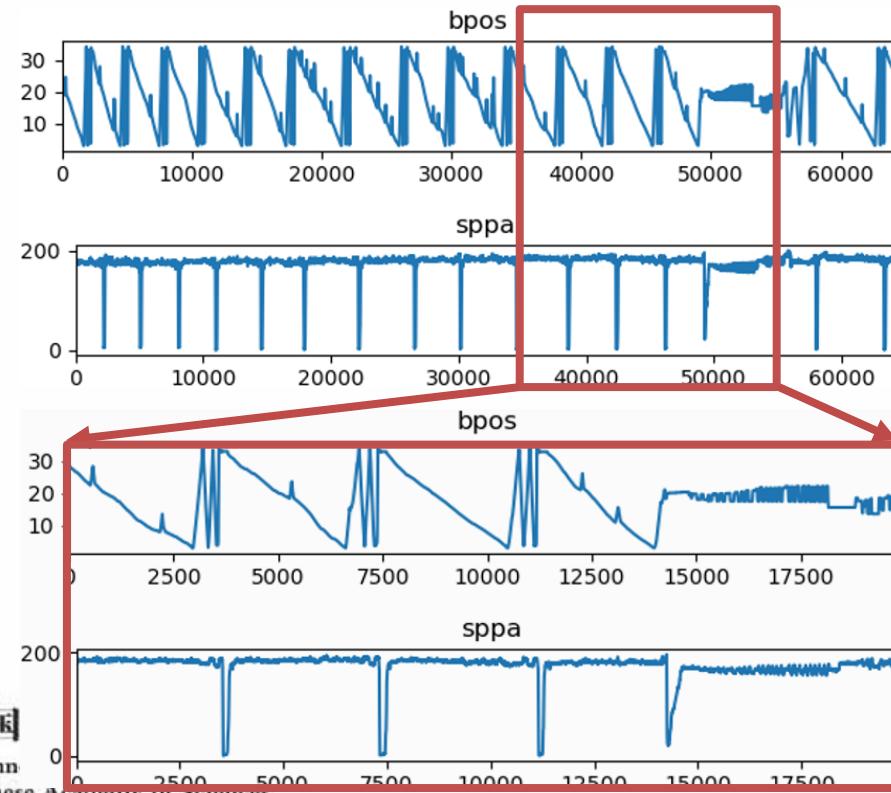
Technological  
Chinese Academy of Sciences



# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B33



中国

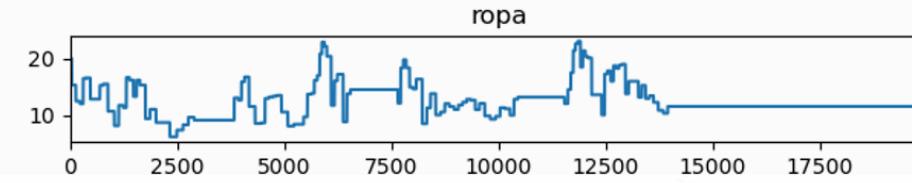
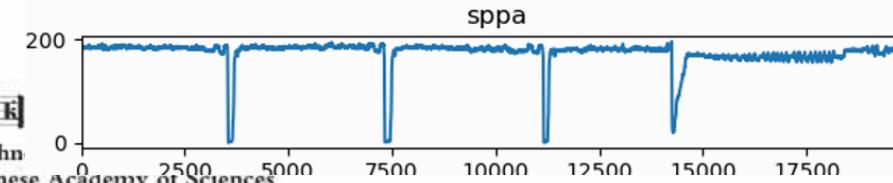
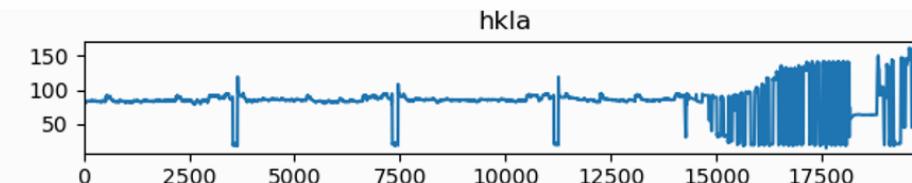
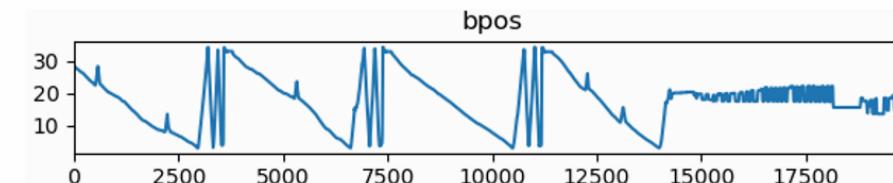
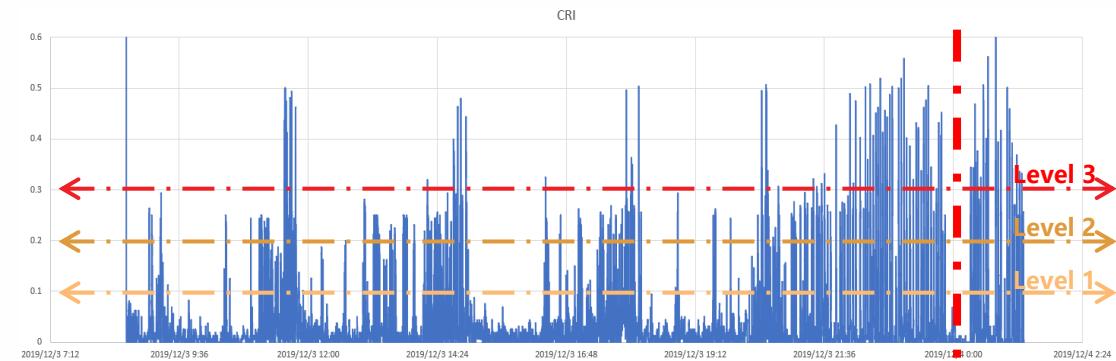
Techn  
Chinese Academy of Sciences



# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B32、B33

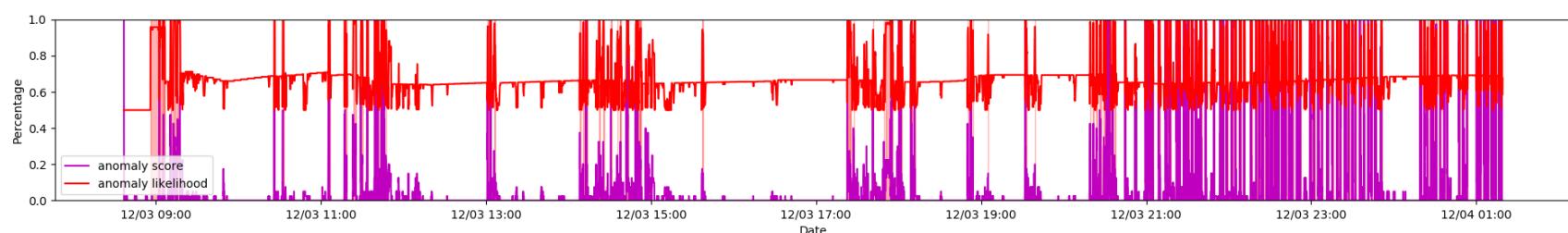
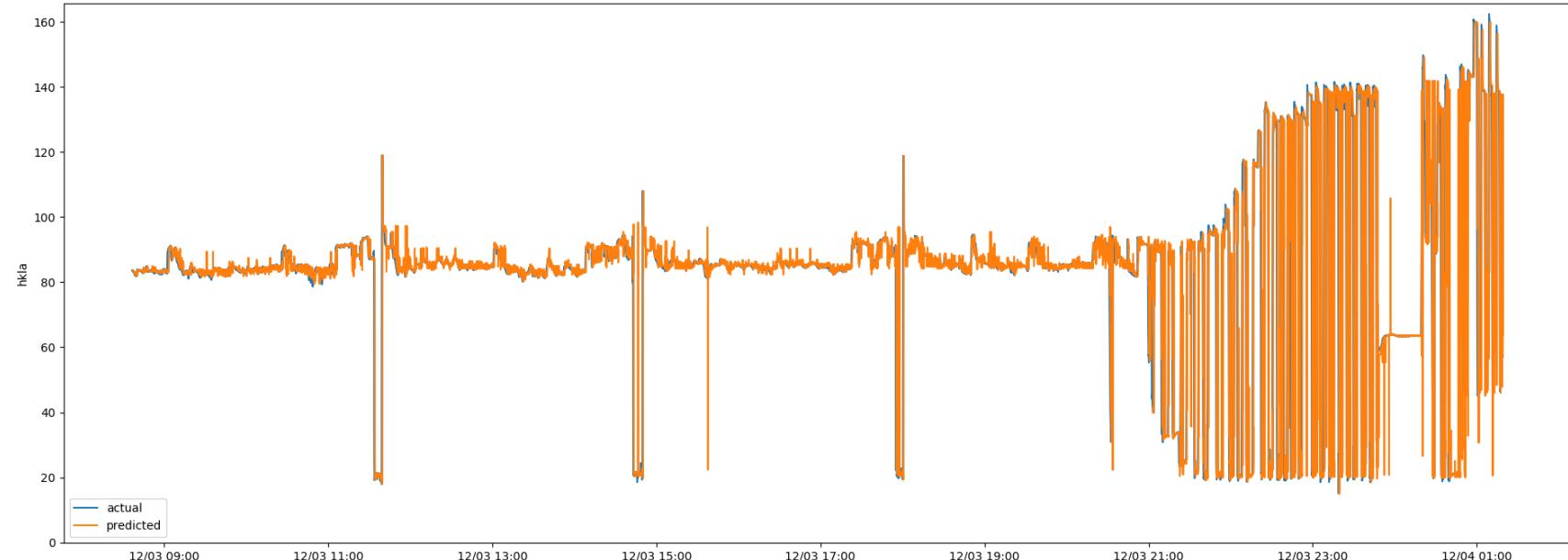
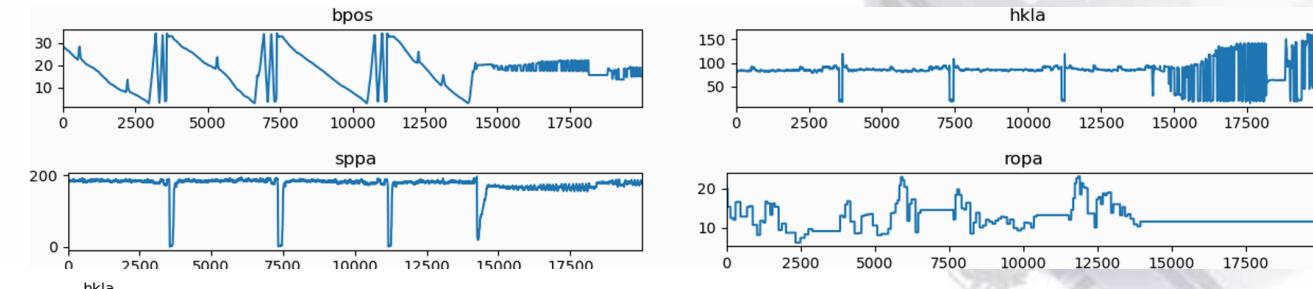




# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B33



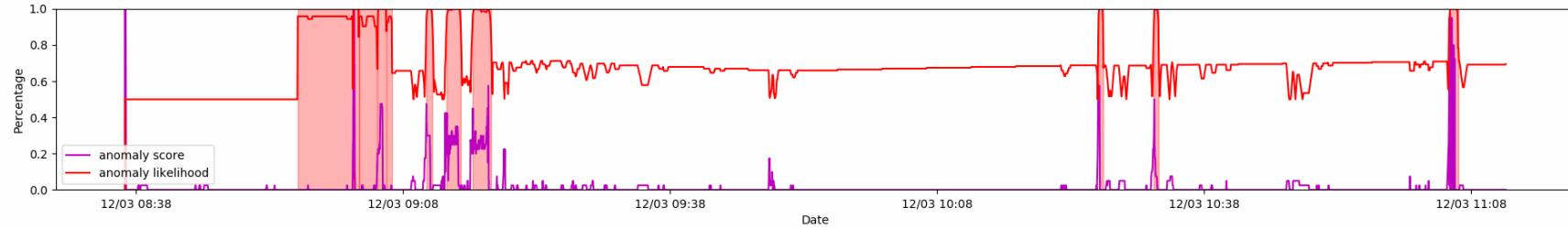
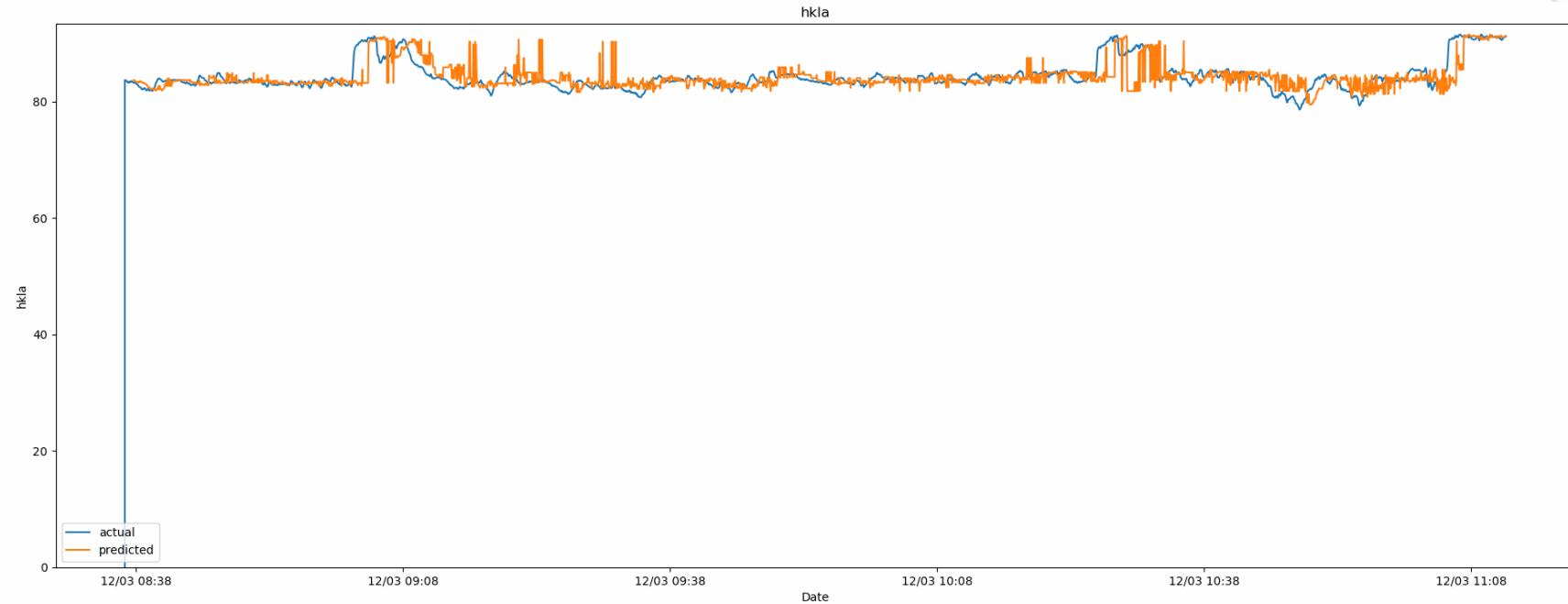
中国科学  
Technology A  
Chinese Academ



# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B33



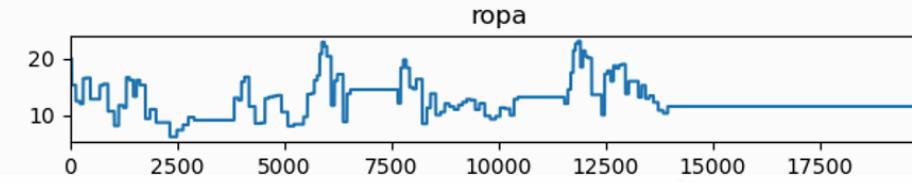
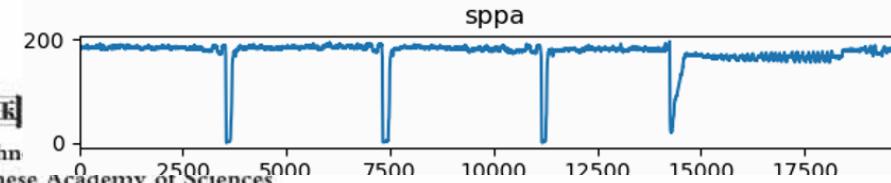
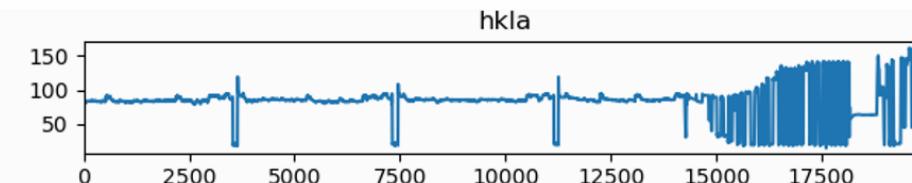
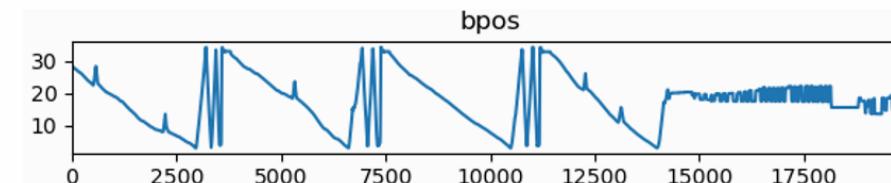
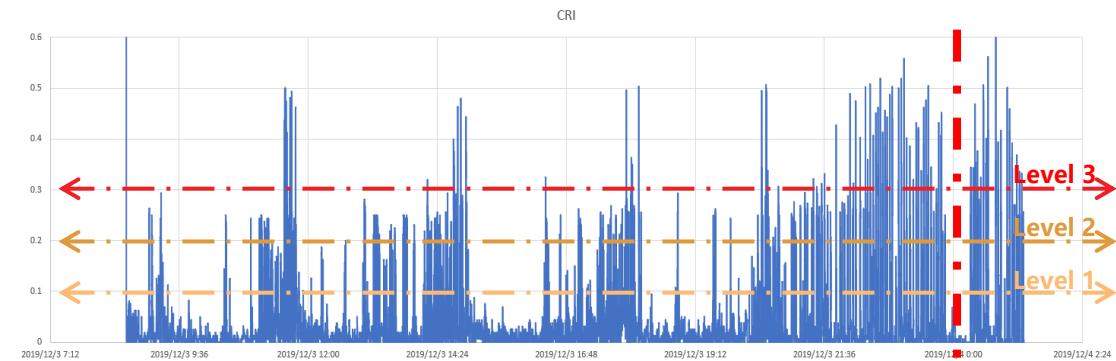
中国科学  
Technology A  
Chinese Academ



## 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B32、B33



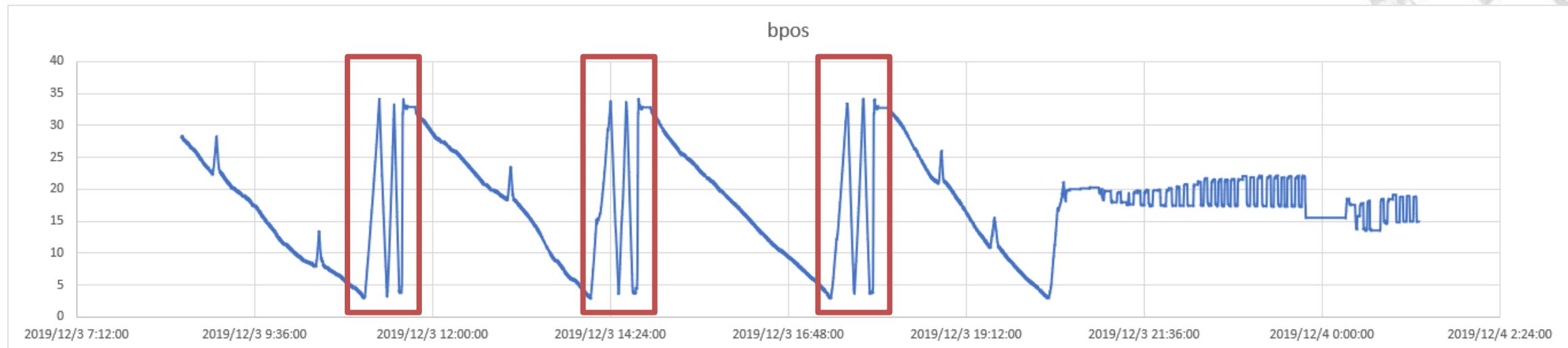


# 1. 钻井事故预警



## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B33

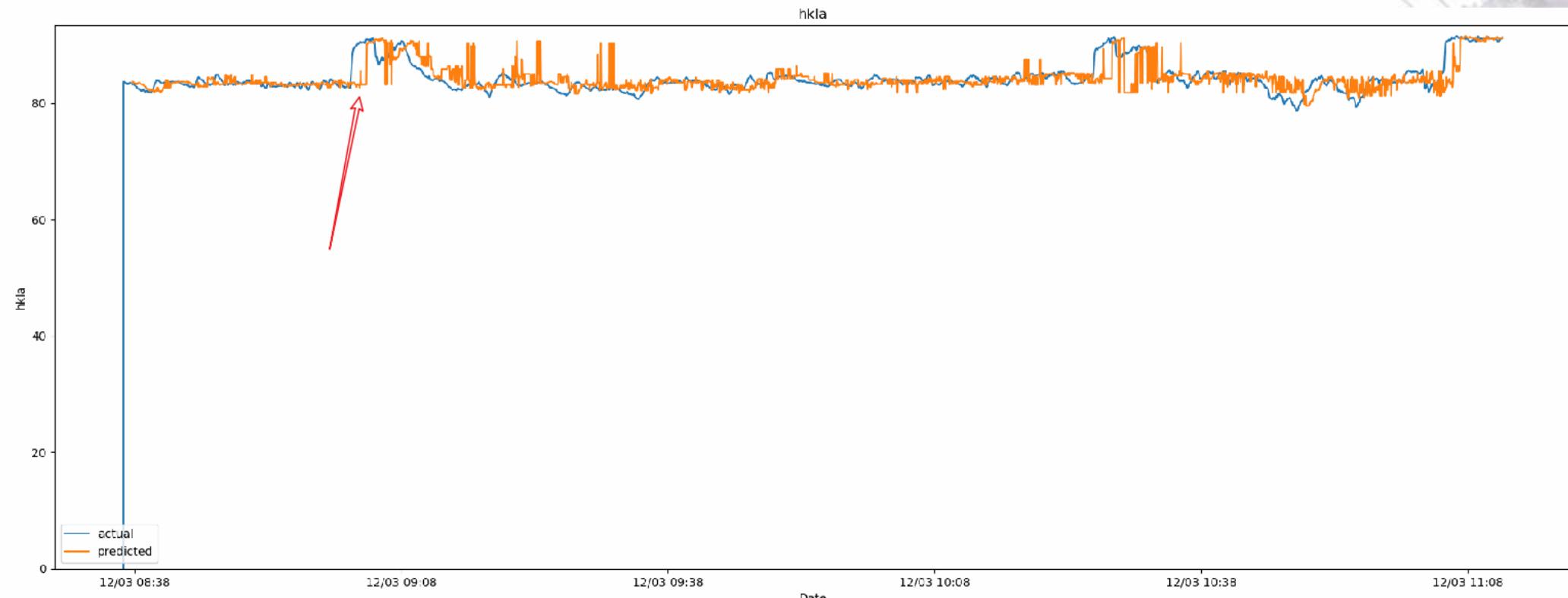




# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B33

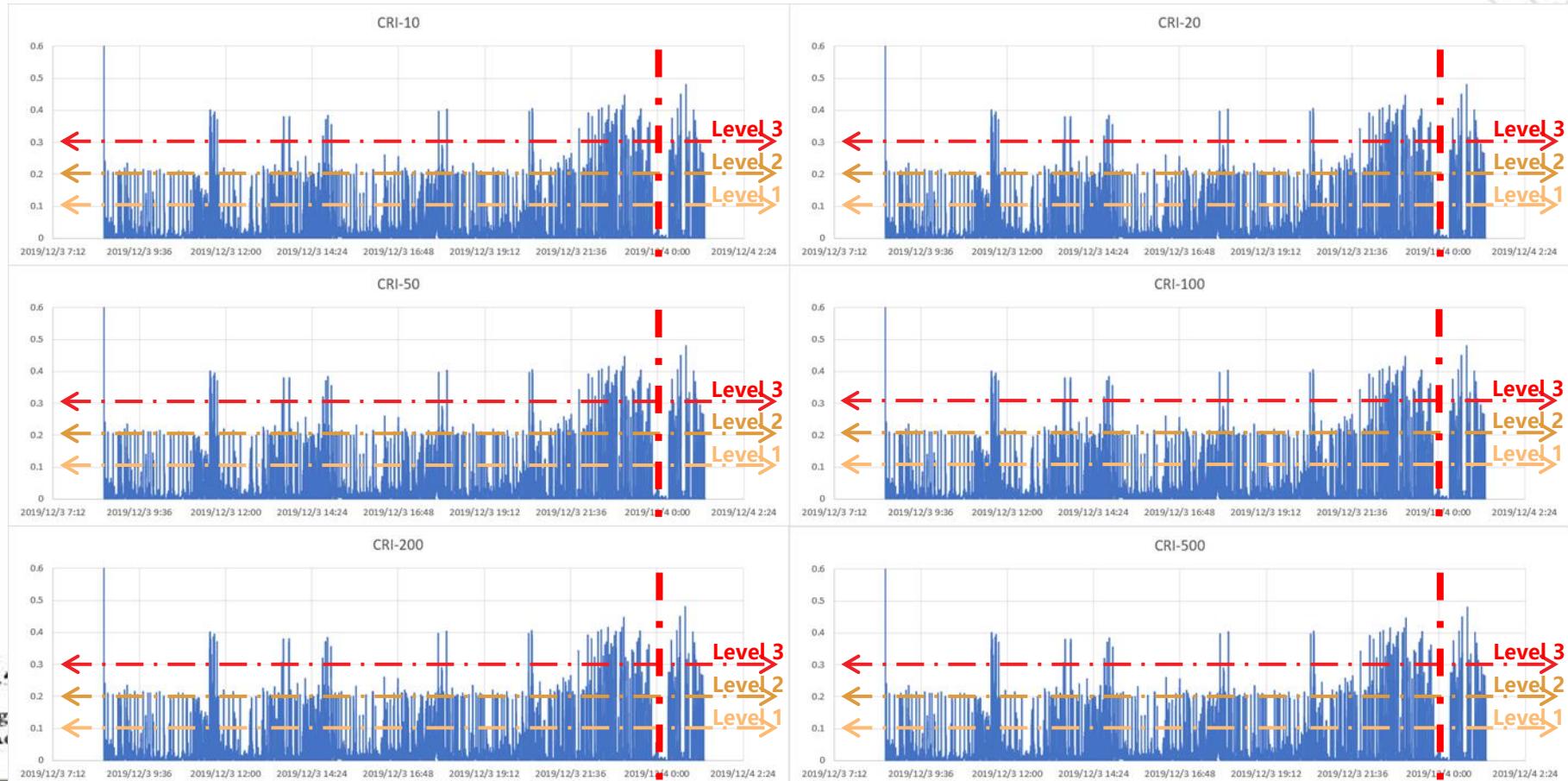




# 1. 钻井事故预警

## 1.3 事故预警系统

- 事故所处工况测试——WZ12-2-B33



中国科  
技学  
院

Technolog  
Chinese Ac



# 下一步计划

1. 异常预警：借助实现的风险评估模型，进一步测试并验证其在其他井的预警效果，积累事故预警模型的井间迁移经验。





谢谢！



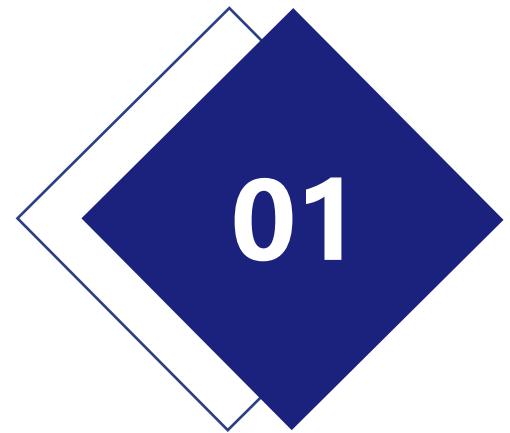
中国科学院空间应用工程与技术中心  
Technology And Engineering Center For Space Utilization  
Chinese Academy of Sciences

# 油田钻井大数据项目进展汇报

---

汇报人：熊雄

2020.08.17



# 前期工作

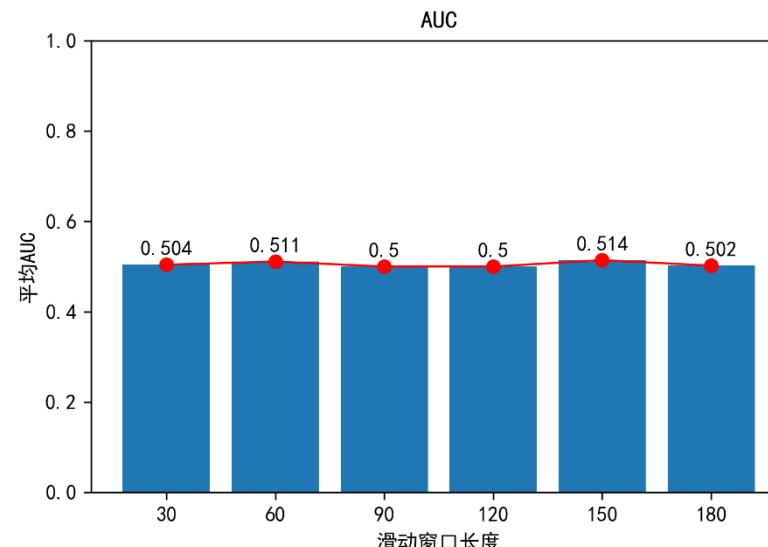
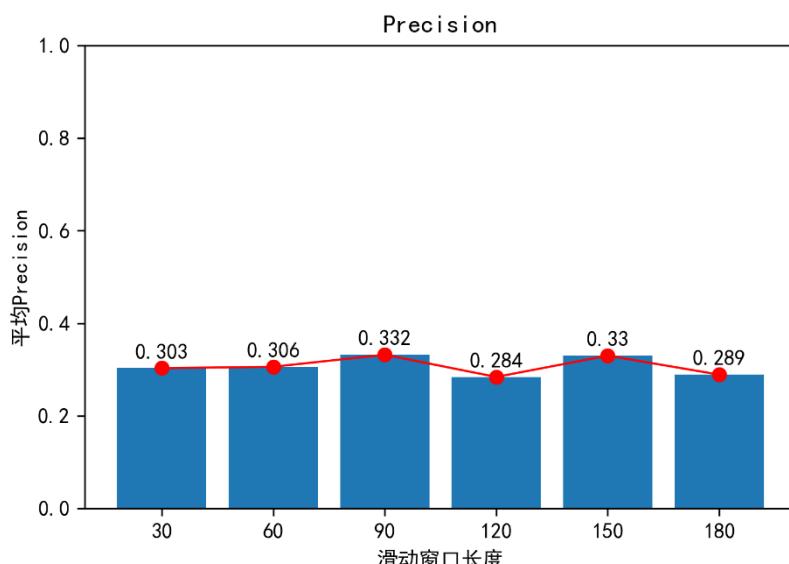
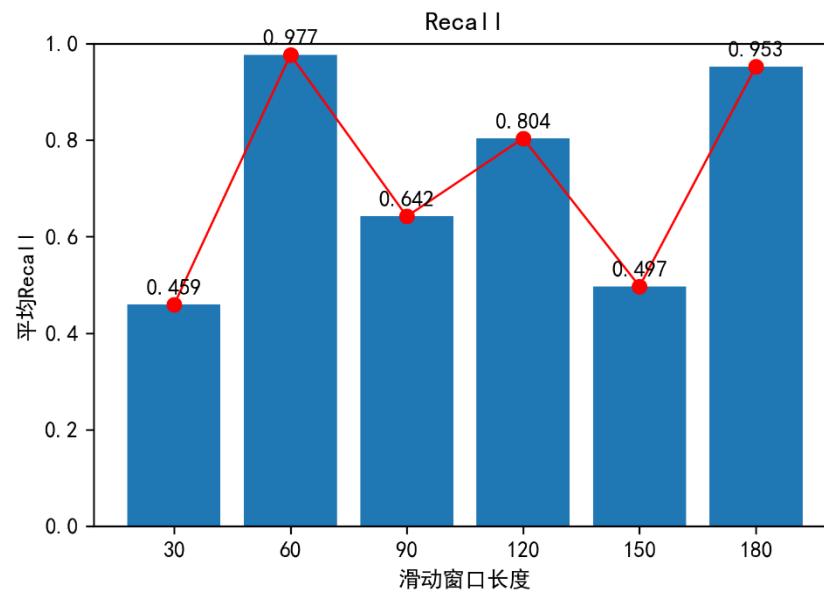
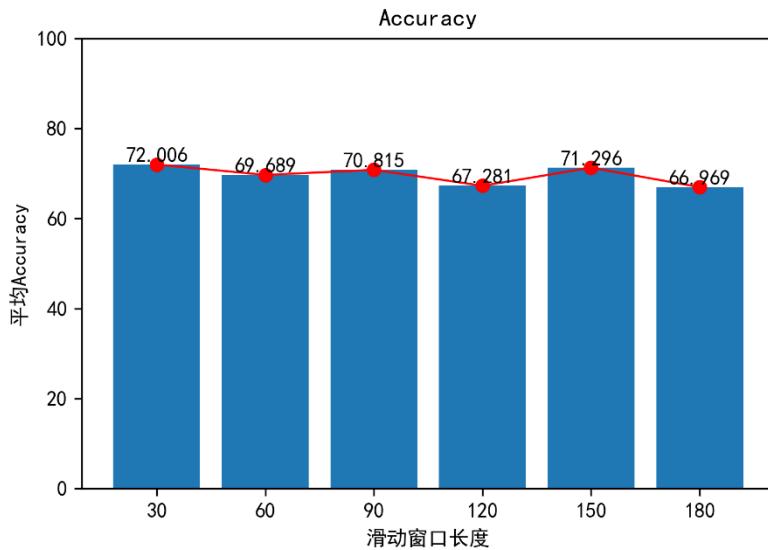
## 前期工作

- ◆利用LSCP异常检测多模型并行集成框架对WZ11-2E-14dsa井的数据进行异常检测实验。
- ◆对WZ11-2E-14dsa井的数据采用MAD-GAN方法进行初步实验探究。

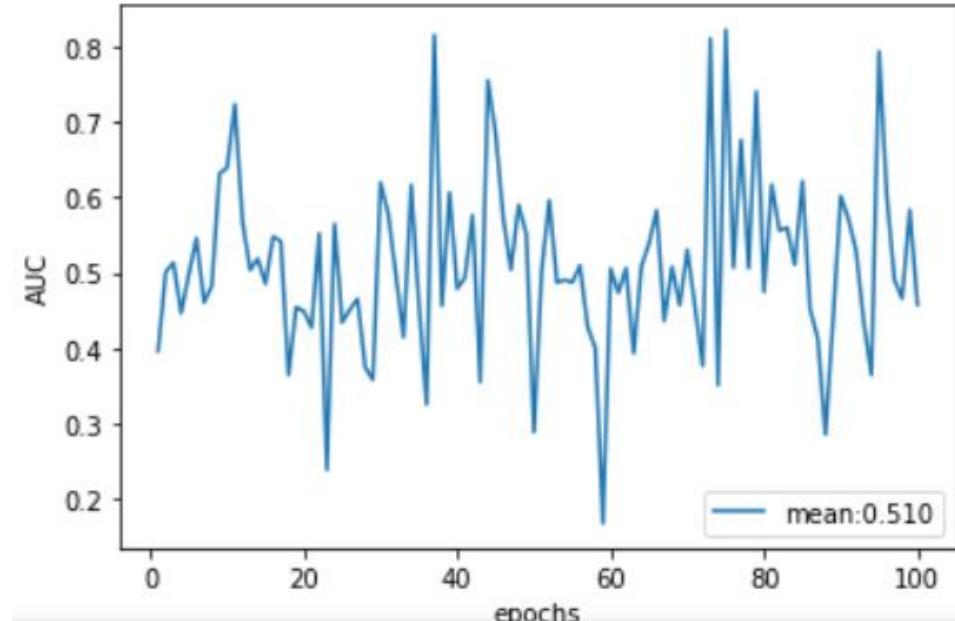
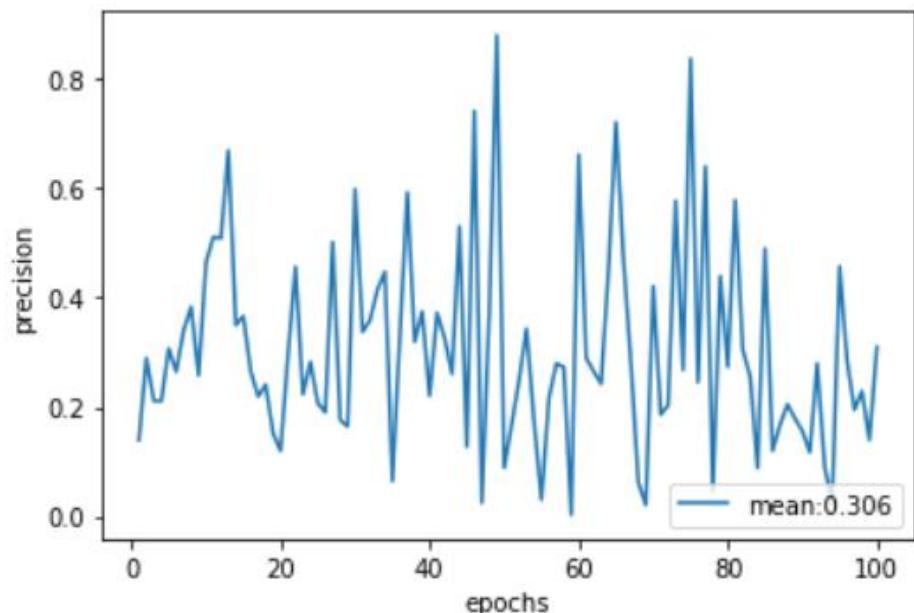
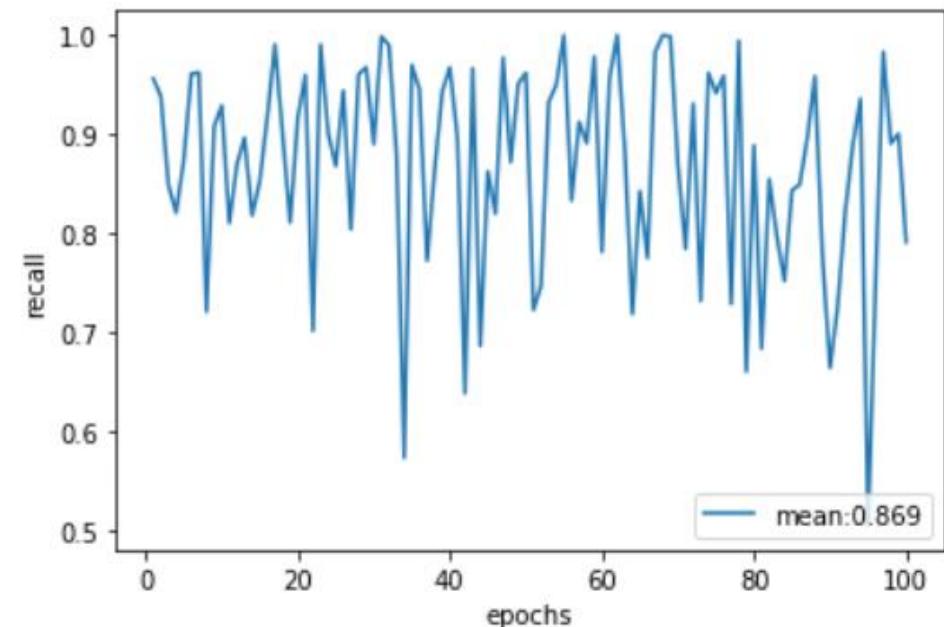
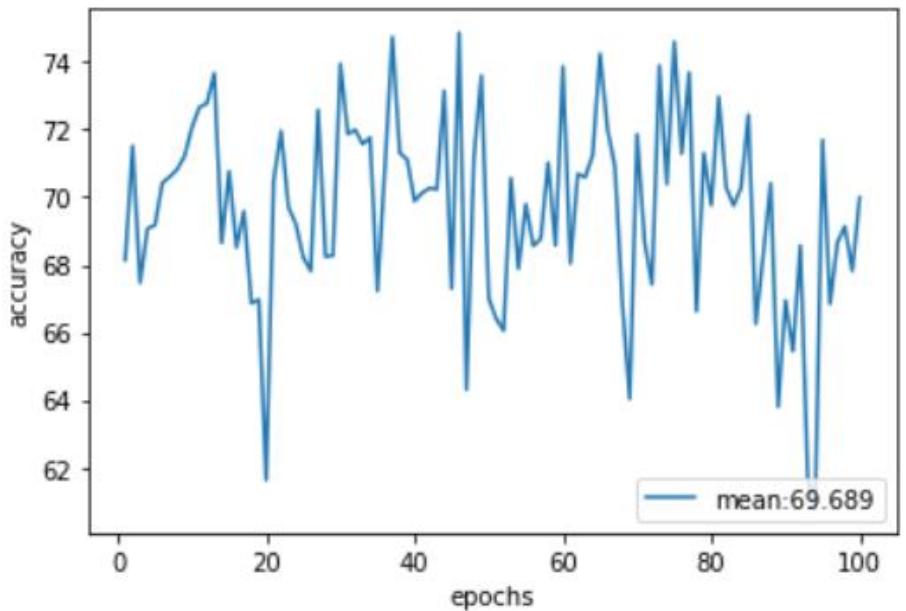
初步实验结果：

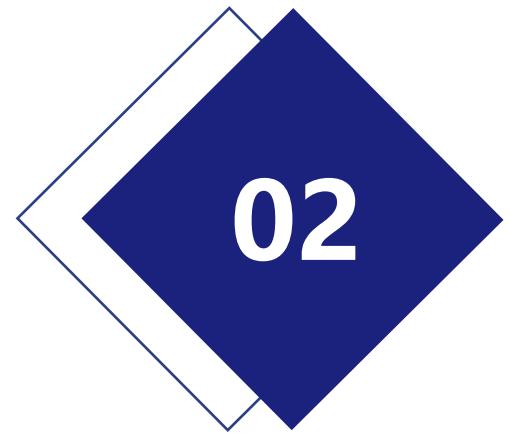
- 滑动窗口长度对检测效果有影响
- AUC指标不稳定

# MAD-GAN实验



# MAD-GAN实验





## MAD-GAN实验

## MAD-GAN实验

### 思考：MAD-GAN实验效果不理想的原因？

- 数据分布：异常数据和正常数据的比例和分布
- 数据特征：数据清洗、特征提取

Table 3: Anomaly Detection Metrics of MAD-GAN at Different PC Resolutions

EM	PC=1	PC=2	PC=3	PC=4	PC=5	PC=6	PC=7	PC=8	PC=9	PC=10	ALL
Pre	16.90	17.76	18.57	14.53	26.56	24.39	13.37	14.78	13.60	13.83	13.95
Rec	72.03	95.34	92.76	91.16	95.27	81.25	91.87	92.02	95.06	96.03	92.96
F_1	0.24	0.23	0.23	0.23	0.37	0.22	0.22	0.22	0.23	0.23	0.23

EM: evaluation metrics. Sub-sequence length equals to 30.

# MAD-GAN实验

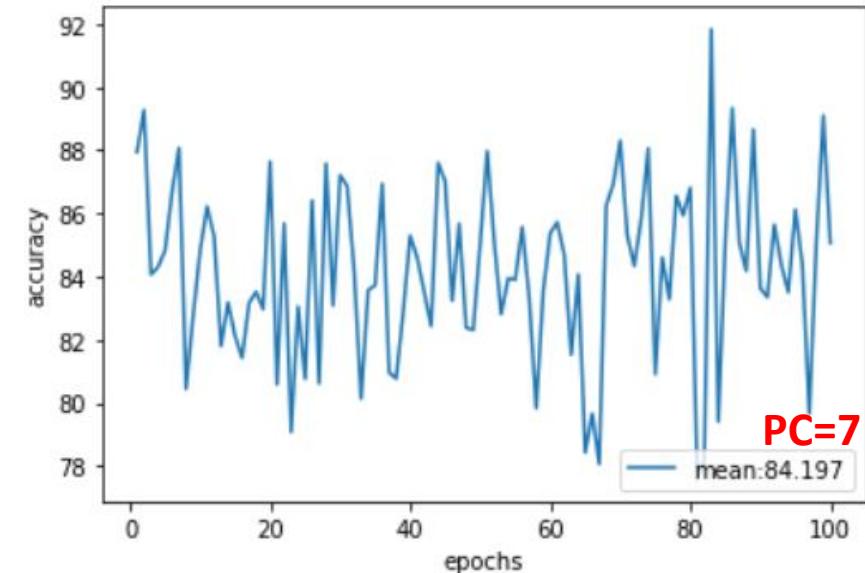
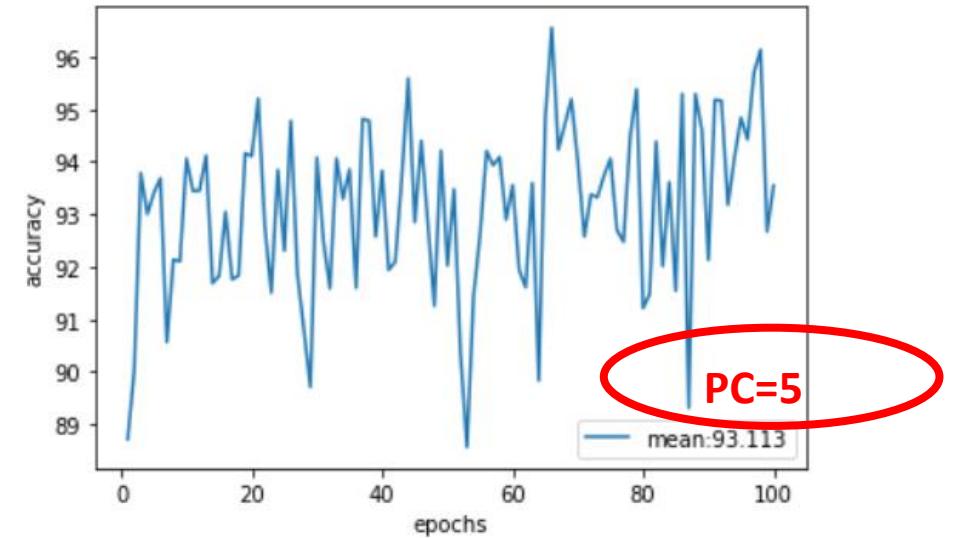
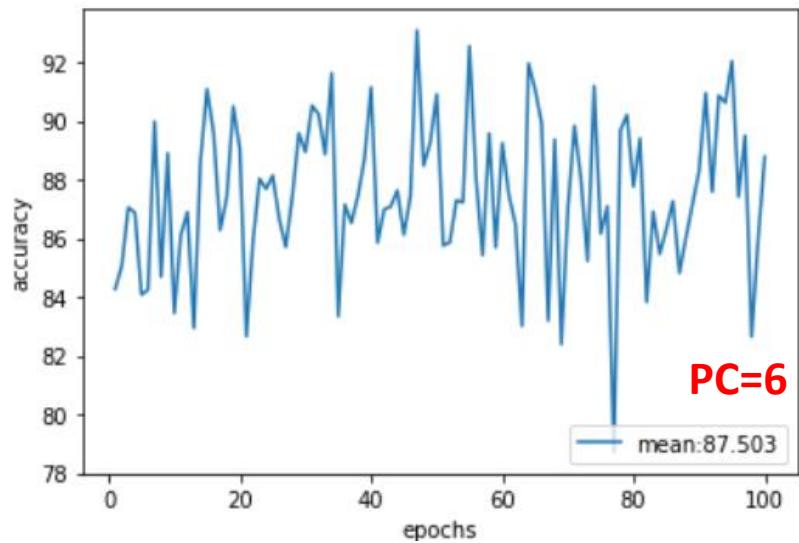
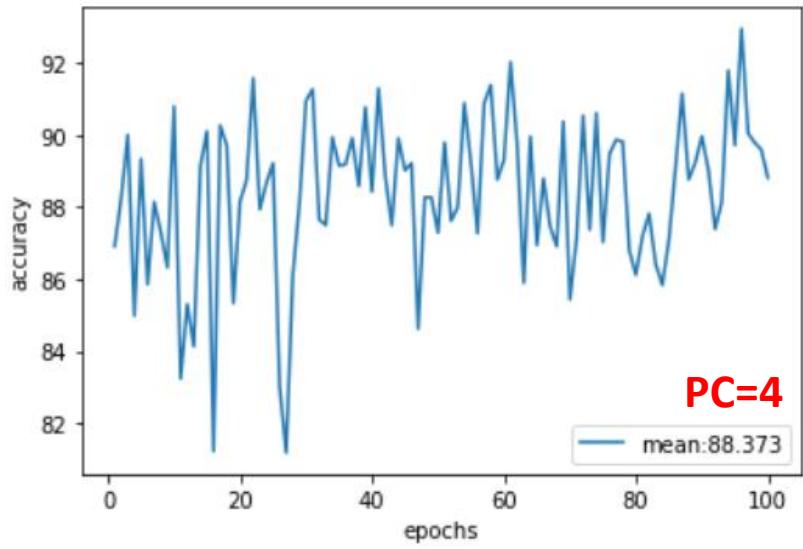
## 数据集构建：WZ12-2-B33井

- 两次卡钻事故开始时刻到结束时刻之间的数据认为是异常数据；
- 第一次卡钻事故开始时刻之前剔除2小时的数据后作为训练集（全部为正常数据，共631735条），第一次卡钻开始之后取409834条数据作为测试集（包含正常和异常）；
- 滑动窗口长度为30，步长为10提取序列数据。
- 对21个因子进行PCA降维（PC = 4, 5, 6, 7）

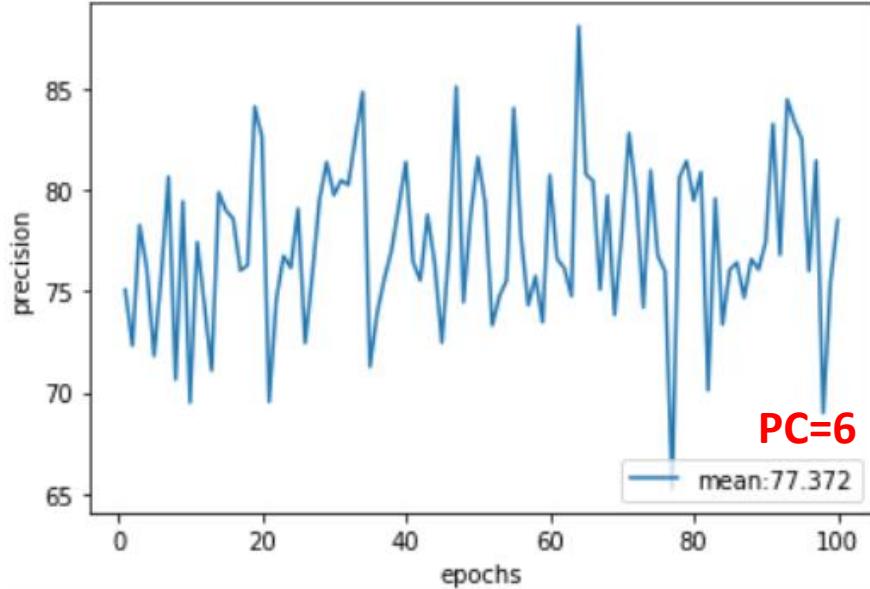
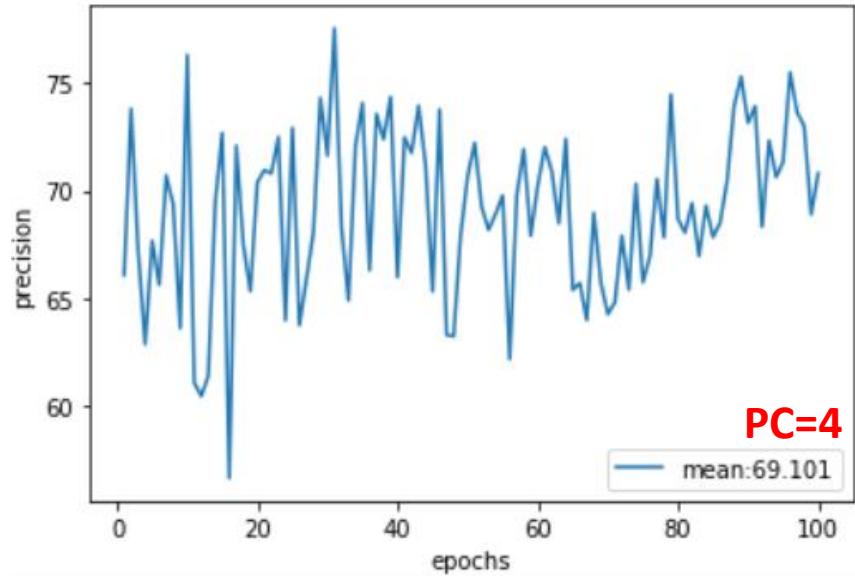
正常井况					备注	复杂情况					备注					
井况开始	结束日期	结束时间	详细井况类型	简单井况类型	在备注里明确解释工况内容，标出特殊点的作业，例如钻进中停泵接立柱的时间单独列出，电测期间的工况事故忽略	复杂情况类别	复杂开始	复杂结束时间	人工操作类型	人工开始	人工开始时	人工结束	人工结束时	复杂结束日	复杂结束时间	备注
时分秒	年月日	时分秒	下拉选择	下拉选择		下拉选择	年月日	时分秒	输入内容	年月日	时分秒	年月日	时分秒	年月日	时分秒	备注
18:55:00	2019/2/15	21:08:00	划眼循环	循环		无										
21:08:00	2019/2/17	5:47:00	倒划眼起钻	湿起下	频繁起钻，遇阻上提下放通过泵，停泵之后憋压、憋扭矩、	无										
5:47:00	2019/2/17	21:04:00	处理卡钻	其它		卡钻	2019/2/16	5:47:00	停水，工具失效	2019/2/16	5:47:00	2019/2/16	21:04:00	2019/2/16	21:04:00	
21:04:00	2019/2/18	0:32:00	循环	循环					具激发震击器，效							
0:32:00	2019/2/18	19:47:00	倒划眼起钻	湿起下	频繁起钻，遇阻上提下放通过泵憋扭矩，降低排量之后憋压	无										
19:47:00	2019/2/18	22:12:00	处理卡钻	湿起下		卡钻	2019/2/18	19:47:00	泵，逐步提高排量	2019/2/18	19:47:00	2019/2/18	22:12:00	2019/2/18	22:12:00	

# MAD-GAN实验

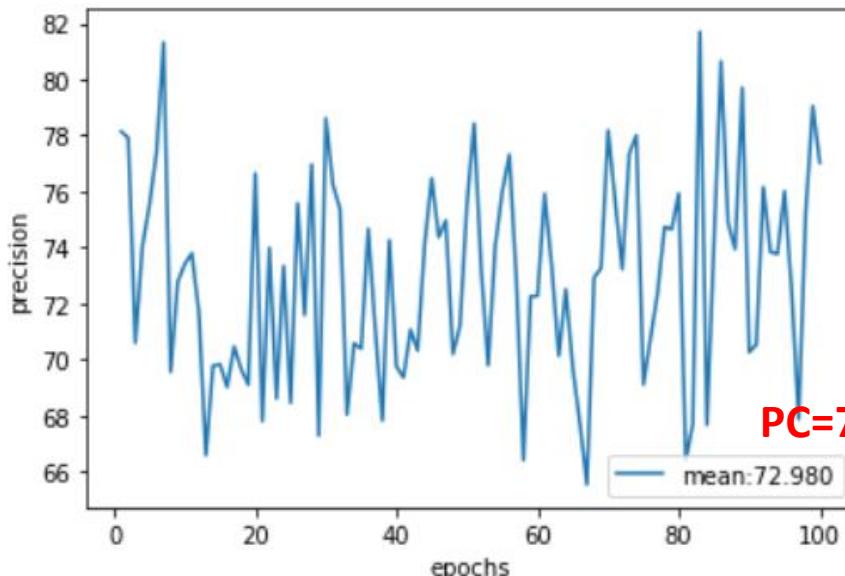
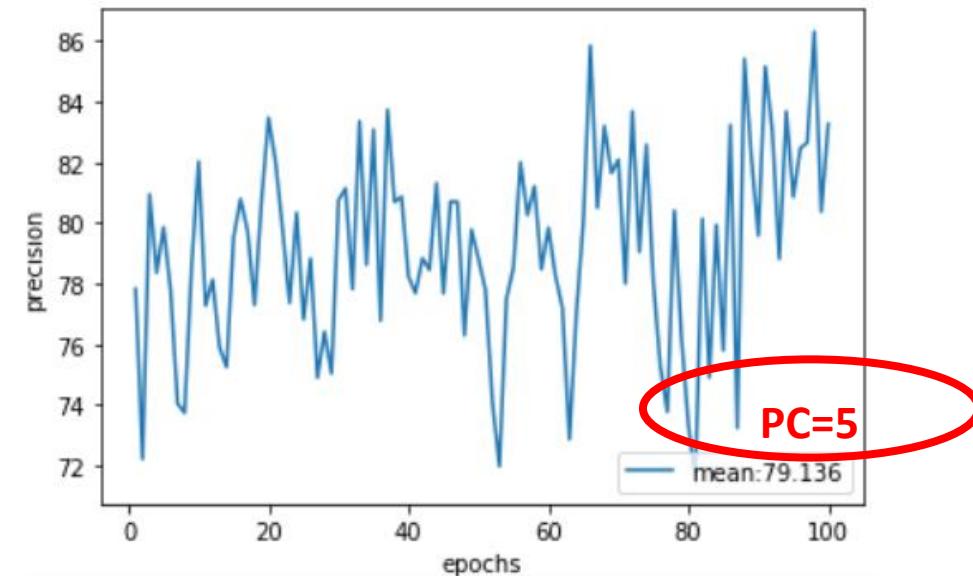
## Accuracy



# MAD-GAN实验

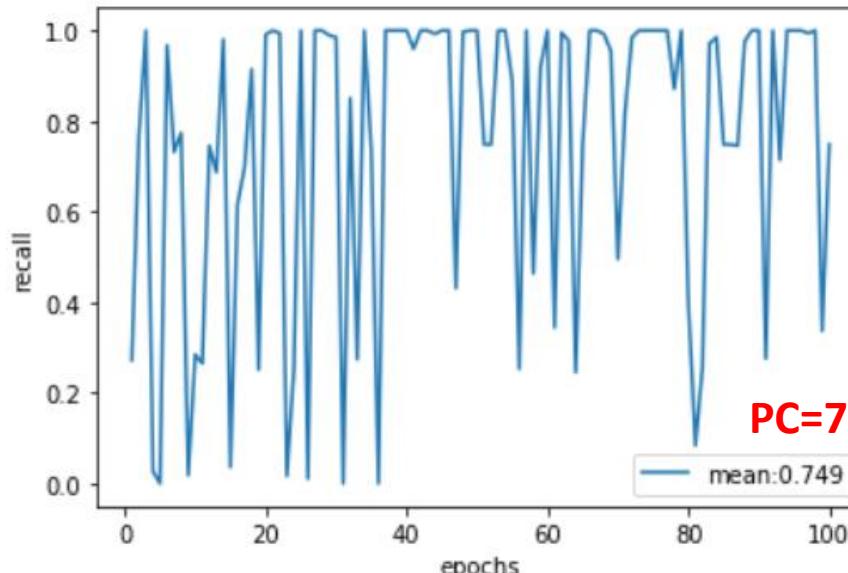
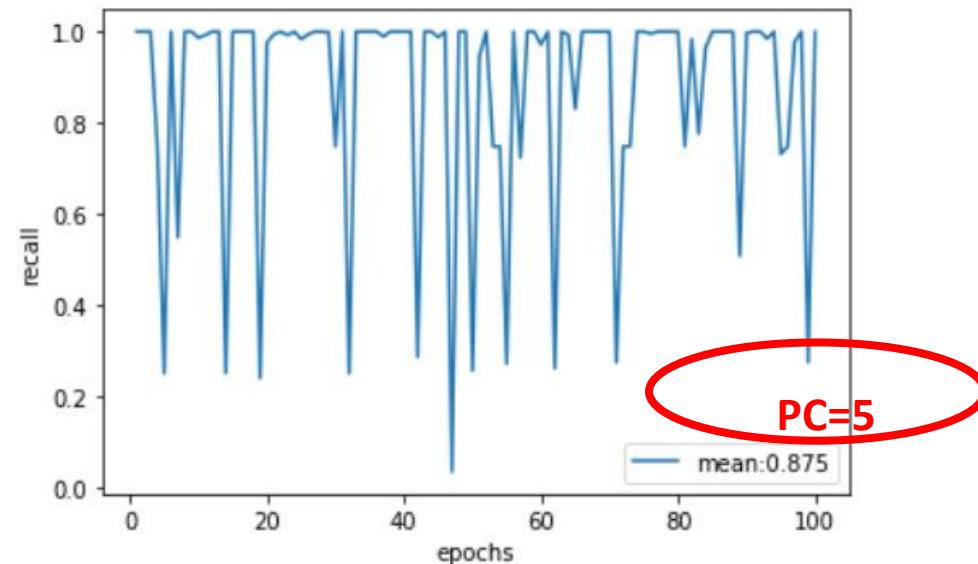
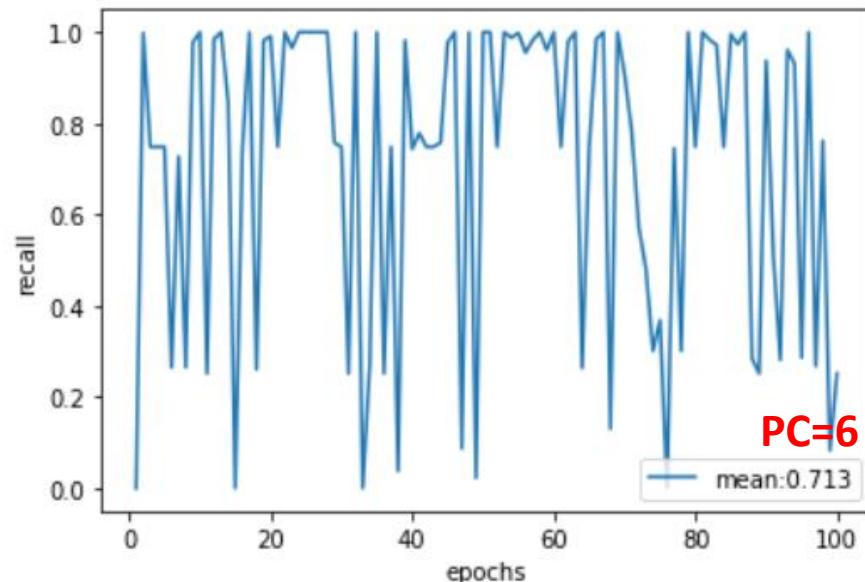
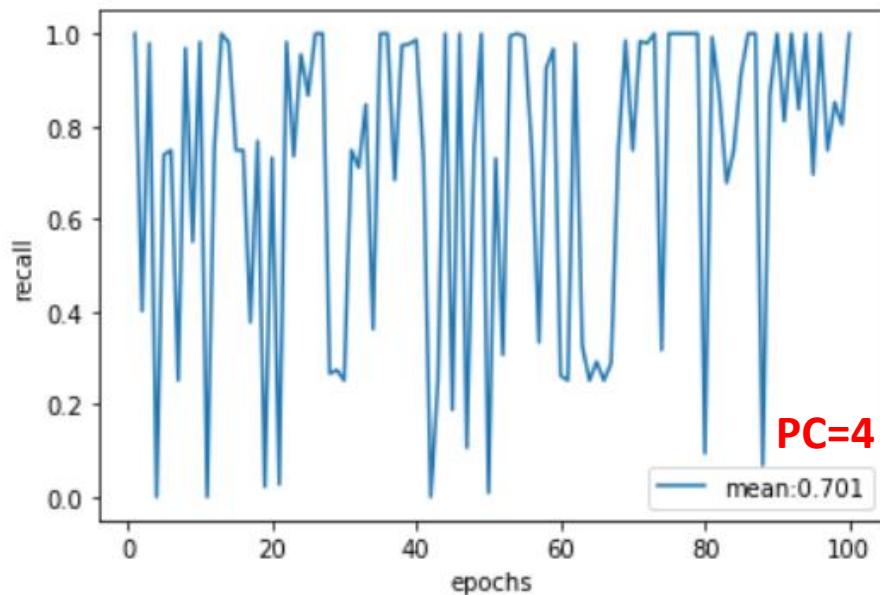


## Precision



# MAD-GAN实验

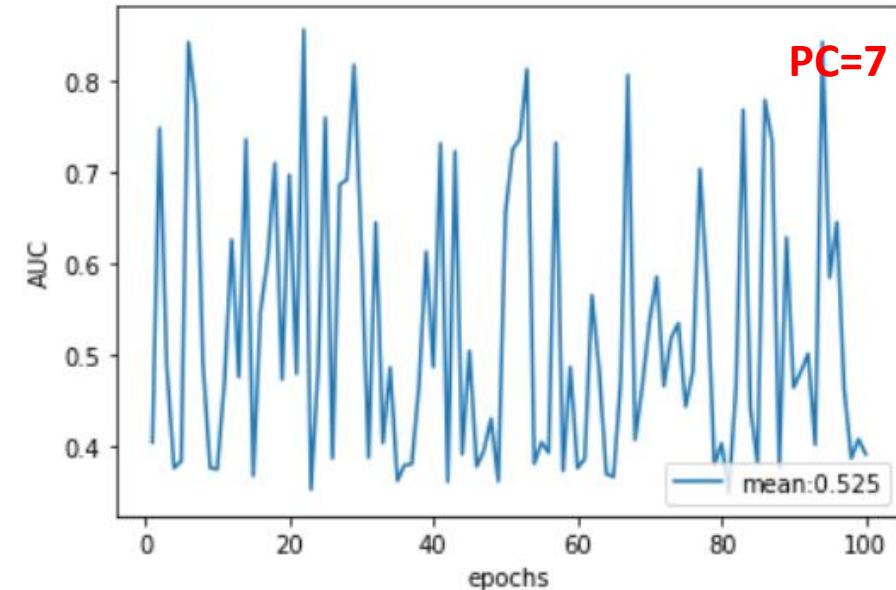
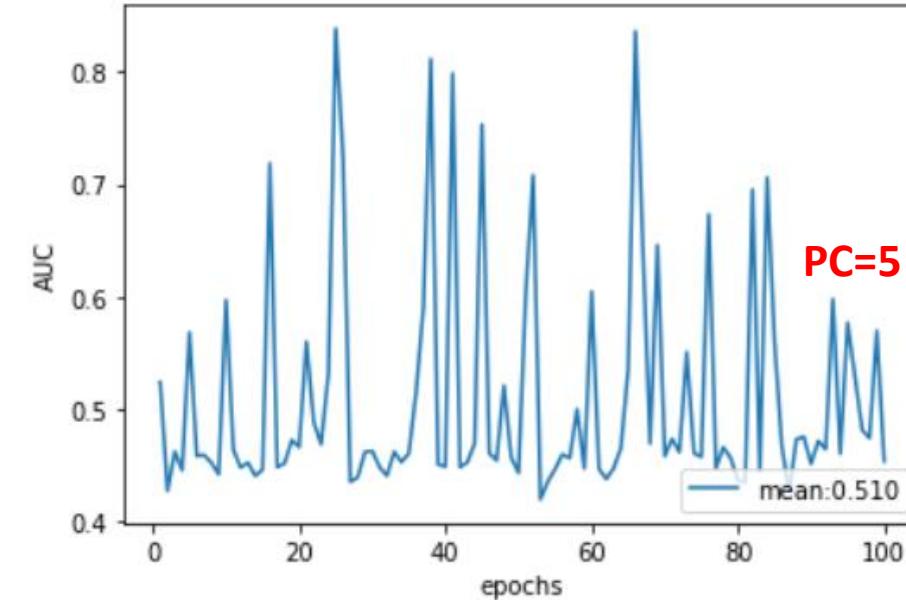
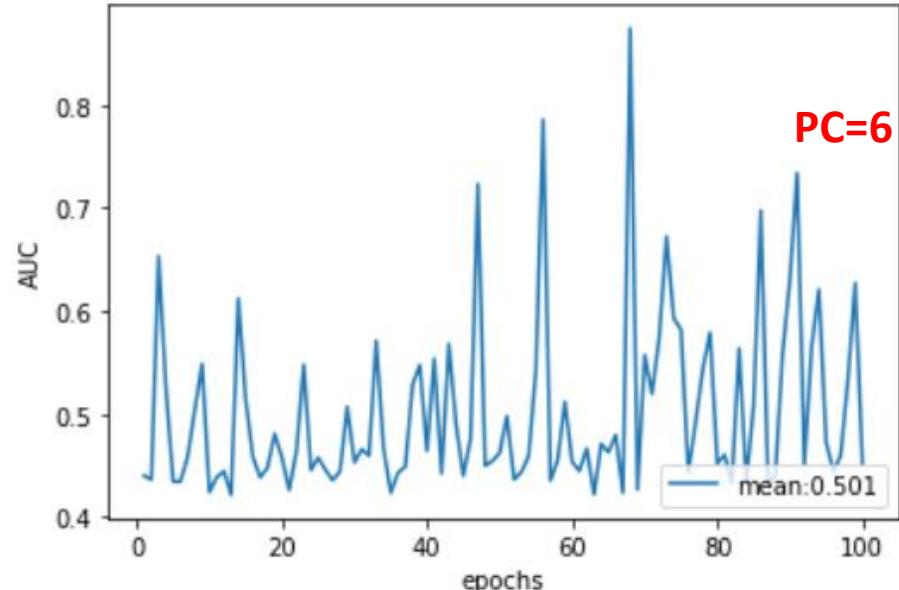
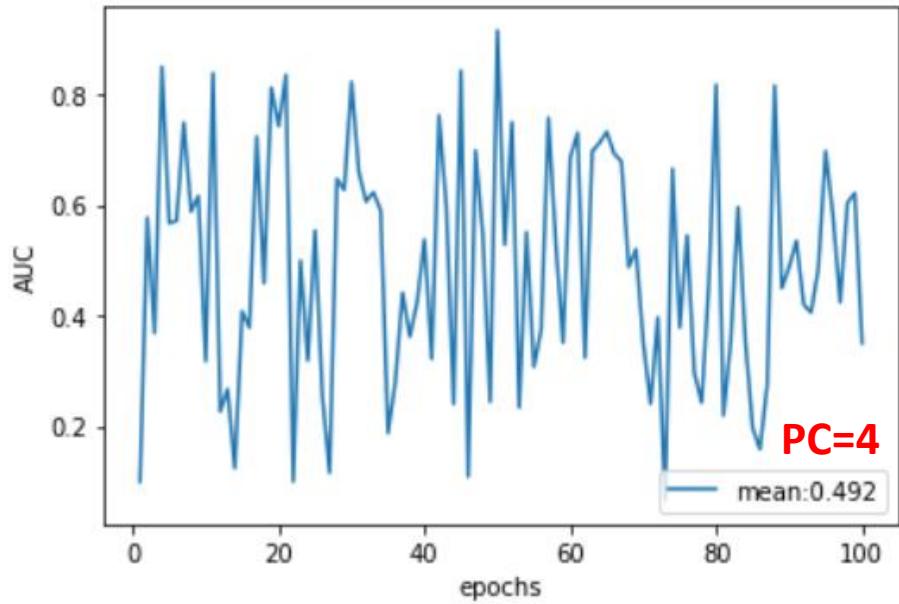
## Recall



# MAD-GAN实验

AUC

AUC指标依然不稳定



# MAD-GAN实验

## 关于AUC指标的思考：

- 排除阈值的影响对分类器进行评估，每个样本根据预测值排好序，然后依次将每个值当做阈值，计算该阈值下的FPR和TPR。
- 衡量的是分类器产生良好的相对实例分数的能力。

Label: 1, 1, 0, 0, 0, 1,

Score:0.9, 0.9, 0.1, 0.1, 0.1, 0.9

✓

Label: 1, 1, 0, 0, 0, 1,

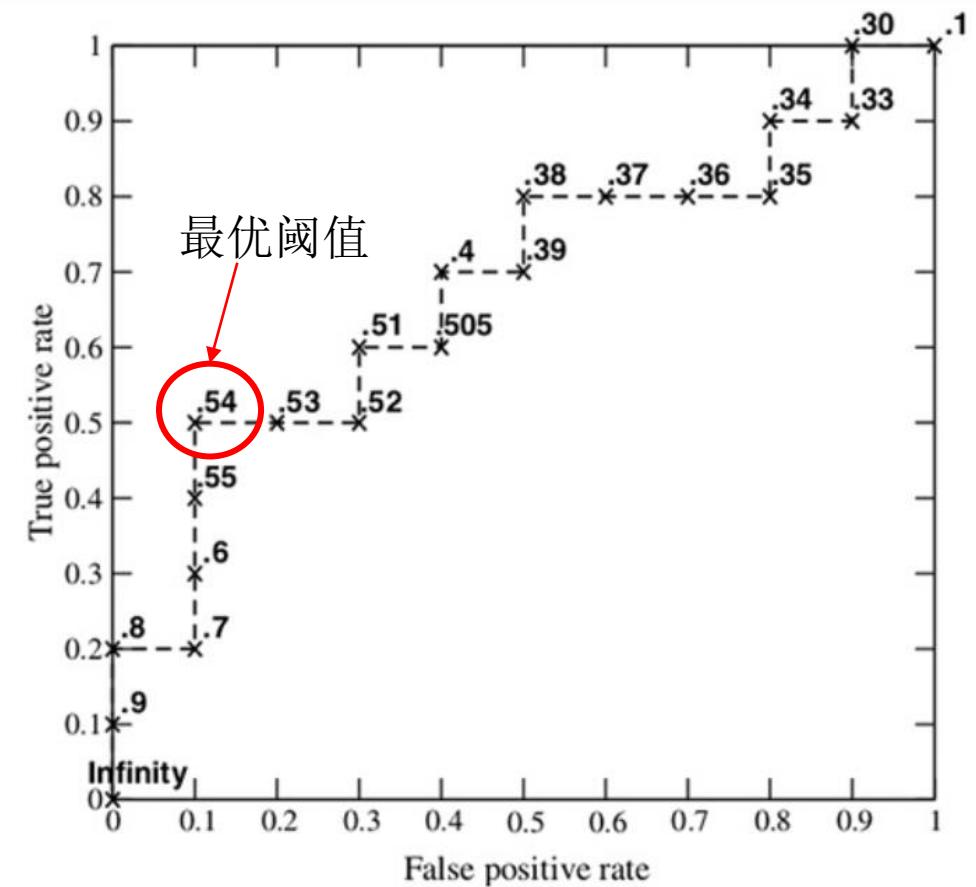
Score:0.5, 0.5, 0.4, 0.4, 0.4, 0.5

✗

- 选择合适的阈值使分类结果最佳
- ROC曲线可视化了一系列阈值所对应的FPR和TPR

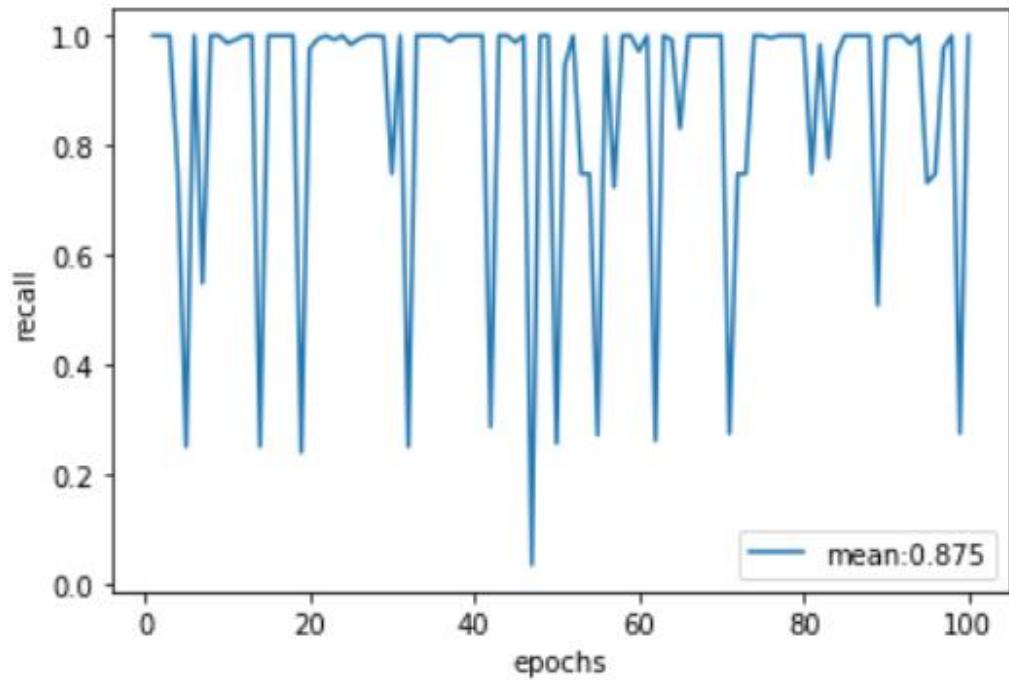


利用ROC曲线找到最优阈值，曲线上的点到(0,1)的距离最短时对应阈值为最优

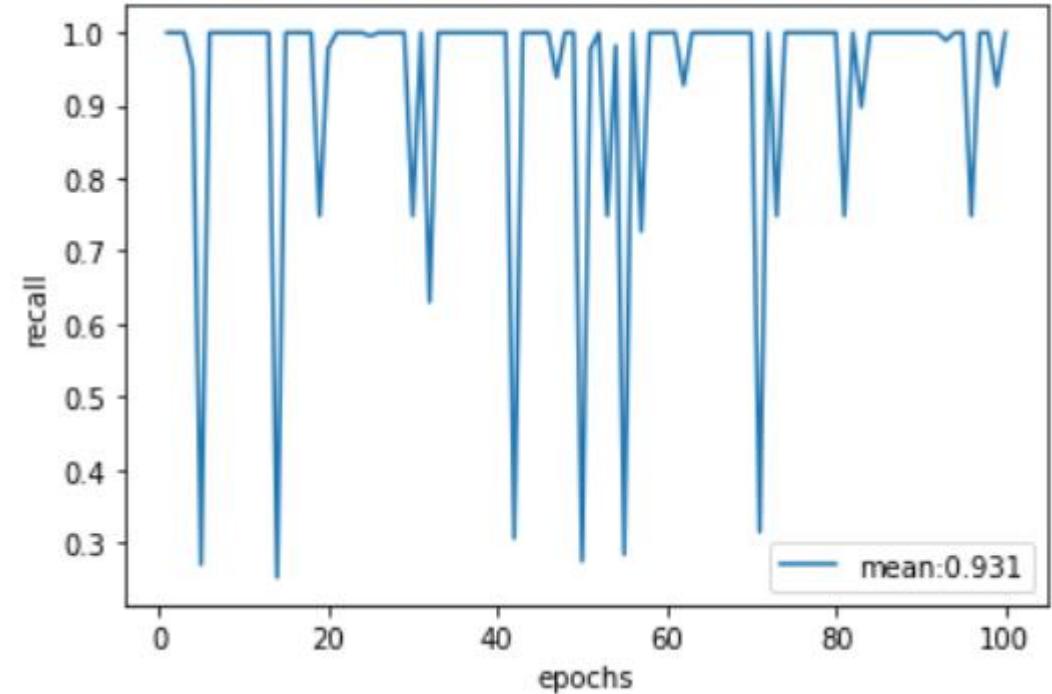


# MAD-GAN实验

## Recall



阈值等于0.5

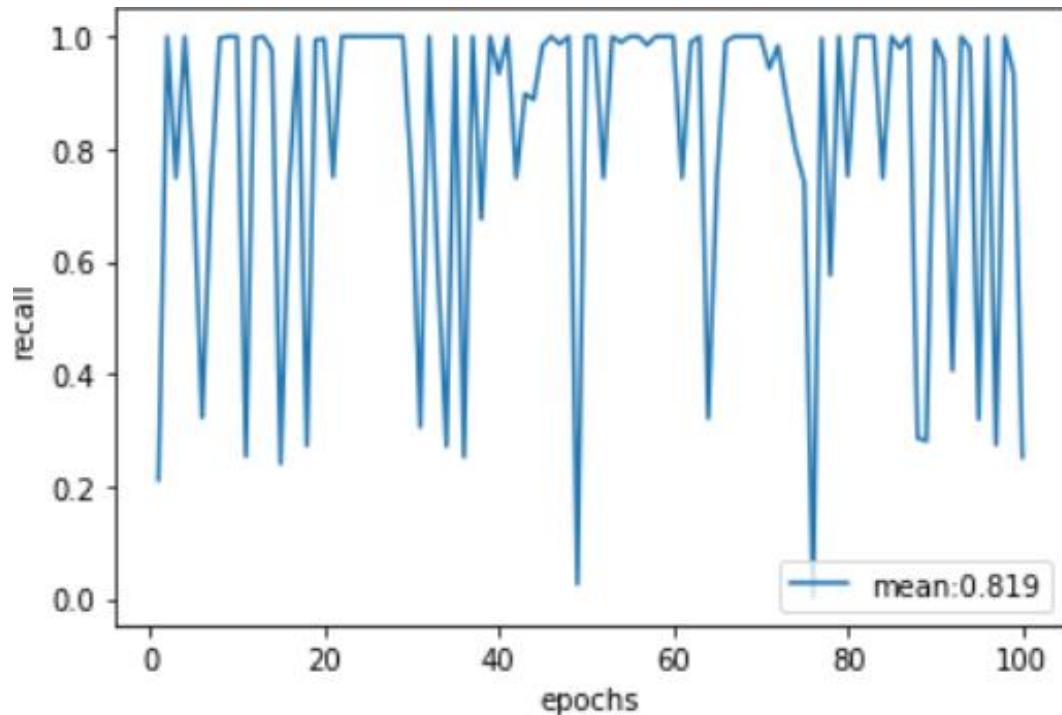


用ROC曲线找到最优阈值

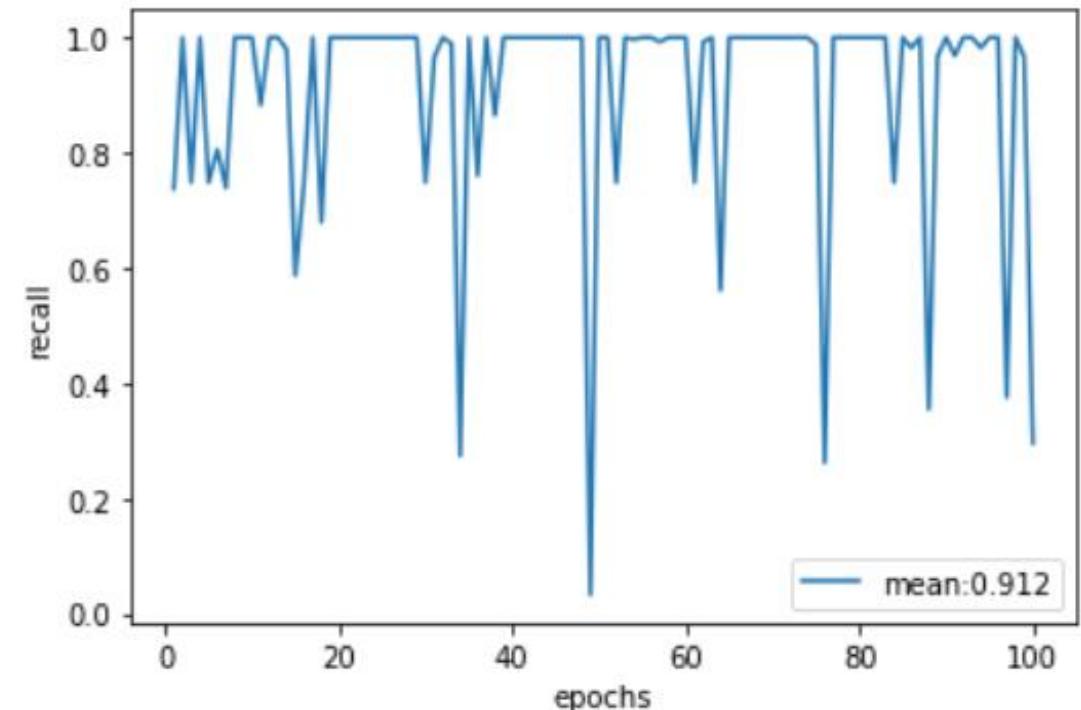
# MAD-GAN实验

WZ12-2-B33井截取另一段数据作为测试集

## Recall



阈值等于0.5



用ROC曲线找到最优阈值

## | 下一步计划

- ◆ 用**MAD-GAN**生成对抗网络的方法对钻井数据进行跨井实验



中国科学院空间应用工程与技术中心

Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences

---

# 钻井汇报

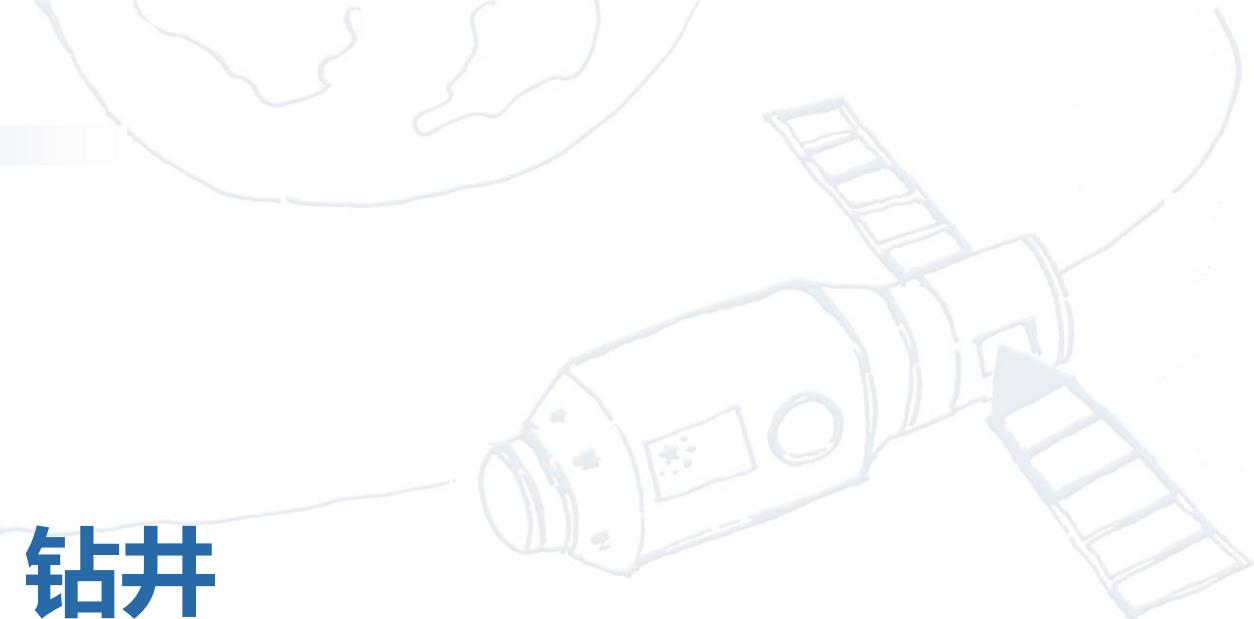
---

郭威龙

20200817

1

# 钻井



## 第一阶段实验

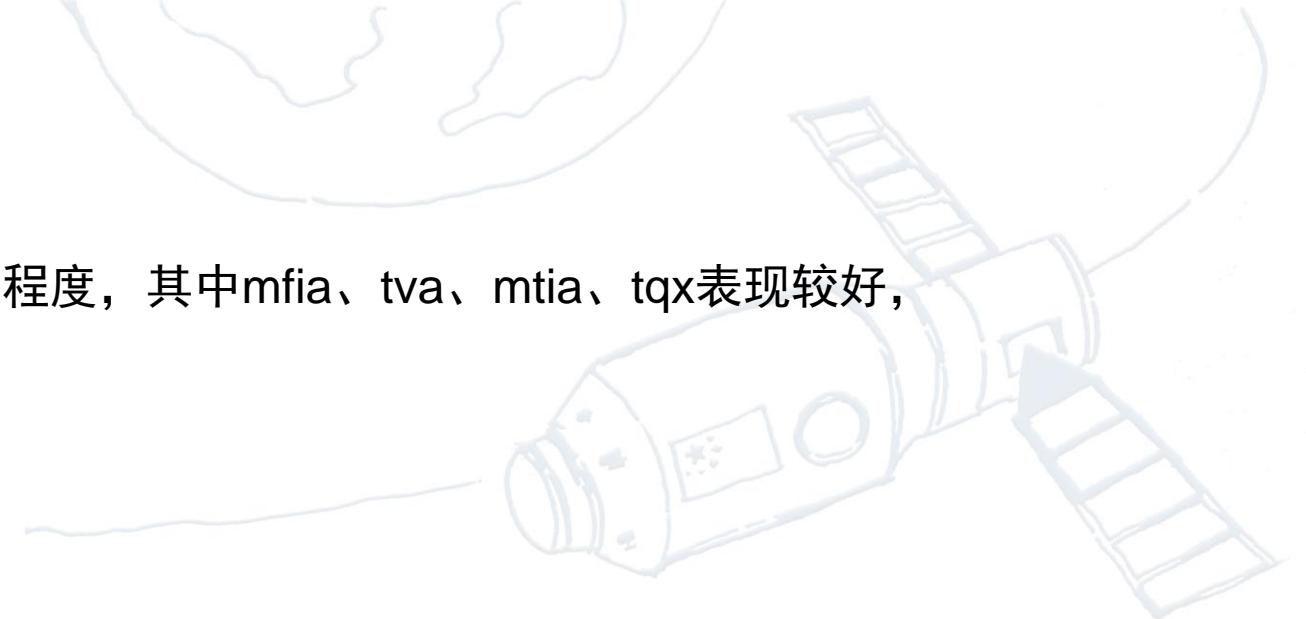
- 在一定程度上说明了哪些参数对卡钻事故的关联程度，其中mfia、tva、mtia、tqx表现较好，dbtm待测试；
  - LSTM的预测性能能够满足需求
- 待解决问题**
- 如何进行多因子对事故的预警研究

## 第二阶段方案

- 观察到：属性之间在纵向相关，单属性参数在横向相关，
- 所寻求的是横纵结合的局部关联特征

### 方案

- 假设当前的预测的数值是准确的，如果能够通过学习局部关联特征，把预测数据能够准确地分为正常状态/事故状态，那么当预测的数据被认为是事故时，在当前时刻即可做出对事故的预警，预警时间最长未预测数据能预测的最长时间，本次预警的事故发生的概率为分类器给出事故状态的得分（也可以基于该得分做其他的换算）
- 第一步：基于第一阶段实验结果，对二个三个参数联合（便于可视化），测试是否存在**分类器**能够正确区分正常状态/事故状态
- 第二步：将数据分块，利用卷积来进行半（14参数）/全参数训练
- 第三步：将LSTM与多因子检测模块连接起来，进行实时数据测试
- **第四步：进行跨井实验**



# 实验

## 总体思路

- 预测+分类；

## 预测

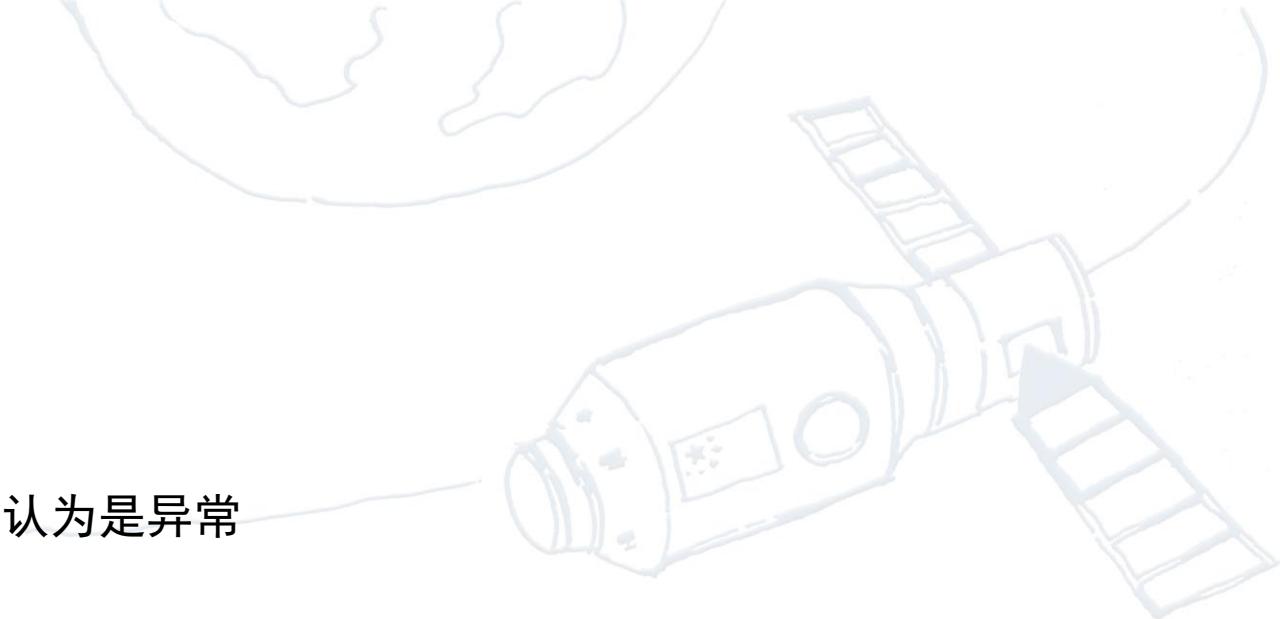
- LSTM

## 分类

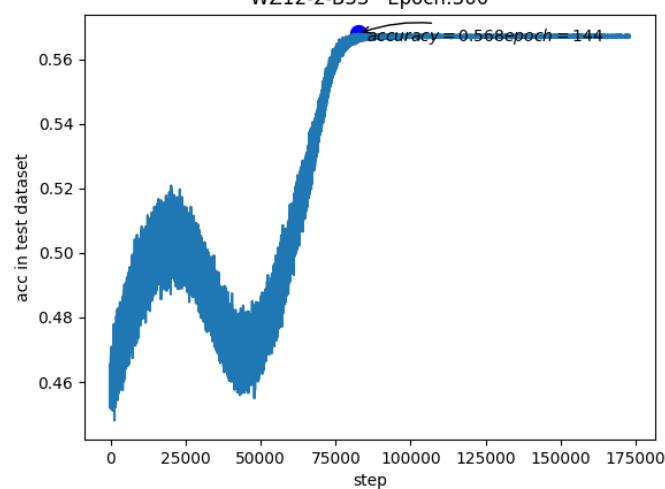
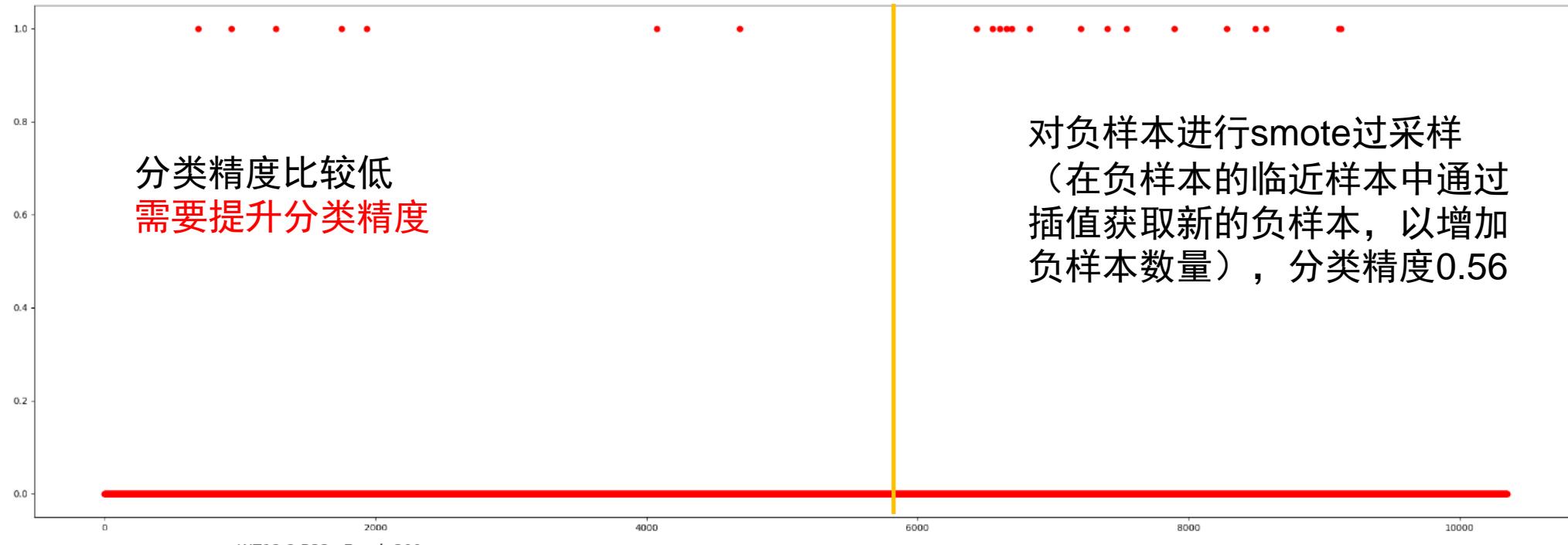
- 将标注文件中，正常状态认为是正常，异常状态认为是异常

## 评价标准

- 主流分类器种所使用的AP，AUC等



## 上次跨井演示中：



训练数据：WZ11-2E-14dsa

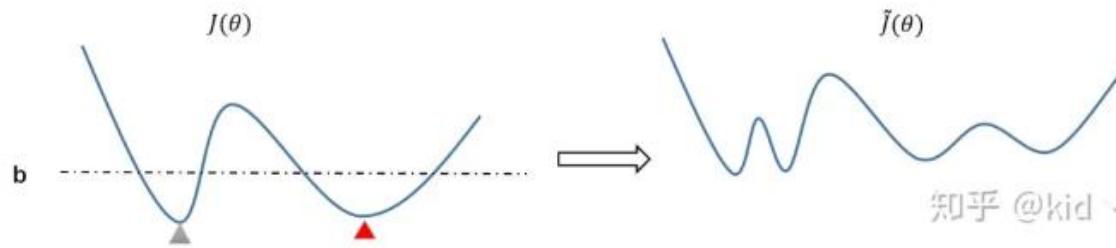
测试数据：WZ12-2-B33

1. 训练集中取一次卡钻前后的数据作为训练集正负样本
2. 测试集中同样选取一次卡钻前后的数据作为测试集的正负样本（大概7万条数据）

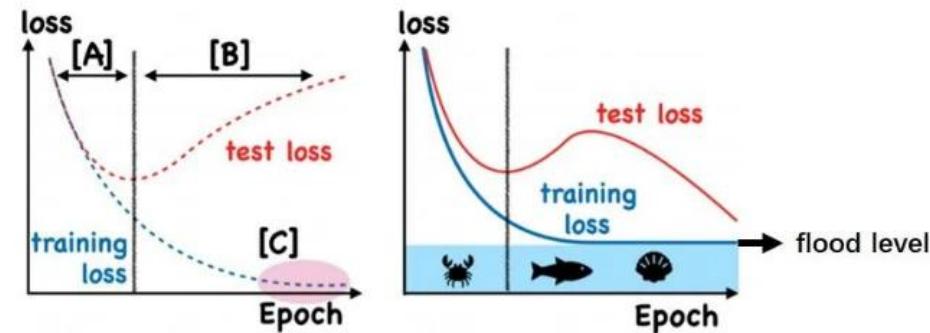
**横轴代表测试集上的样本，纵轴0为正常，1为事故**

# 实验

$$\tilde{J}(\theta) = |J(\theta) - b| + b,$$



知乎 @kid



(a) w/o Flooding

(b) w/ Flooding

(c) C10 w/o Flooding

(d) C10 w/ Flooding

知乎 @kid

```
outputs = model(inputs)
loss = criterion(outputs, labels)
flood = (loss-b).abs() + b # This is it!
optimizer.zero_grad()
flood.backward()
optimizer.step()
```

训练数据: WZ11-2E-14dsa

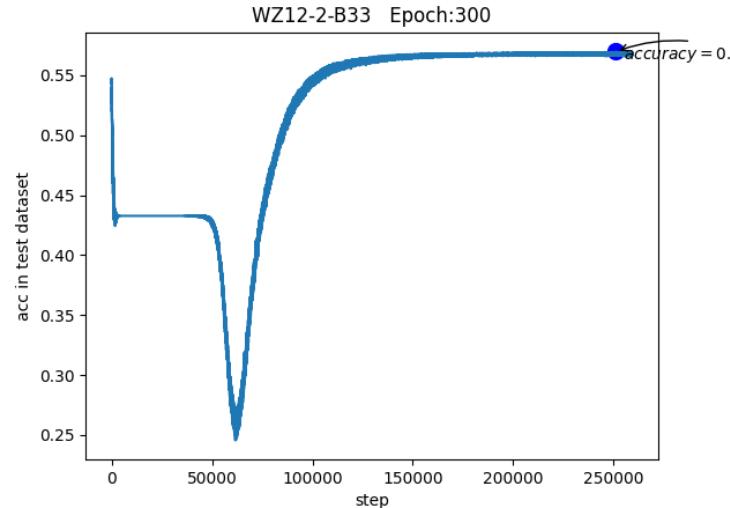
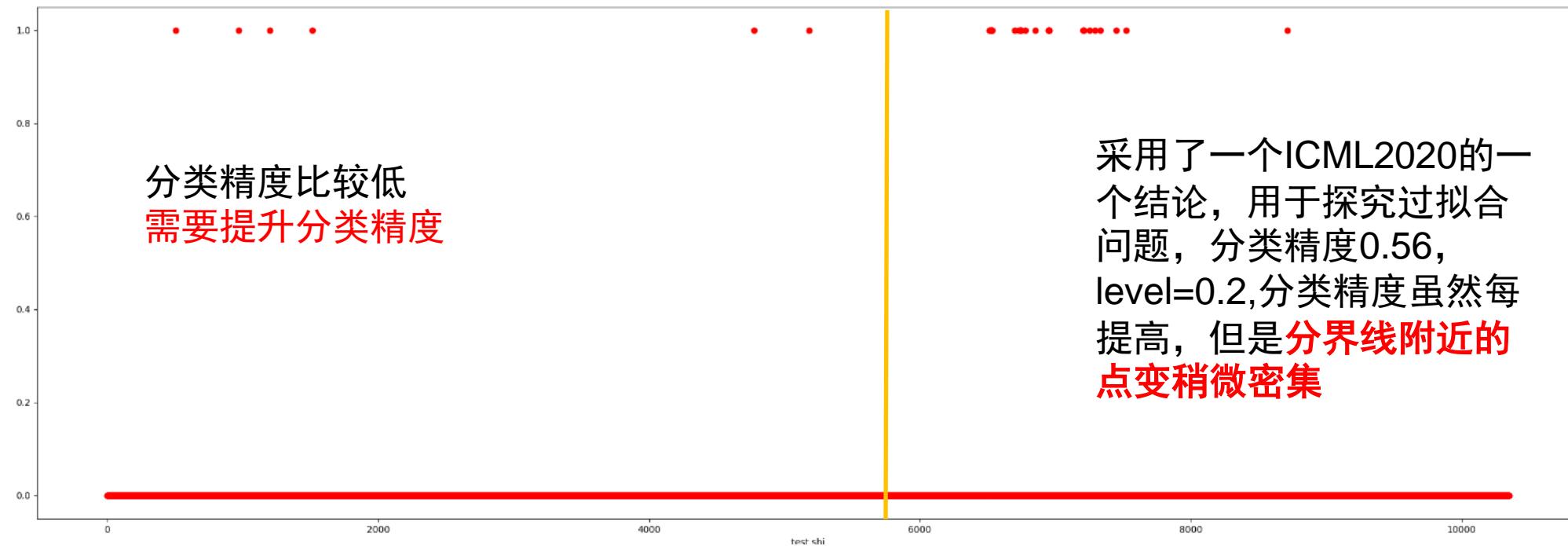
测试数据: WZ12-2-B33

1. 训练集中取一次卡钻前后的数据作为训练集正负样本
2. 测试集中同样选取一次卡钻前后的数据作为测试集的正负样本（大概7万条数据）

**横轴代表测试集上的样本，纵轴0为正常，1为事故**

51

# 实验



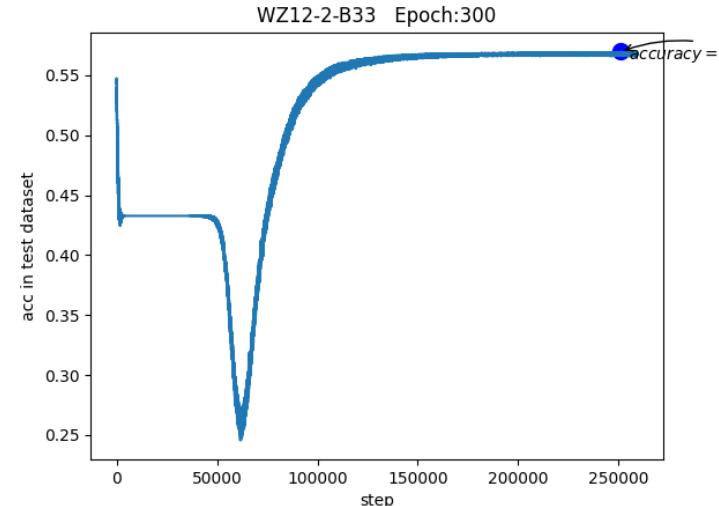
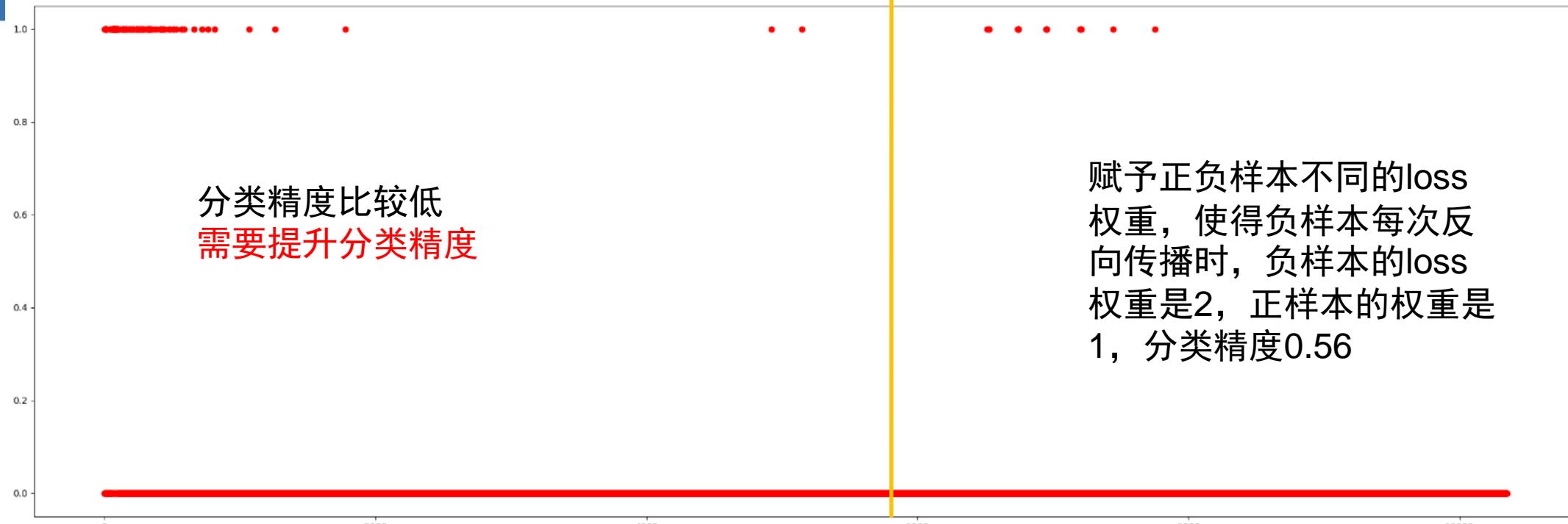
训练数据：WZ11-2E-14dsa

测试数据：WZ12-2-B33

1. 训练集中取一次卡钻前后的数据作为训练集正负样本
2. 测试集中同样选取一次卡钻前后的数据作为测试集的正负样本（大概7万条数据）

**横轴代表测试集上的样本，纵轴0为正常，1为事故**

# 实验



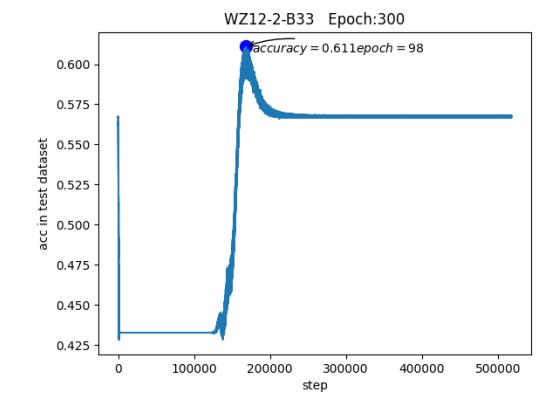
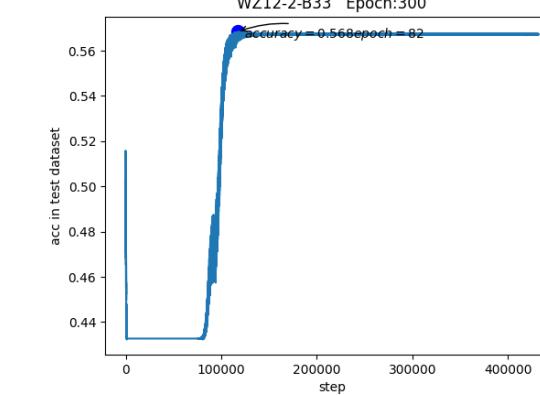
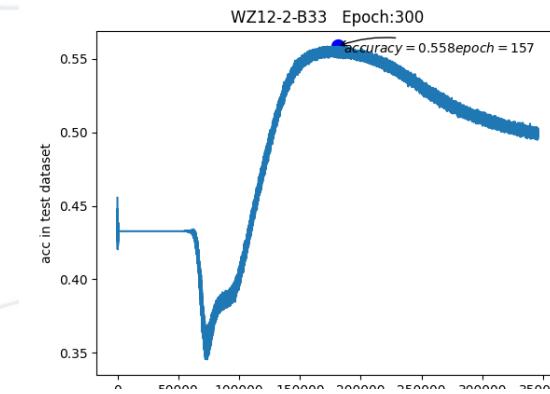
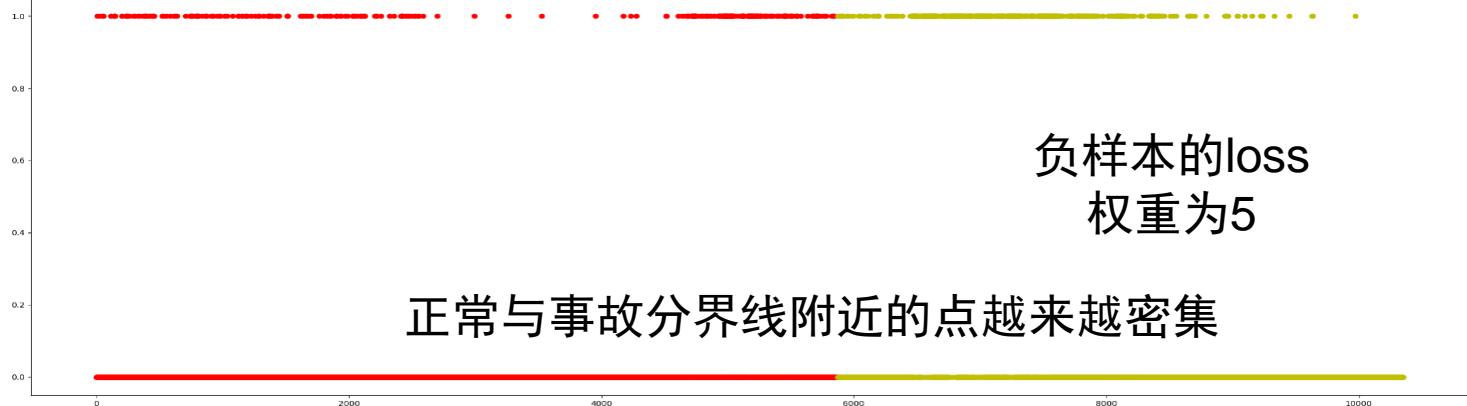
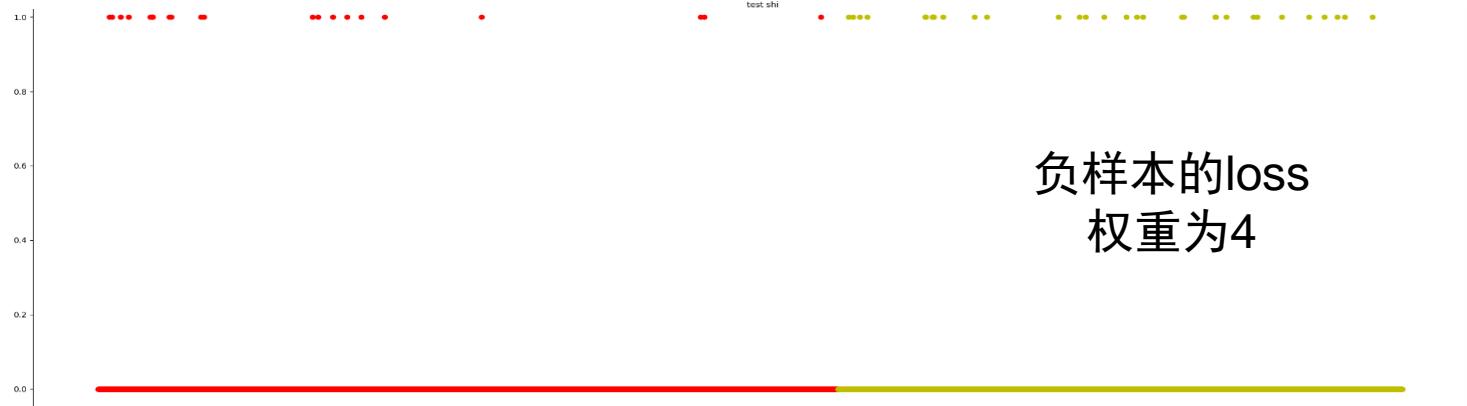
训练数据：WZ11-2E-14dsa

测试数据：WZ12-2-B33

1. 训练集中取一次卡钻前后的数据作为训练集正负样本
2. 测试集中同样选取一次卡钻前后的数据作为测试集的正负样本（大概7万条数据）

横轴代表测试集上的样本，纵轴0为正常，1为事故

# 实验



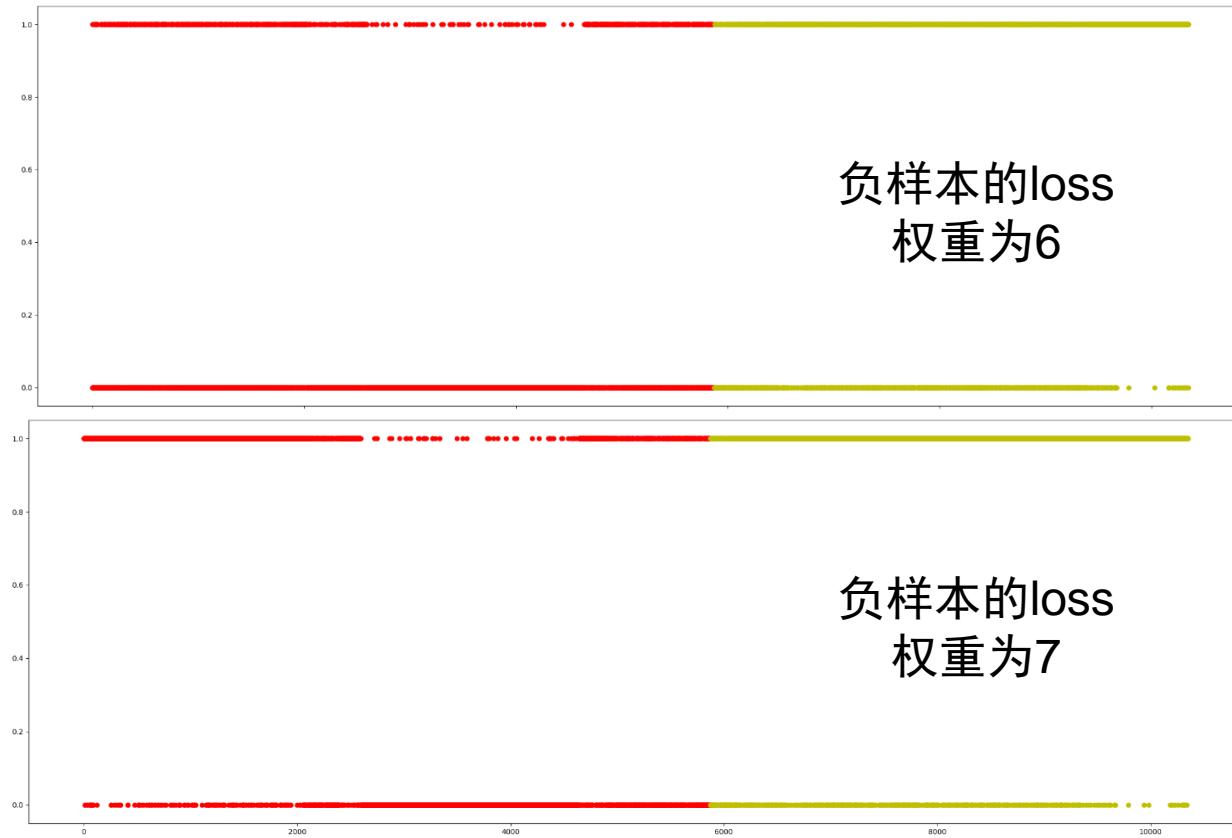
0.55

0.56

0.61

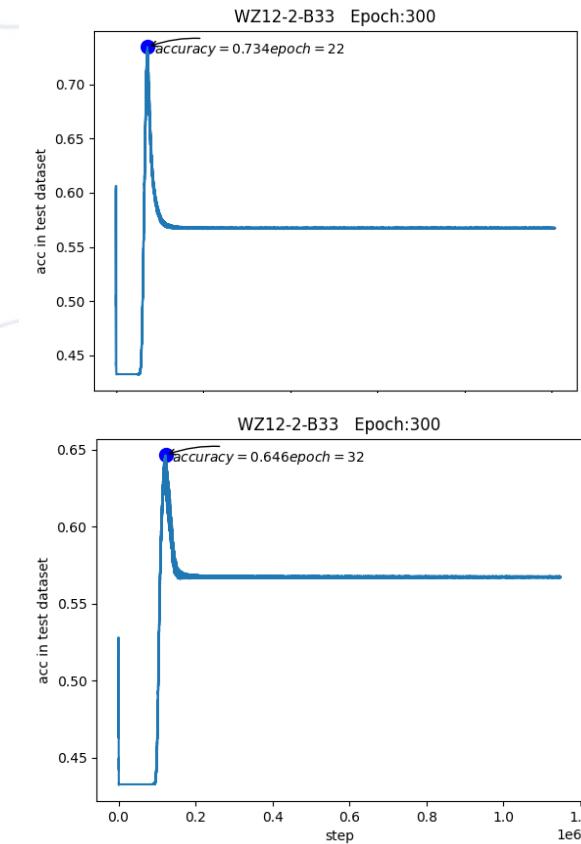
54

# 实验



负样本的loss  
权重为6

负样本的loss  
权重为7



0.734

0.646

基本上在权重为6的时候，精度上已经算  
比较高了，但是从分类结果的分布上来说，  
却离项目需求，越来越远

# 总结

需求：做预警--寻找在钻井过程中，**特征信息由正常变为异常的时间点**

问题：

1. 追求分类器的精度是没有意义的（这也是衡量标准的问题）

2. 对于分类器来说，我们寻找的异常点，

➤ 不应该包括事故已经发生的点，这样可能会使网络学习到一些，对于寻找异常点来说错误的信息，（标注的事故阶段为处理事故的阶段，其中很可能存在一些各属性指标正常的时间点，把这些信息都认为是事故，会是网络学习到错误的判别信息）

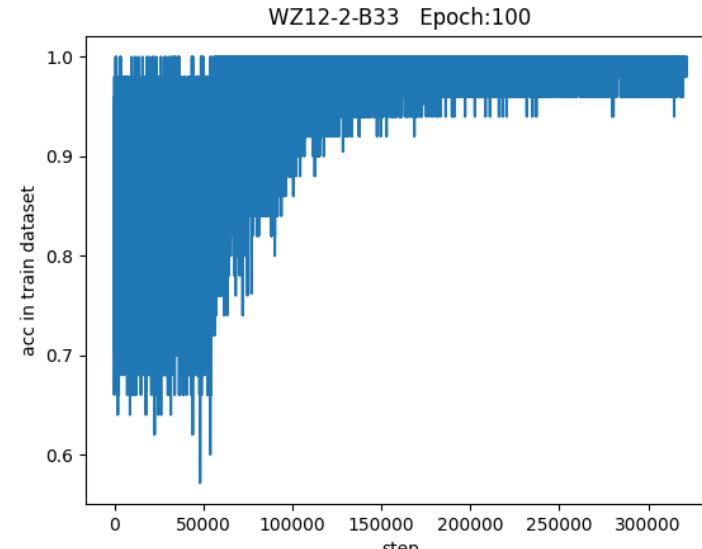
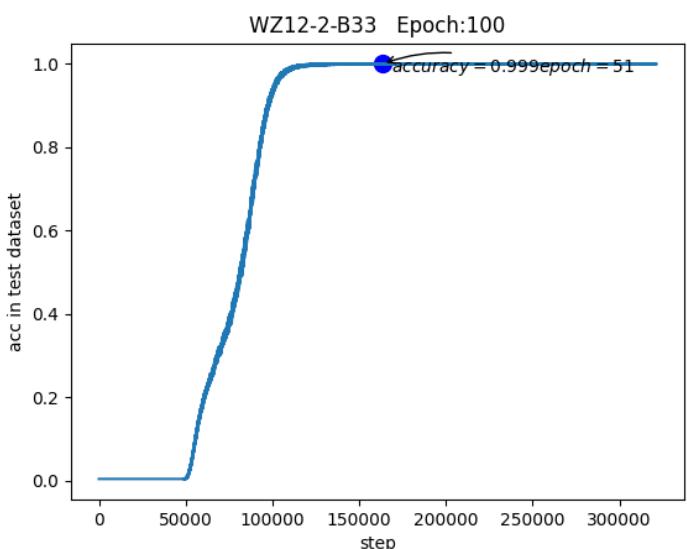
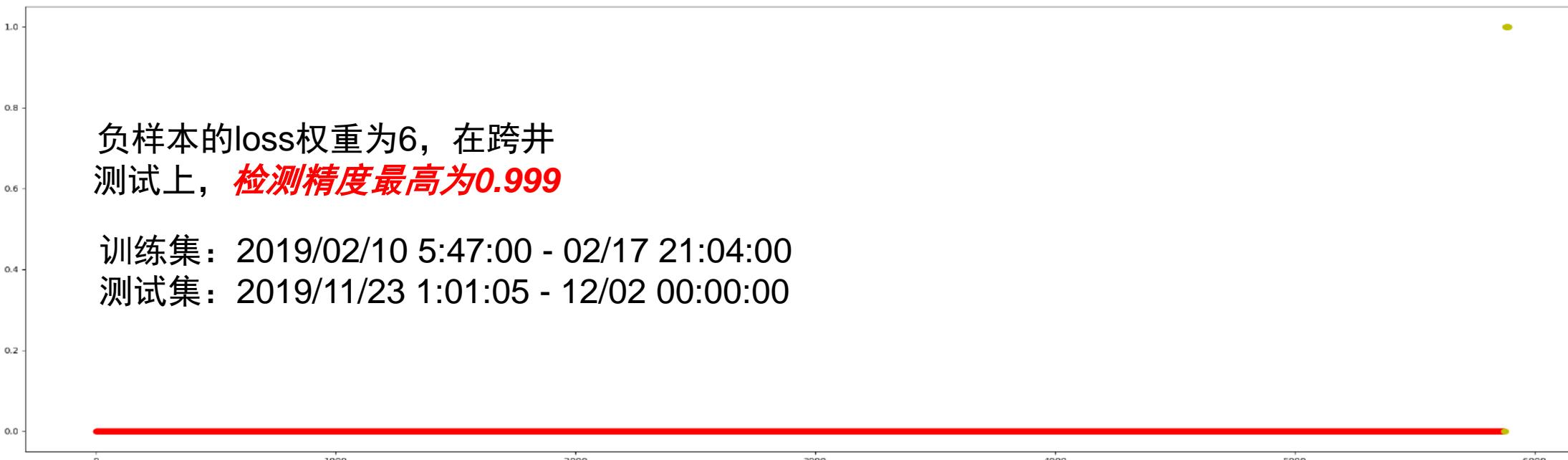
解决方案：

1. 仍然看作是二分类的问题，但是训练时，事故发生时间点一小段之后的数据，丢弃，不参与训练，对于测试集中，事故发生之后，处理事故阶段的数据，随机分类

2. 看作是三分类问题，正常状态，事故将要发生点，事故处理状态，

3. 或者对事故处理阶段的数据进行稀疏采样，仍然作为负样本

# 实验-解决方案1：只把事故发生起点附近密集采样，作为负样本



训练数据：WZ11-2E-14dsa  
测试数据：WZ12-2-B33  
1. 训练集中取一次卡钻前后的数据作  
为训练集正负样本  
2. 测试集中同样选取一次卡钻前后的  
数据作为测试集的正负样本（大概7  
万条数据）  
**横轴代表测试集上的样本，纵轴0为  
正常，1为事故**

## 下一步计划

- 钻井

尝试方案1，2，3进行试验和结果分析

