



基于kde与xgboost的故障检测和根因定位

浦智运维团队

浦发银行信用卡中心技术运营部

2022 CCF国际AIOps挑战赛决赛暨AIOps研讨会

目录 CONTENTS

第一节 团队介绍

第二节 挑战与应对

第三节 总体框架

第四节 总结与展望

第一章



团队介绍

团队介绍



浦智运维战队

浦发银行信用卡中心 技术运营部

队员：王楠 冯耀 鲁羽佳 刘敏敏 朱祥

复赛成绩：第二名



团队设计产品——AI智能平台



模型依据规则智能识别日志冗余、语义并量化得分

智能识别异常日志并快速告警，同时引导标注，实现模型知识库增加，识别能力增强



第二章节



挑战与应对

总体思路

为实现实时，准确的预测故障分类，采用无监督学习算法（核密度估计kde）进行异常检测，识别异常指标，kde异常检测的结果经过加工后作为有监督学习算法（xgboost）的输入，进行故障分类。

Kde异常检测目标

近3分钟指标取值

kde训练集

近33分钟至近3
分钟指标取值

kde异常
检测

异常检测
结果处理

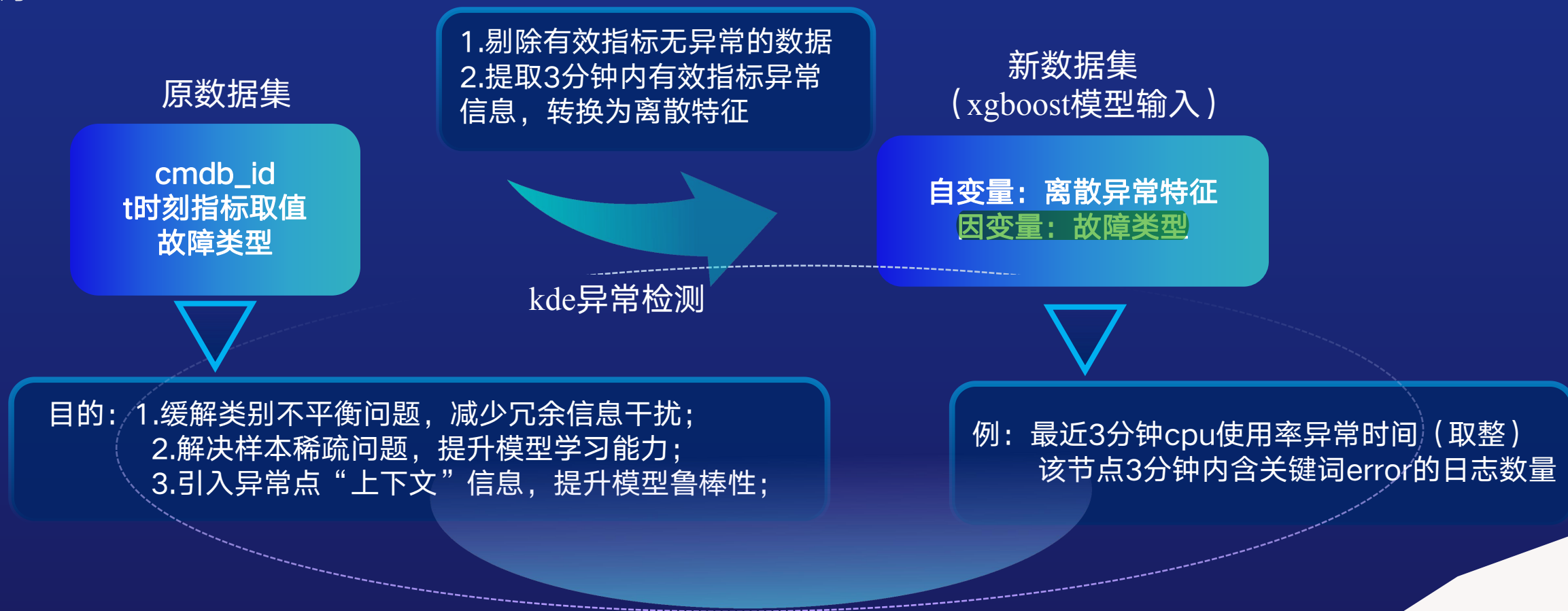
xgboost
故障分类

故障分
类结果



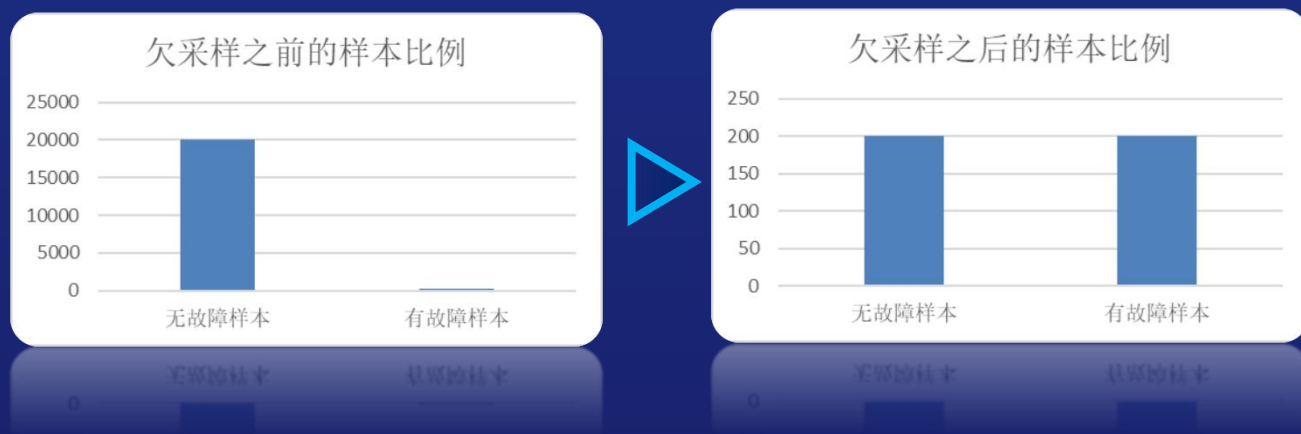
挑战1： 原数据集直接转化为有监督学习的训练集存在样本稀疏、类别不平衡问题，同时故障标注数据只有故障开始时间，不能直接作为有监督学习模型的因变量；

应对1：

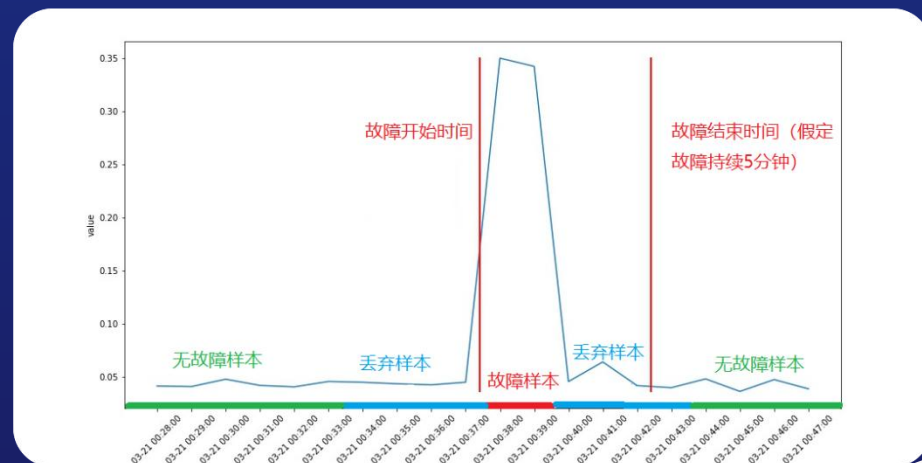


挑战1： 原数据集直接转化为有监督学习的训练集存在样本稀疏、类别不平衡问题，同时故障标注数据只有故障开始时间，不能直接作为有监督学习模型的因变量；

应对2： 通过欠采样处理，进一步缓解类别不平衡问题；



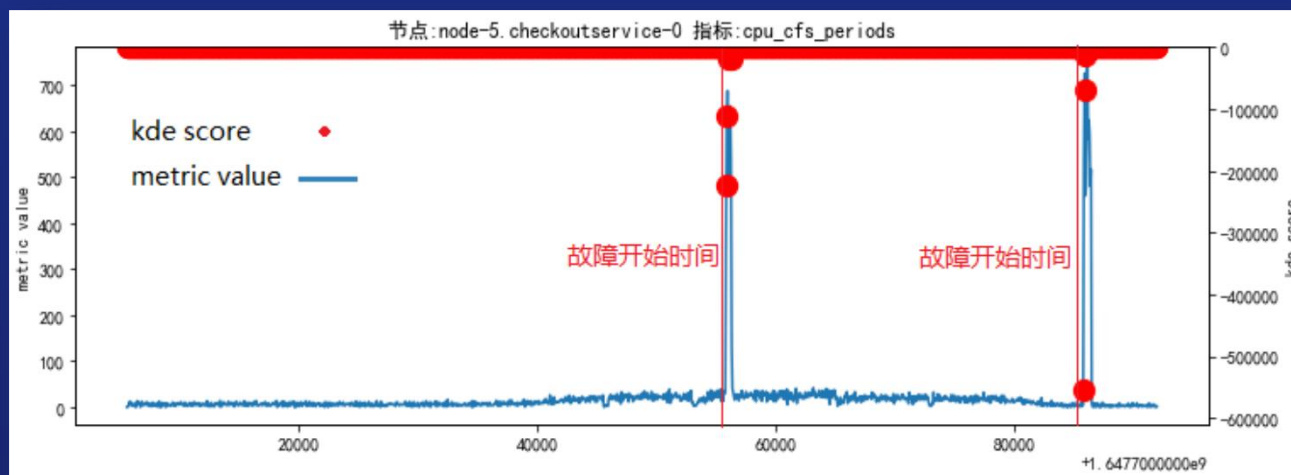
应对3： 丢弃故障初期、末期样本，避免指标异常不明显，污染有故障样本；丢弃故障结束后一段时间内样本，避免指标异常在故障结束后未立即恢复，污染无故障样本；



挑战2：无监督学习算法的通用性要求

应对1：故障样本与非故障样本kde score存在显著差异的指标作为有效指标；

应对2：根据故障样本kde score的分布自动生成异常检测阈值；



挑战3：对故障定位高实效性要求

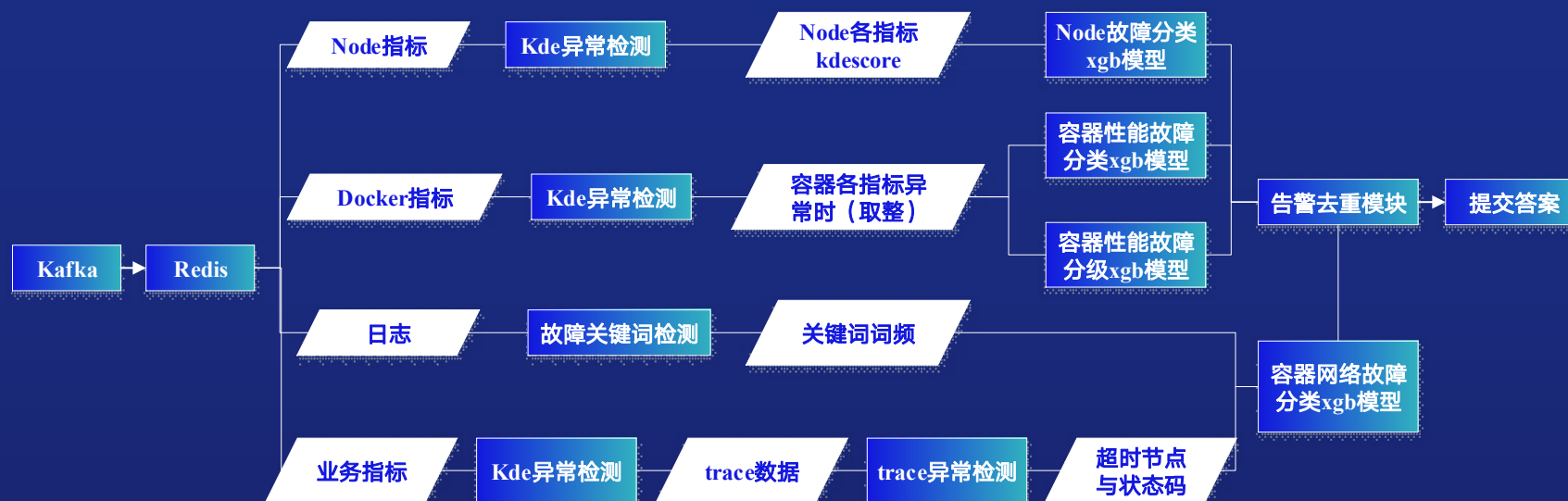
应对1：多线程并行化计算，每个线程单独处理一类数据与故障；

应对2：高效无监督学习算法前置，只有检测到关键指标出现异常时才唤起有监督学习算法；

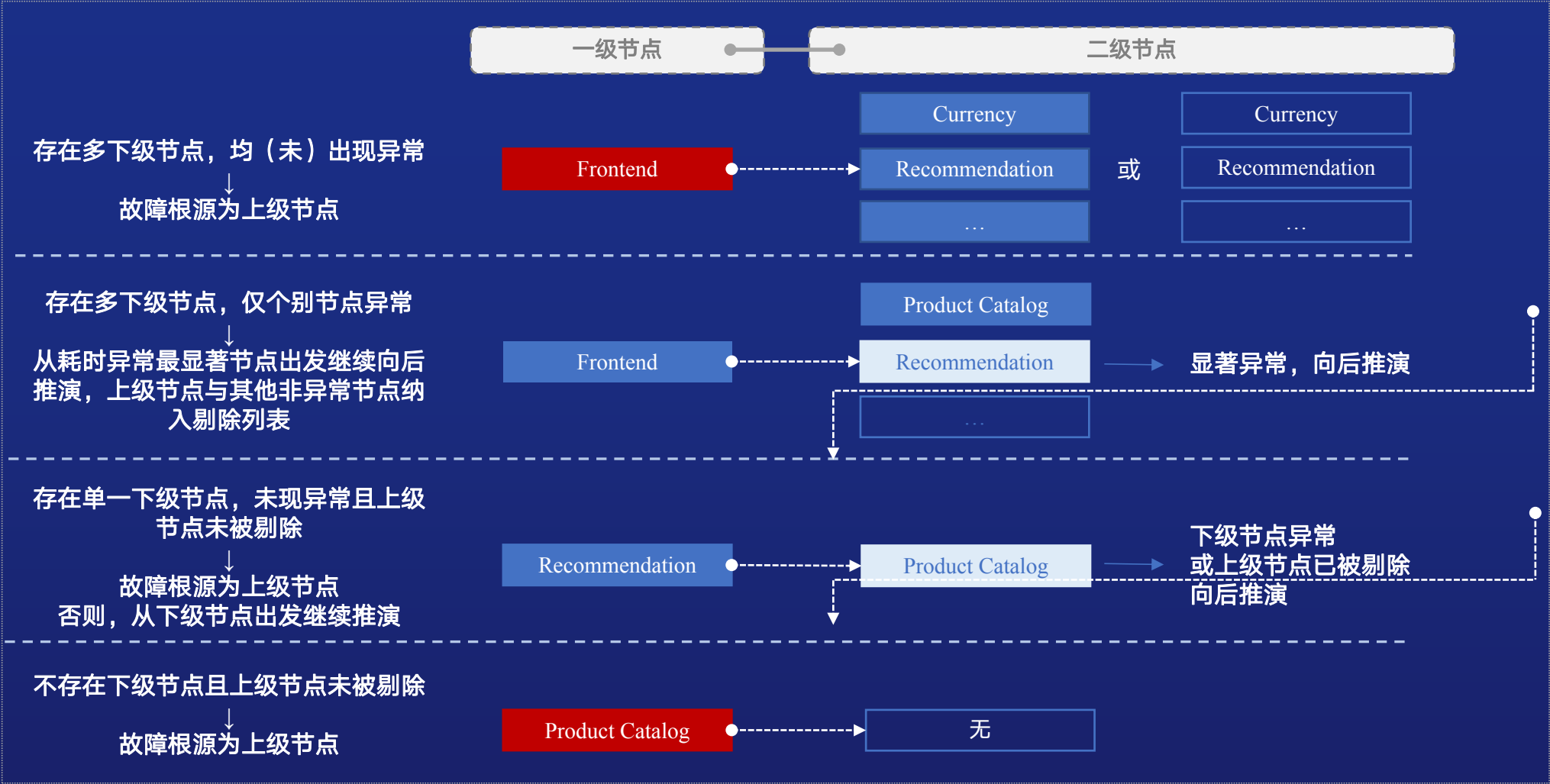
第三章



总体架构



trace异常检测



第四章



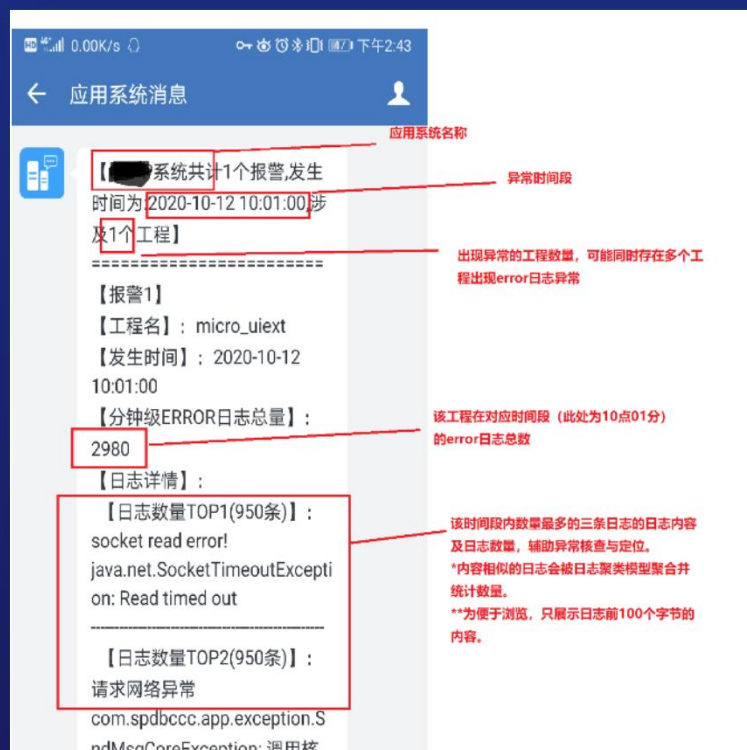
总结与展望

总结

团队复赛排名第二，容器性能故障识别召回率93%；服务器性能故障识别召回率86%。

落地应用优化方向

1) 利用浦发信用卡基于日志数据的故障检测模型，提取日志异常特征，改善日志特征的通用性；



应用系统消息

应用系统名称

【系统】系统共计1个报警,发生时间为2020-10-12 10:01:00,涉及1个工程

异常时间段

出现异常的工程量,可能同时存在多个工程出现error日志异常

【报警1】

【工程名】: micro_uiext

【发生时间】: 2020-10-12 10:01:00

【分钟级ERROR日志总量】:

2980

该工程在对应时间段 (此处为10点01分) 的error日志总数

【日志详情】:

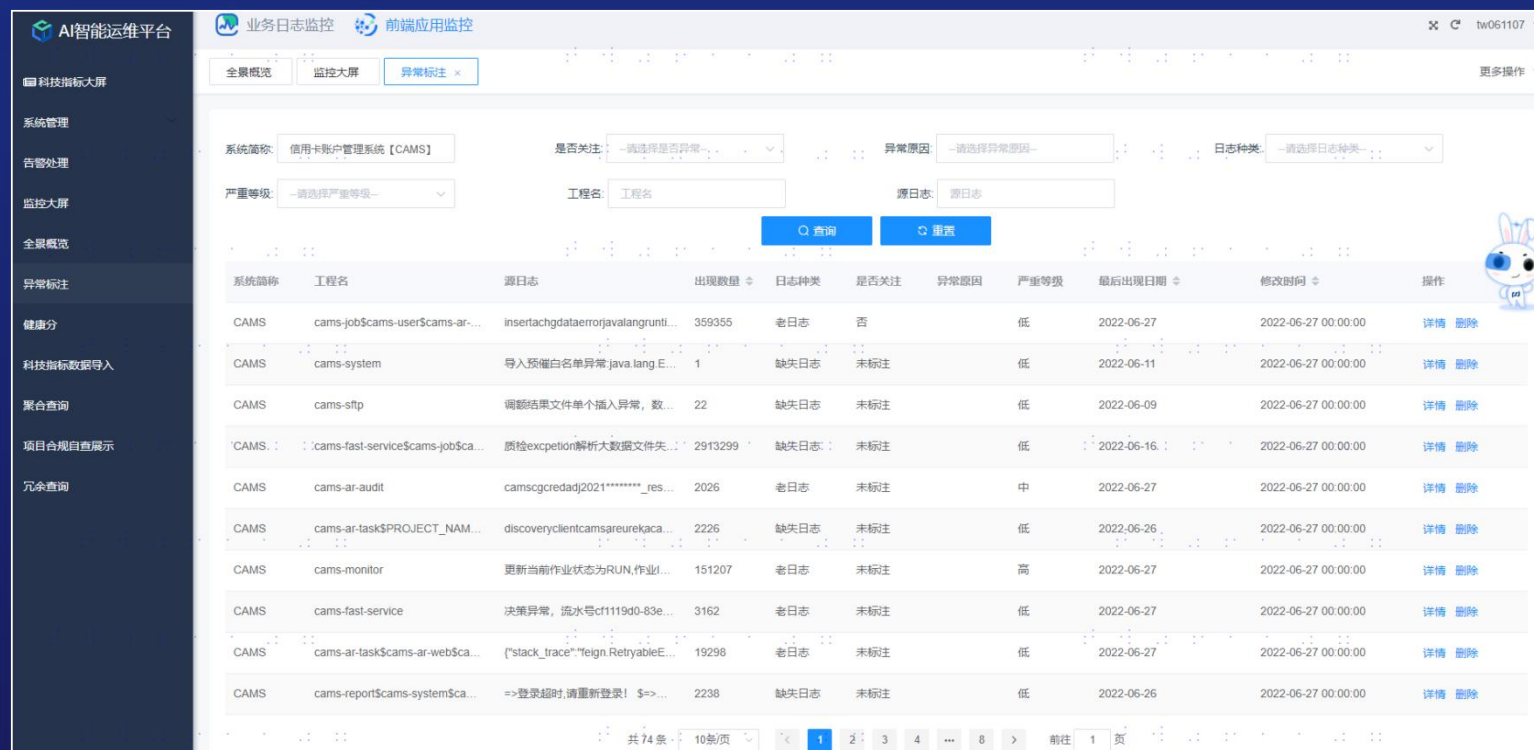
【日志数量TOP1(950条)】:

socket read error!
java.net.SocketTimeoutException: Read timed out

该时间段内数量最多的三条日志的日志内容及日志数量,辅助异常核查与定位。
*内容相似的日志会被日志聚类模型聚合并统计数量。
**为便于浏览,只展示日志前100个字节的内容。

【日志数量TOP2(950条)】:

请求网络异常
com.spdbccc.app.exception.SndMsgCoreException: 调用核



AI智能运维平台

业务日志监控 前端应用监控

全景概览 监控大屏 异常标注

系统名称: 信用卡账户管理系统【CAMS】 是否关注: 请选择是否异常 异常原因: 请选择异常原因 日志种类: 请选择日志种类

严重程度: 请选择严重程度 工程名: 工程名 源日志: 源日志

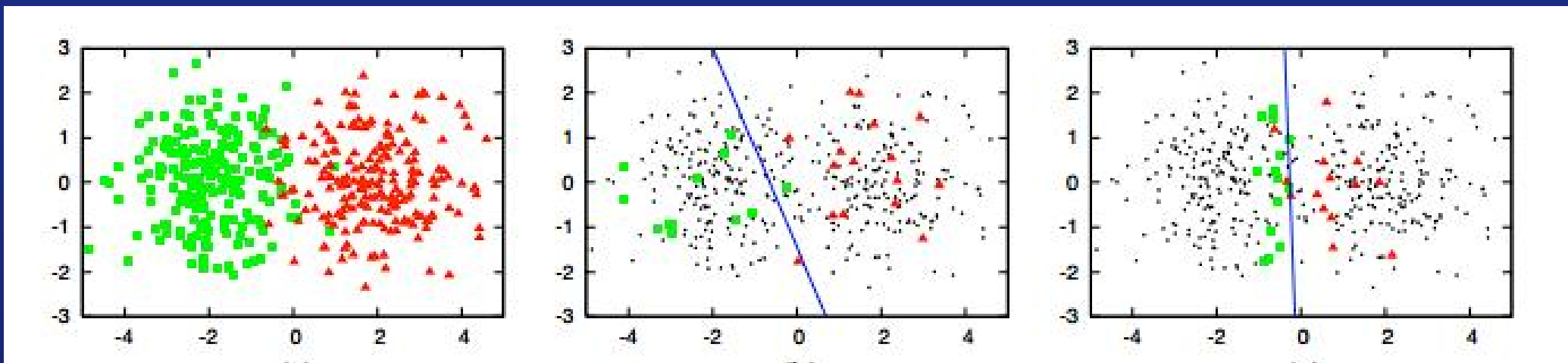
查询 重置

系统名称	工程名	源日志	出现数量	日志种类	是否关注	异常原因	严重程度	最后出现日期	修改时间	操作
CAMS	cams-jobScams-userScams-ar...	insertchgdataerrorjavlangrunti...	359355	老日志	否		低	2022-06-27	2022-06-27 00:00:00	详情 删除
CAMS	cams-system	导入预催白名单异常 java.lang.E...	1	缺失日志	未标注		低	2022-06-11	2022-06-27 00:00:00	详情 删除
CAMS	cams-sftp	调额结果文件单个插入异常,数...	22	缺失日志	未标注		低	2022-06-09	2022-06-27 00:00:00	详情 删除
CAMS	cams-fast-serviceScams-jobSc...	质控exception解析大数据文件失...	2913299	缺失日志	未标注		低	2022-06-16	2022-06-27 00:00:00	详情 删除
CAMS	cams-ar-audit	cmsgcredad2021*****_res...	2026	老日志	未标注		中	2022-06-27	2022-06-27 00:00:00	详情 删除
CAMS	cams-ar-taskPROJECT_NAM...	discoveryclientcamsareurekaca...	2226	缺失日志	未标注		低	2022-06-26	2022-06-27 00:00:00	详情 删除
CAMS	cams-monitor	更新当前作业状态为RUN,作业...	151207	老日志	未标注		高	2022-06-27	2022-06-27 00:00:00	详情 删除
CAMS	cams-fast-service	决策异常,流水号cf119d0-83e...	3162	老日志	未标注		低	2022-06-27	2022-06-27 00:00:00	详情 删除
CAMS	cams-ar-taskScams-ar-webSc...	("stack_trace":"feign.RetryableE...	19298	老日志	未标注		低	2022-06-27	2022-06-27 00:00:00	详情 删除
CAMS	cams-reportScams-systemSc...	=>登录超时,请重新登录! \$=>...	2238	缺失日志	未标注		低	2022-06-26	2022-06-27 00:00:00	详情 删除

共 74 条 - 10 条/页

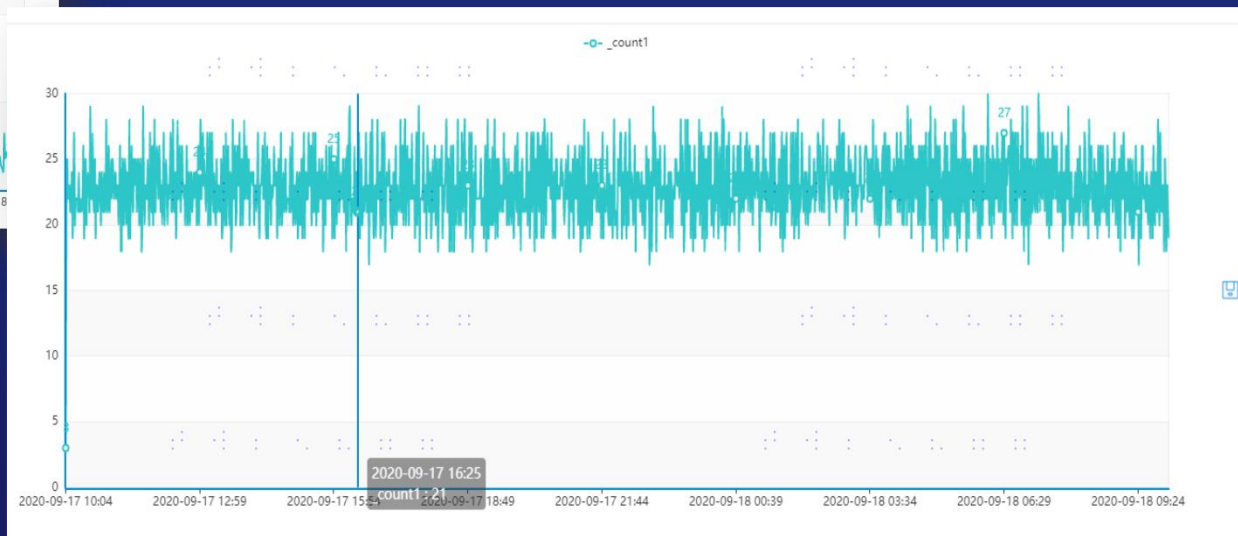
落地应用优化方向

2) 采用聚类算法，对故障样本进行预分类再进行标注，或采用主动学习，减少专家标注成本；



落地应用优化方向

3) 指标异常检测算法优化, 根据指标数据特征, 如是否具有周期性、是否稳定等对指标进行分类, 不同类型的指标采用不同的异常检测算法;





2022 CCF国际AIOps挑战赛决赛暨AIOps研讨会

THANKS