OpenAIOps

AIOPS | 2025 CCF国际AIOps挑战赛
2025 CCF International AIOps Challenge

基于大模型智能体的微服务根因定位

队伍：xwy6475　　杨康

内蒙古大学

主办单位：中国计算机学会（CCF）
承办单位：中国计算机学会互联网专委会、中国科学院计算机网络信息中心、中国移动研究院、清华大学
协办单位：华为2012实验室、阿里云、中兴通讯、中国移动九天团队、南开大学、西安电子科技大学、清华大学计算机科学与技术系、神州灵云

OpenAIOps AIOPS | 2025 CCF国际AIOps挑战赛
2025 CCF International AIOps Challenge

目录
CONTENTS

第一章节

准确性

准确性

模型：QWQ-32B

初赛成绩

排名	团队	分数
1	FastReject	71.97
2	我们是有技术的	51.95
3	hwlyyzc	51.33
4	xwy6475	50.68
5	Holmes	50.32
6	Roborock-LLM	48.39
7	AISecOps	47.82
8	小鸟吃大蒜(蒜鸟)	46.87
9	十二楼	45.55
10	男团910	45.01

复现成绩

排名	团队	分数
1	FastReject	73.95
2	xwy6475	54.62
3	hwlyyzc	54.60
4	我们是有技术的	49.64
5	Holmes	48.54
6	男团910	48.52
7	小鸟吃大蒜(蒜鸟)	45.82
8	十二楼	45.68
9	可观测战队	44.77
10	baseline----	39.91

第二章节

方案设计

总体方案设计

01 **异常检测预处理**

融合Metrics、Logs、Traces数据，通过离线学习正常样本特征，实现故障期间数据的异常检测与统一告警。

02 **实时拓扑构建**

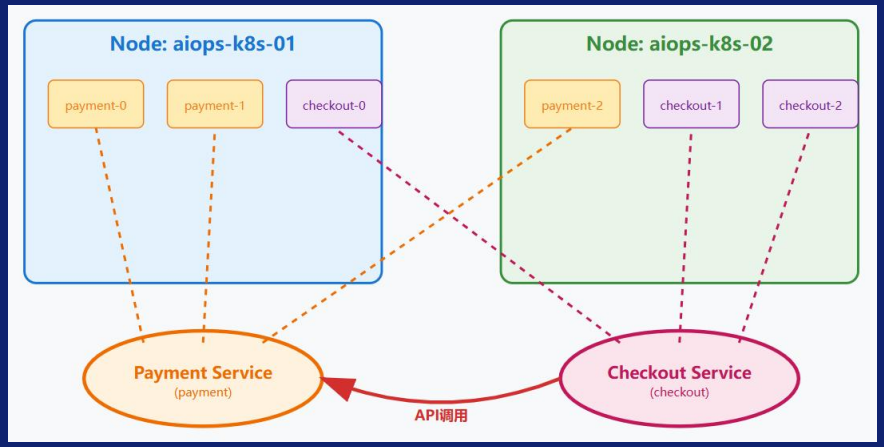
挖掘Pod-Node实时运行关系，构建Service-Pod-Node三层拓扑结构。

03 **告警拓扑根因定位**

通过组件依赖关系，分析故障传播路径，定位引发连锁故障的源头。

04 **智能体自动排障**

采用Agent/LangGraph框架，实现端到端的自主排障。



MCP 工具

Trace异常检测

- Trace串联
- Span自耗时突变检测（基线对比）
- 调用链错误span定位

Log异常检测

- 正常样例Drain聚类建立模板基线
- 新颖性检测（new template识别）
- 模板频率突增检测（基线q95+Δ）

Metric异常检测

- Mann-Whitney U / Welch t-test统计检验
- 基线阈值越界检测（均值±k · std）
- 标准化偏差与方差比检验
- STL季节性分解剥离趋势与周期

实时拓扑获取

- Pod-Node运行时关系挖掘
- Service-Service调用关系
- 三层拓扑图构建与更新

Agent 智能编排架构设计

核心组件架构（单Agent）

ReAct Agent核心

- LangGraph工作流编排
- 状态管理与节点控制

工具注册机制

- 异常检测工具包装
- Pod-Node关系获取

会话管理

- 多轮对话追踪
- Thread ID状态保持

确定性推理

- Temperature=0
- Seed=0固定输出

推理策略与优化

- **多模数据分析**：整合trace、log、metric三类监控数据进行综合故障定位
- **专有排障规则**：判定Pod、Service、Node级故障的方法，建立清晰的故障层次模型
- **优先级决策规则**：基于异常数量和严重程度制定组件识别的优先级策略
- **内置排障知识库**：预定义CPU过载、JVM压力、网络故障等常见故障类型和诊断规则
- **微服务架构映射** - 详细描述服务调用关系和Pod-Node部署关系，用于根因推理
- **标准化输出要求**：包含组件、原因、推理链路三要素，输出JSON

案例

question: Determine the root cause of the system fault during 2025-06-06T13:03:21Z to 2025-06-06T13:03:22Z.

trace 异常：


```
[
  {
    "anomaly_type": "error",
    "span_kind": "client",
    "component": "recommendationservice",
    "count": 54,
    "avg_cost_ms": 58868.15,
    "messages": ["context canceled"]
  }
]
```


log 异常：


```
{
  "pod": "recommendationservice-0",
  "logs": [
      "initializing conf files",
      "execute pod-init.sh ...",
      "product catalog address: productcatalogservice:3550",
      "initializing recommendationservice",
      "listening on port: 8080",
      "TINGYUN_APP_NAME=recommendationservice(hipstershop)"
  ],
  "count": 16
},
{
  "pod": "recommendationservice-1",
  "logs": [
      "initializing conf files",
```


诊断结果：


```
{
  "component": "recommendationservice",
  "reason": "pod kill: execute pod-init.sh, context canceled",
  "reasoning_trace": [
    {
      "step": 1,
      "action": "TraceAnalysis",
      "observation": "recommendationservice has 54 trace errors with context canceled"
    },
    {
      "step": 2,
      "action": "LogAnalysis",
      "observation": "all 3 pods show pod-init.sh execution logs"
    }
  ]
}
```

第三章节

创新性和实用性

创新性

多模数据异常检测

- Metric、Log、Trace统一处理为异常告警
- 故障前后数据突变分析，全面捕获根因异常特征
- 多检测器投票机制：集成Robust Z-Score、ESD、频域残差等多种检测方法，少数服从多数

实时拓扑根因定位

- 实时拓扑构建：挖掘Pod-Node运行关系，构建Service-Pod-Node三层图结构
- 图上证据传播：在拓扑结构上传播异常置信度，结合相关性、时间一致性进行评分
- 多维度归因：支持服务级、Pod级、节点级的分层归因与降级策略

Agent设计

- 专业算法集成：Agent可调用Metric、Log、Trace专门的异常检测算法
- 自主代码生成（备选）：当未发现明显异常时，Agent可自行编写代码分析数据
- 原始数据直接分析（备选）：Agent直接对Trace、Log、Metric原始数据片段进行自主分析

实用性

生产落地适应性

双模式支持：实时在线推理与离线批量评测一体化

云原生兼容：深度适配K8s环境，支持容器化部署

稳定可靠：多检测器投票机制，显著降低误报率

高效分析能力

快速响应：单次分析10-30秒，大幅提升故障处理效率

并行处理：多模态数据并行检测，充分利用计算资源

智能降级：数据缺失时自动切换备用策略

运维友好特性

模块化架构：组件可插拔，便于功能迭代升级

可视化输出：结构化结果与图形界面直观验证

易于集成：标准API接口，快速接入现有监控体系

第四章节

未来展望

未来展望

Human in Loop

- 反馈闭环：集成专家标注与运维反馈，持续优化模型准确率
- 交互式分析：支持人机协同，运维专家可实时修正分析结果

RAG领域知识增强

- 知识库构建：集成运维手册、故障案例等领域知识
- 上下文增强：基于历史故障模式，提供更精准的根因推理

多Agent协同优化

- 专家分工：不同Agent负责特定模态或业务域分析
- 层次协作：一个主管Agent协调多个专业Agent

异常检测算法升级

- 深度学习：引入时序Transformer、图神经网络等先进算法
- 自适应学习：动态调整检测阈值，适应业务模式变化

OpenAIOps AIOPS 2025 CCF国际AIOps挑战赛
2025 CCF International AIOps Challenge

THANKS

主办单位：中国计算机学会（CCF）

承办单位：中国计算机学会互联网专委会、中国科学院计算机网络信息中心、中国移动研究院、清华大学

协办单位：华为2012实验室、阿里云、中兴通讯、中国移动九天团队、南开大学、西安电子科技大学、清华大学计算机科学与技术系、神州灵云