AIOps Challenge

基于时间序列的解决方案

benjili20224109/李冕正
微众银行

2022 CCF国际AIOps挑战赛决赛暨AIOps研讨会

团队介绍

来自微众银行智能运维团队，队员从左到右分别是：全栈工程师-曹旭东、全栈工程师-林佳宇、全栈工程师-张紫婷、全栈工程师-李冕正、全栈工程师-王国峰

赛题分析及解决方案

AIOps Challenge

原始数据

日均1500万+日志，1000万+调用链消息、700万+metric数据

曲线异动

对时间序列进行异常检测，提取其中的异动，作为异常决策和根因定位的证据

时间序列

用时间序列描述从原始数据中提取出来的所有特征信息，并进行时间序列异常检测

故障根因

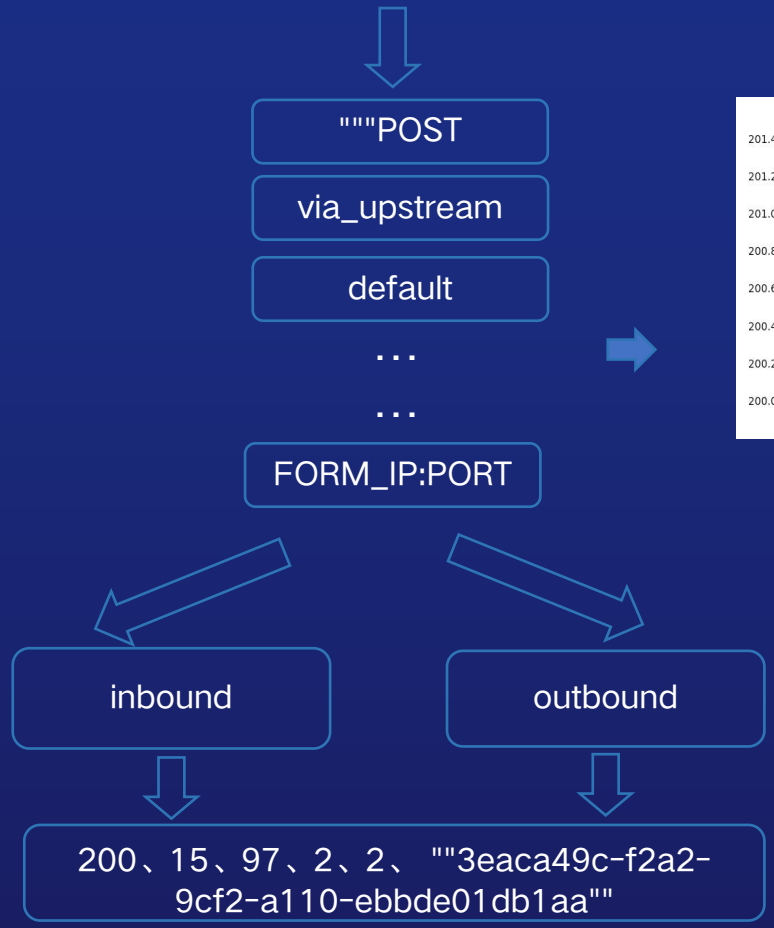
综合链路异动、日志异动、KPI异动以及其上下游关系，推断最终的故障根因

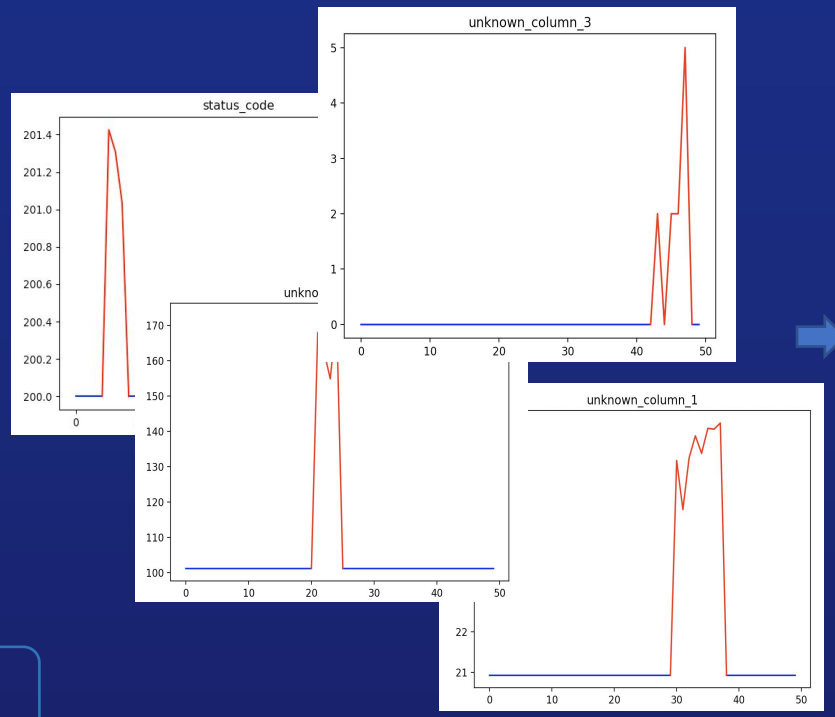
日志信息挖掘

AIOps Challenge

利用词频创建特征树，将特征树转化为时间序列

"""POST /hipstershop.AdService/GetAds HTTP/2"" 200 - via_upstream - ""-"" 15 97 2 2 ""-"" ""grpc-go/1.31.0"" ""3eaca49c-f2a2-9cf2-a110-ebbde01db1aa"" ""adservice:9555"" ""172.20.3.4:9555"" inbound|9555|| 127.0.0.6:60627 172.20.3.4:9555 172.20.3.12:52936 outbound_.9555_._.adservice.ts.svc.cluster.local default"
"""POST /api/v2/spans HTTP/1.1"" 202 - via_upstream - ""-"" 306 0 0 0 ""-"" ""okhttp/3.14.7"" ""2eda6e95-b457-92b1-bcaa-565437d2865f"" ""jaeger-collector:9411"" ""172.20.4.66:9411"" outbound|9411||jaeger-collector.ts.svc.cluster.local 172.20.3.4:38766 10.68.163.189:9411 172.20.3.4:52446 - default"
"warning envoy config StreamAggregatedResources gRPC config stream closed: 14, connection error: desc = ""transport: authentication handshake failed: context deadline exceeded"""

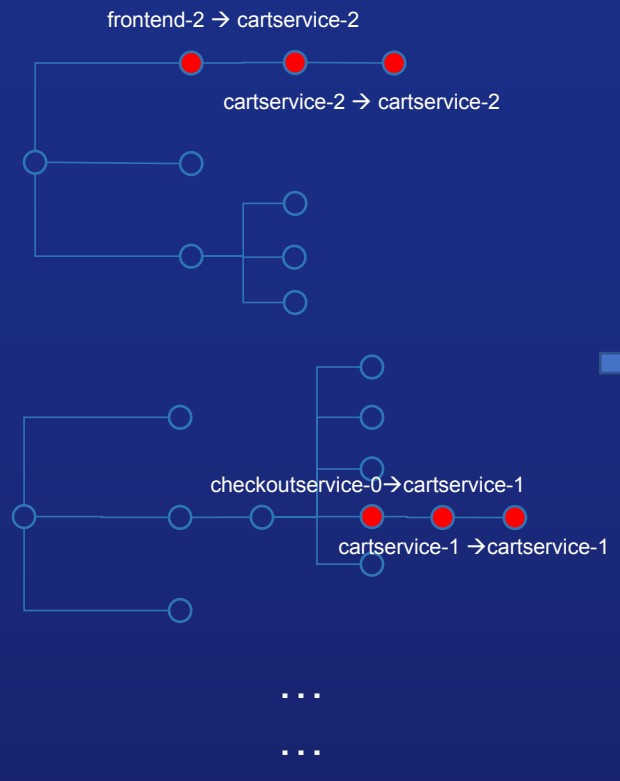


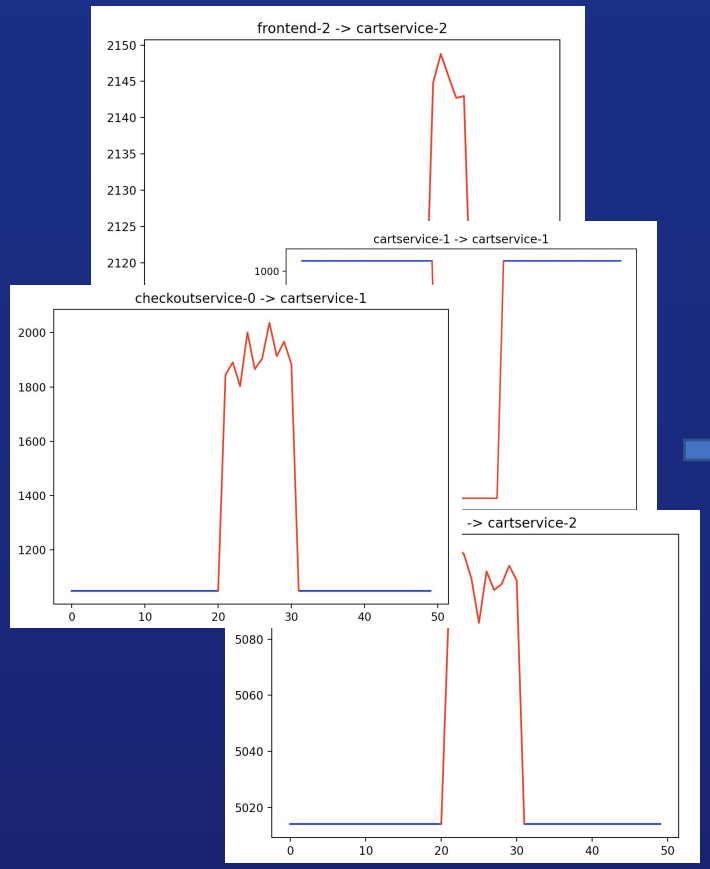


	10:01	10:02	10:03
status_code	⇧ 1.4	⇧ 1.2	⇧ 1.1
unknown_column_3	⇧ 2	---	⇧ 2
unknown_column_1	⇧ 170	⇧ 160	⇧ 150
...	...	...	...

调用链信息挖掘

AIOps Challenge



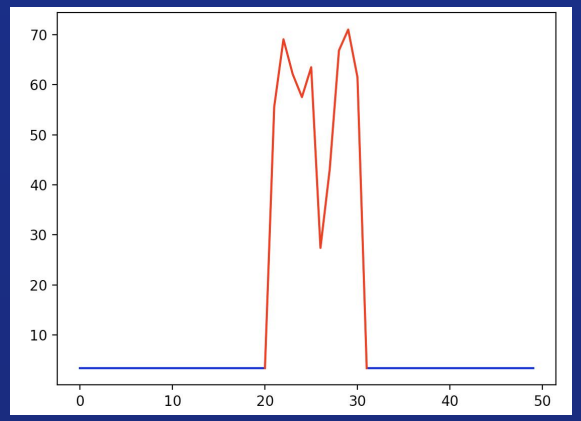


	10:01	10:02	10:03	10:04
frontend-2 → cartservice-2	⇑1390	⇑1380	⇑1370	⇑1370
cartservice-2 → cartservice-2	⇑1200	⇑1100	⇑1300	⇑1000
cartservice1→ cartservice1	⇓1000	⇓1000	⇓1000	⇓1000
...	...	...	...	...

1.并非所有异常都体现在本节点上
2.上游耗时包含下游耗时

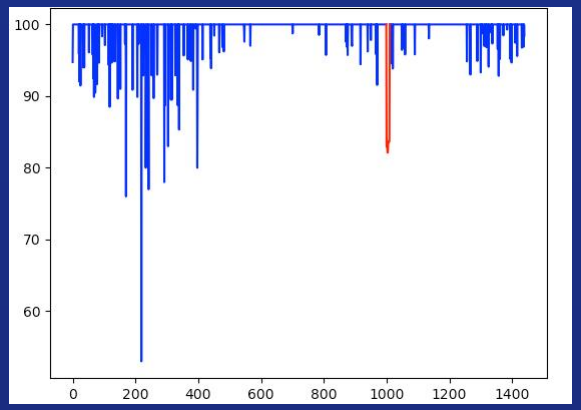
1.以消息对形式检测
2.上游节点耗时减掉所有下游节点耗时之和

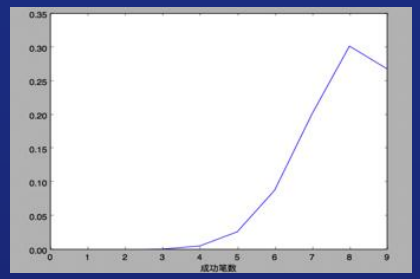
时间序列异常检测

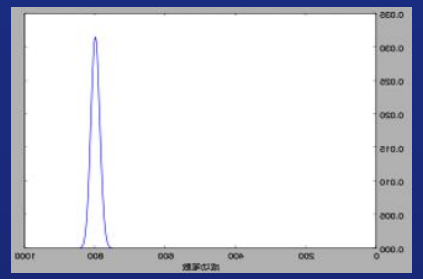


$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

0







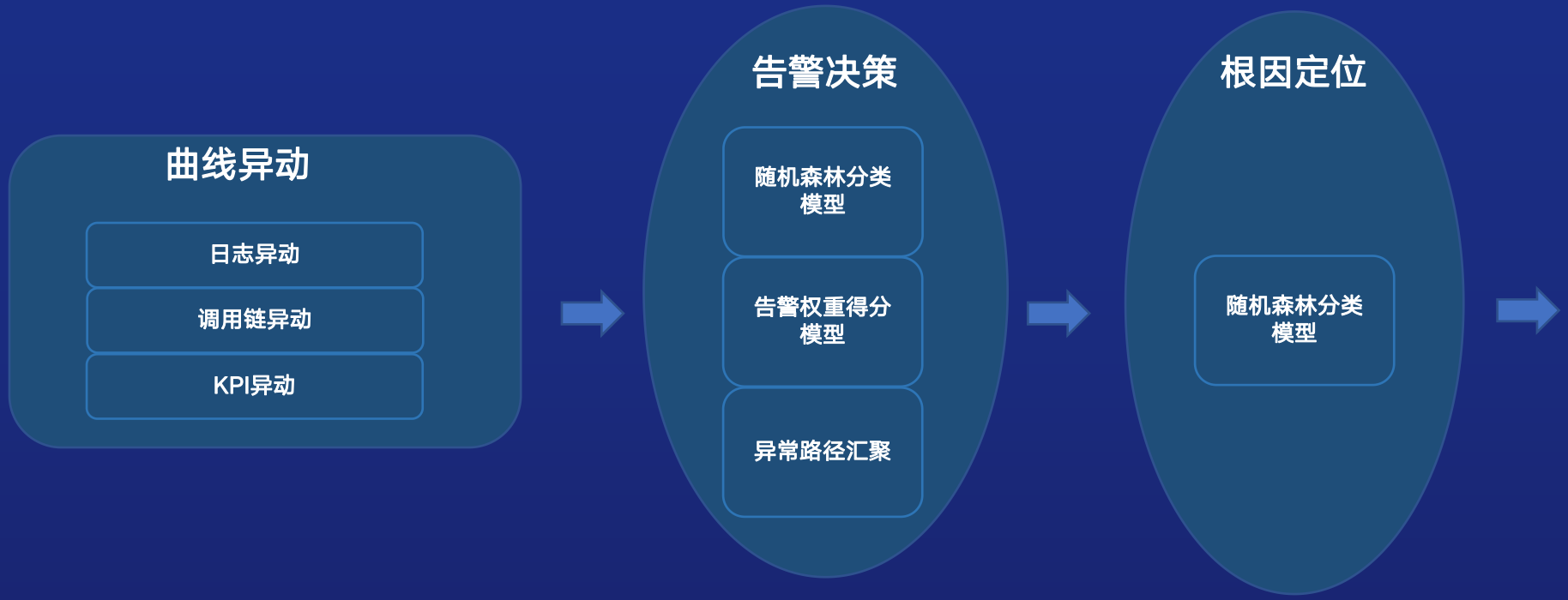
高斯分布

采用普通高斯分布，用数据分布的概率进行异常检测

伯努利二项分布

部分布尔型或者按照少量种类分布的数据，使用伯努利二项分布，可以更好地放大异常。

告警决策及根因定位



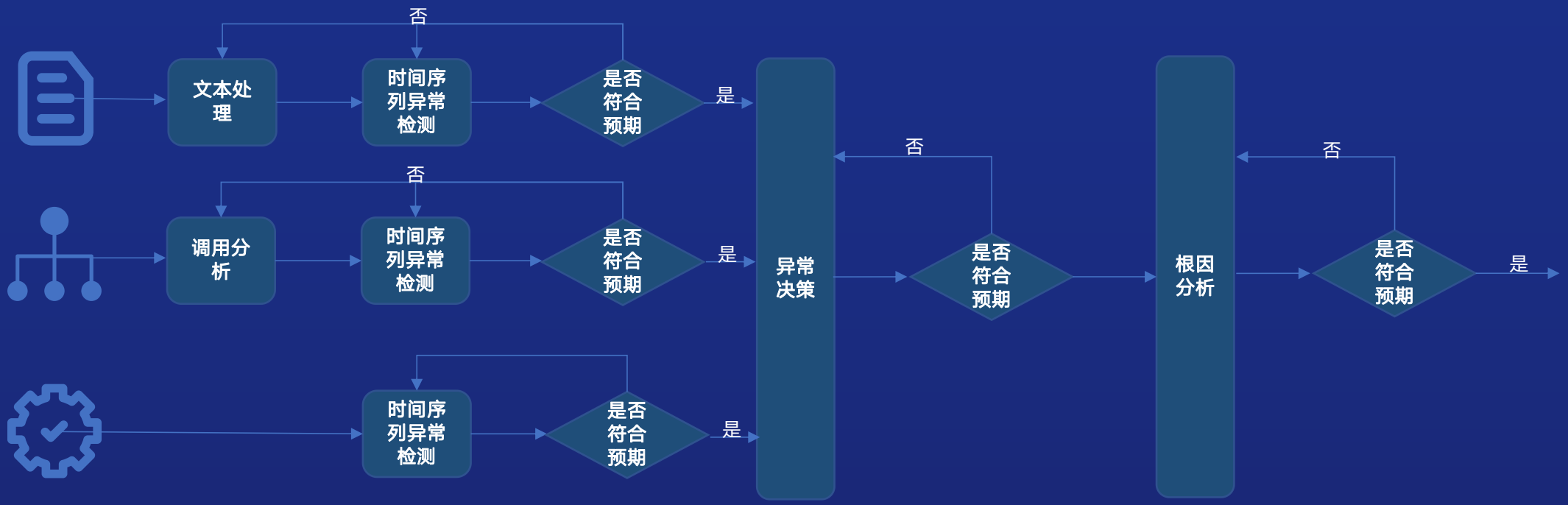
告警权重得分模型

自定义得分模型，根据指标的历史异动情况计算其权重，跟历史故障匹配度越高，同时跟非异常情况下抖动次数越少,则权重越高

异常链路汇聚

该方法只用于调用链判断，主要是根据调用链上下游的抖动情况，计算cmdb_id节点的汇聚度，通过汇聚度决策是否异常。

总体运转流程



1.松耦合、多闭环，隔离影响同时提升算法调试效率
2.多种数据源统一转化为时序检测，既保证了上层的灵活性，也兼顾开发的便捷性
3.全流程可视化，数据分析更加直观

不足及改进

AIOps Challenge

1.数据特征需要进一步扩展

2.融入专家经验,结合系统框架进行针对性监控

3.注重时效、细化策略，将数据特征融入到告警决策中

AIOps Challenge

2022 CCF国际AIOps挑战赛决赛暨AIOps研讨会

THANKS