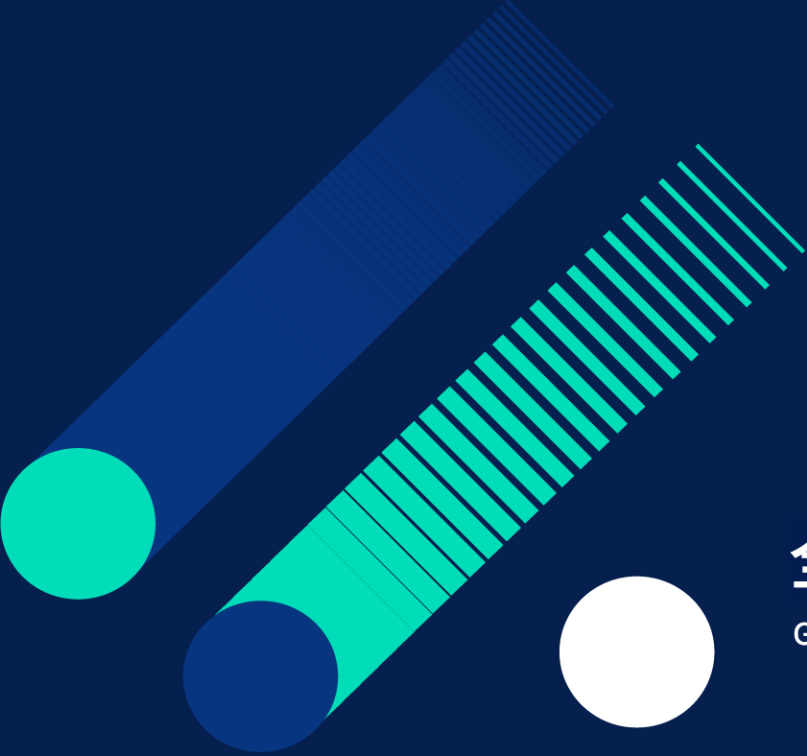




GIAC

全球互联网架构大会

GLOBAL INTERNET ARCHITECTURE CONFERENCE

AIOps入门与实践案例分享

李剑锋 腾讯CSIG Metis智能运维团队负责人

100

第七届 全球软件案例研究峰会

运营增长

组织发展

爆款架构

人工智能

测试实践

运维体系

产品创新

研发效能

数据平台

AI/数据驱动

测试体系

AI/ops DevOps

体验设计

团队管理

工程实践

AI实践

大前端

区块链



扫描二维码
查看最值得学习的100+创新案例

2018年11月30日-12月3日
北京 | 国家会议中心

主办方 **msup**

上海精品公开课

管理3.0认证课程

时间：12月22-23日 | 地点：上海 | 讲师：林伟丹



大数据及AI挖掘技术

时间：12月22-23日 | 地点：上海 | 讲师：风清扬



高可用架构与设计

时间：01月12-13日 | 地点：上海 | 讲师：沈老师



K8S与service mesh

时间：01月12-13日 | 地点：上海 | 讲师：Jim



备注：扫码查看课程详情，两人以上报名有优惠，详情咨询：15802217295

msup[®]



目录



1

AIOps入门

2

腾讯AIOps实践案例

3

学件与Metis



AI Ops是什么 (VS DevOps)

- ◆ 目的都是为了从“人肉”运维到“自动”运维的转变
- ◆ 实现方式不一样，传统DevOps基于程序逻辑规则，AI Ops基于数据模型
- ◆ DevOps需要先挖掘发现逻辑，然后实现规则，AI Ops通过数据自动寻找逻辑规则
- ◆ 两者并不矛盾，AI Ops是DevOps的高阶实现，不能也不会取而代之



机器学习关键要素

- ◆ 数据
 - ◆ 对AIOps来说，最重要的是数据，运营环境长期积累的数据是AIOps的前提
 - ◆ 质量上报数据、程序日志、访问数据等等
- ◆ 特征
 - ◆ 用来区分事物的维度
 - ◆ 量化衡量事物的量化标准
 - ◆ 大小（长宽高）、重量、距离、角度等等
- ◆ 算法/模型
 - ◆ 用数学函数表达事物间存在的内在逻辑关系



AIOps算法举例

时间序列异常检测

无监督 **EWMA算法 ; 多项式拟合算法 ; 3-Sigma算法 ; Isolation Forest算法**

有监督 **GBDT ; XGBoost算法**

时间序列趋势预测

Holt-Winters ; ARIMA ; LSTM等

告警收敛

Apriori (啤酒尿布)

根因分析

交叉熵 ; 决策树

根源分析

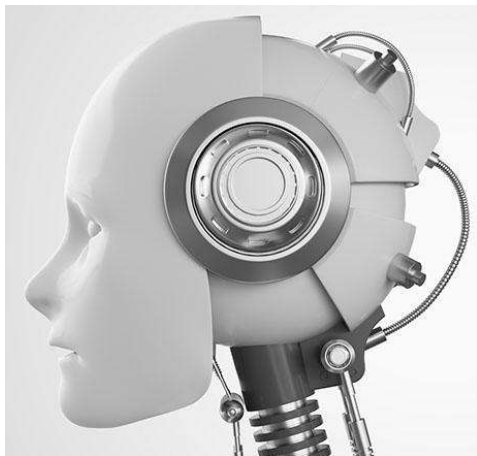
FastUnfolding社团划分算法等

服务器&数据库异常检测

Kmeans聚类 ; DBSCAN聚类 ; Isolation Forest



AIOps场景应用场景



01

质量保障

异常检测、故障诊断、故障预测、故障自愈

02

效率提升

智能变更、智能问答、智能决策、容量预测

03

成本管理

成本优化、资源优化、容量规划、性能优化



AI Ops在场景问题解决中的衡量标准

实际\预测	真	假	汇总
真	A	B	A+B
假	C	D	C+D
汇总	A+C	B+D	总样本

- ◆ **精准率**：分类正确的样本数占总体样本数的比例 $\text{Accuracy} = (A+D)/\text{总样本}$
- ◆ **召回率**：预测结果为1的样本数占真实标签为1的样本数的比例 $\text{Recall} = A/(A+B)$
- ◆ **准确率**：预测结果为1的样本数占总共预测结果为1的比例 $\text{Precision} = A/(A+C)$



AIOps能力框架





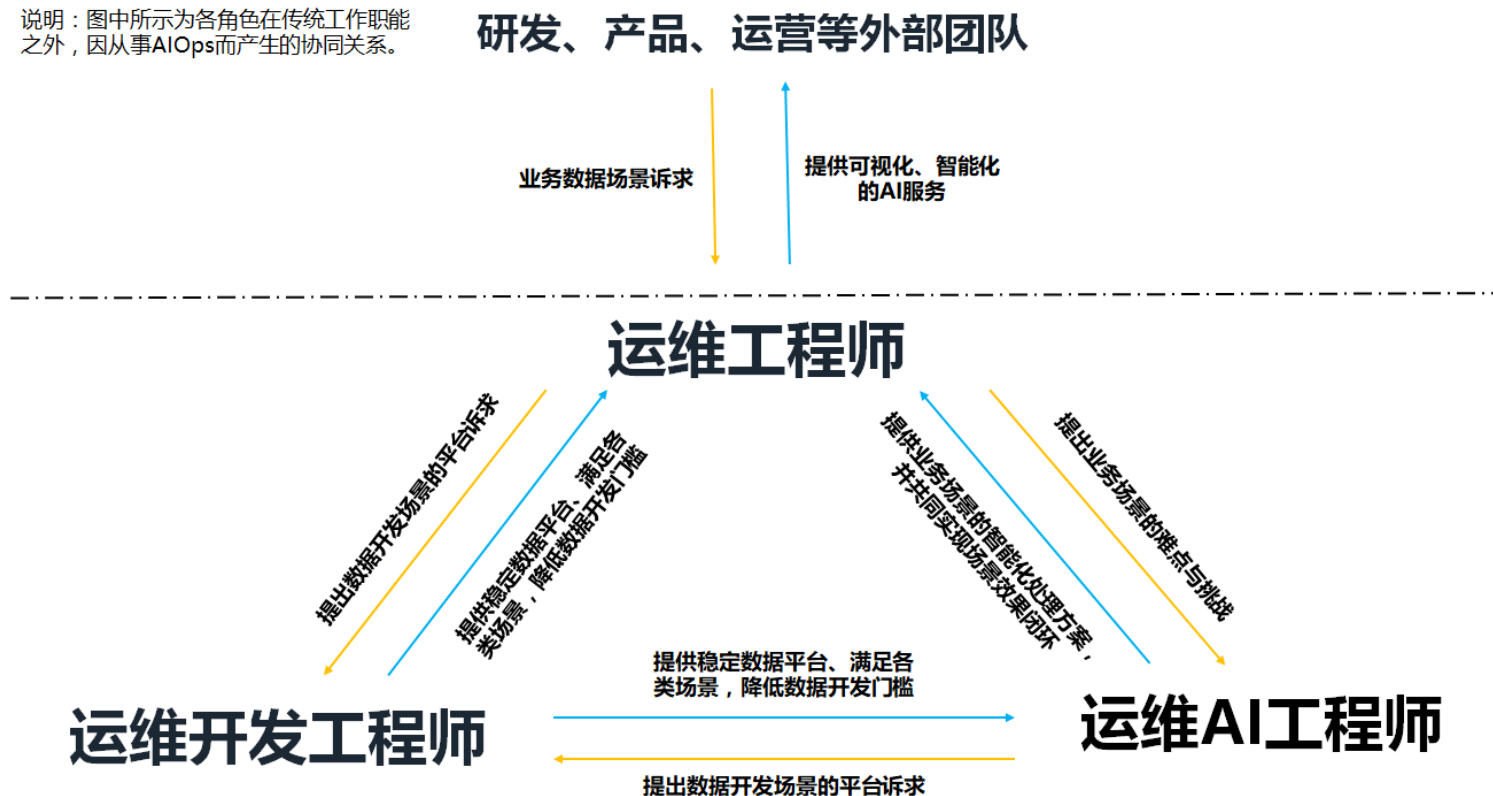
AIOps能力分级





AIOps团队角色

说明：图中所示为各角色在传统工作职能之外，因从事AIOps而产生的协同关系。



图片摘自：《企业级AIOps实施建议白皮书》



目录

1

AIOps入门



2

腾讯AIOps实践案例

3

学件与Metis

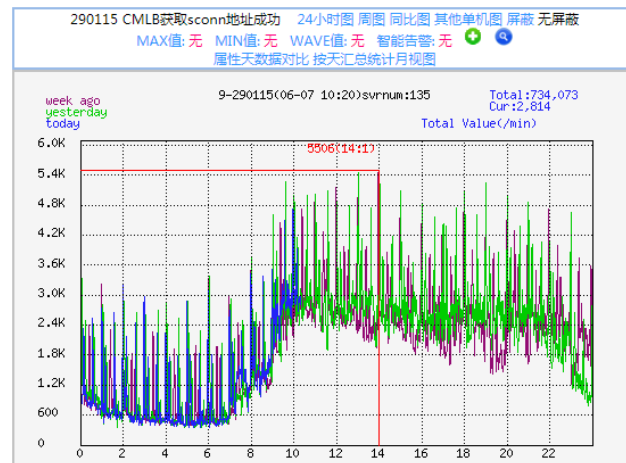
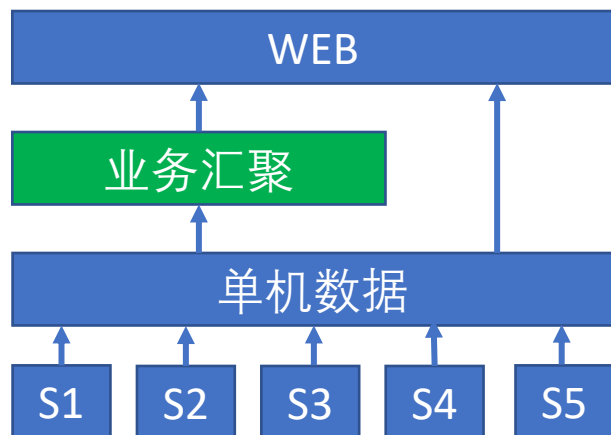


监控指标异常检测-起源

20W+服务器
240W+汇聚指标



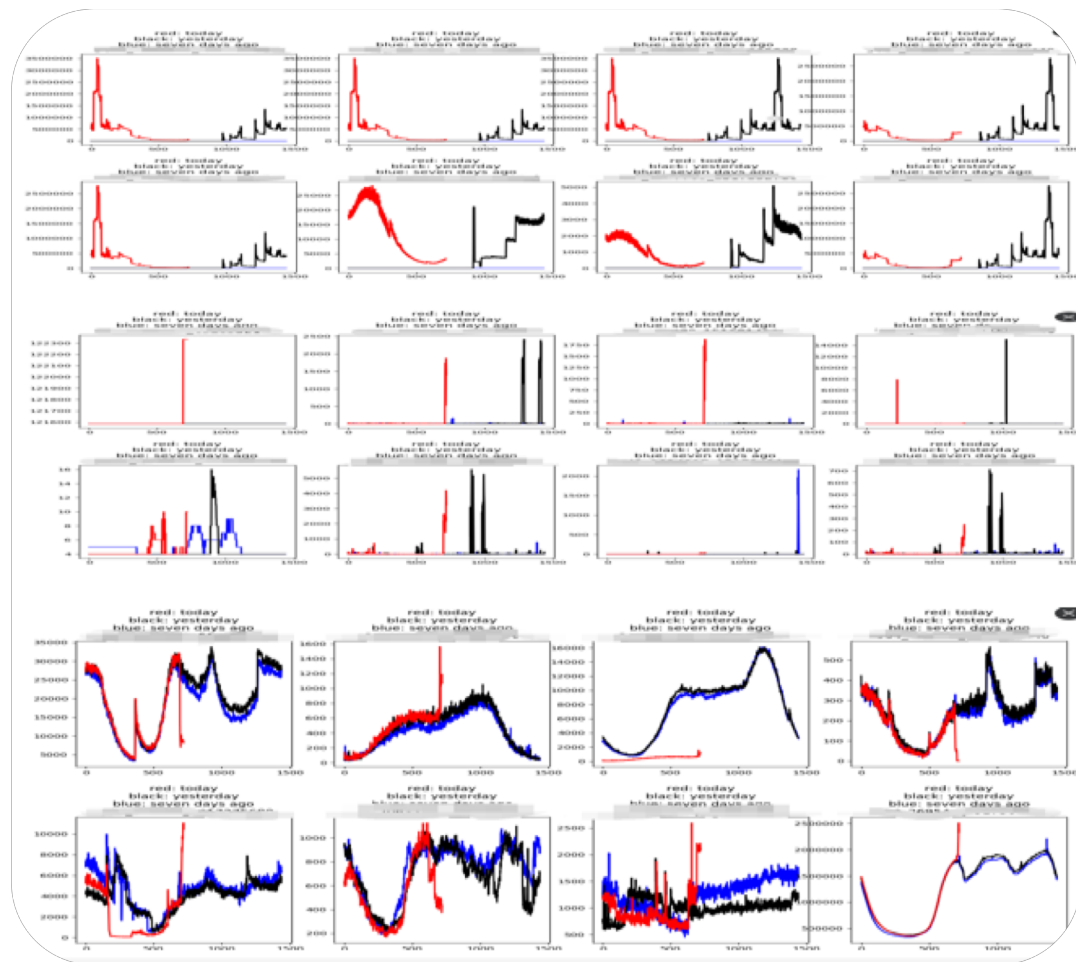
腾讯Monitor流式监控系统基本框架



- ◆ 原有告警方式：阈值设置（最大值、最小值、波动值）
 - ◆ 准确率低（形态各异）
 - ◆ 维护成本高



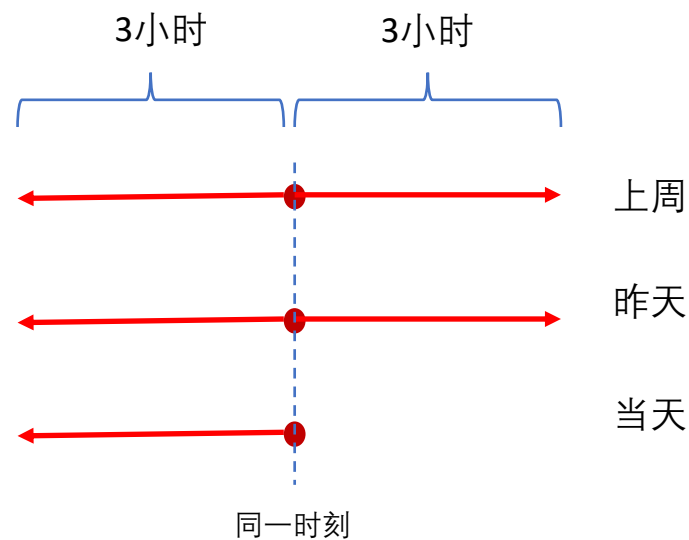
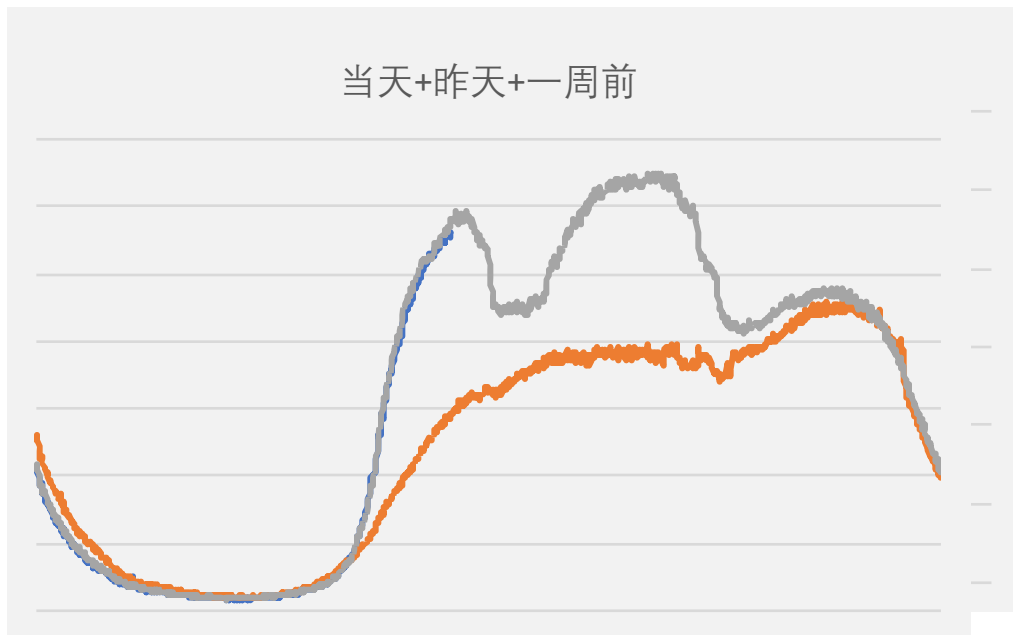
监控指标异常检测-解决思路



- ◆ 曲线复杂，异常情况比较多
- ◆ 核心点：特征的选取
- ◆ AI：让机器模拟人的思维让机器来判断



监控指标异常检测-解决思路

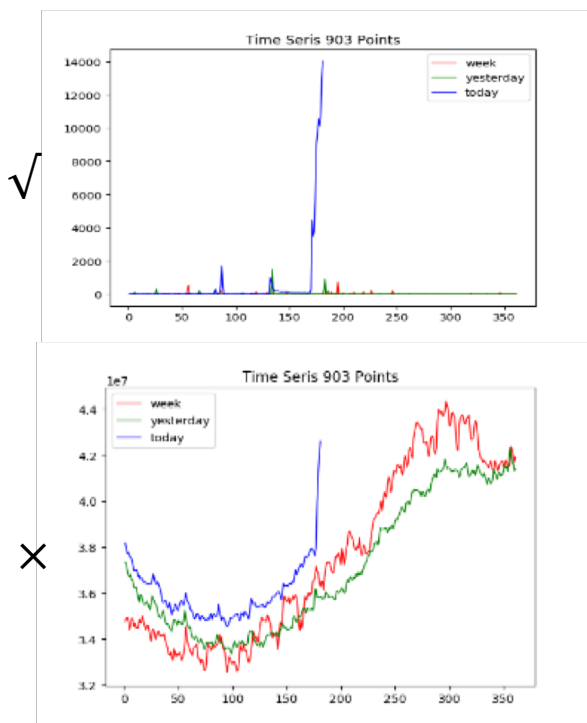


特征数据窗口选择



监控指标异常检测-算法

◆ 传统算法



数据提取

- 以当前时刻为标准
- 七天前后三小时 + 昨天前后三小时 + 今天前三小时

Grubbs

- 3sigma原理

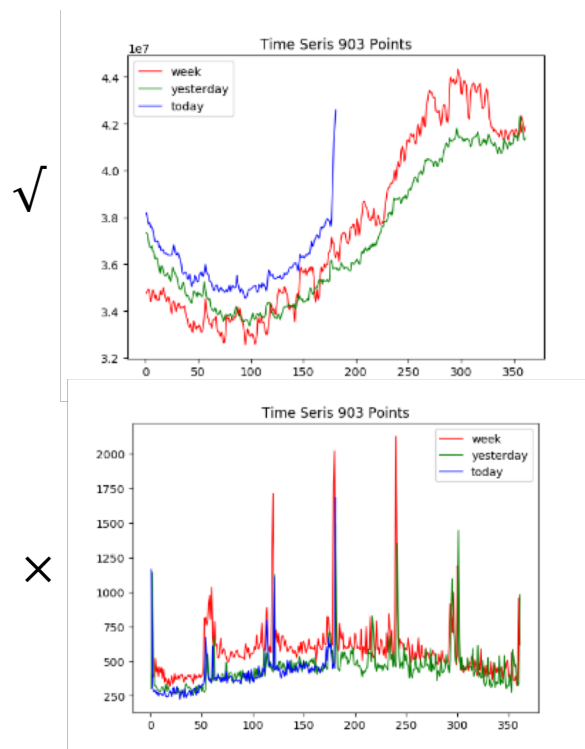
控制图

- 移动平均算法
- 指数移动平均算法



监控指标异常检测-算法

◆ 无监督学习算法



数据提取

- 以当前时刻为标准
- 七天前后三小时 + 昨天前后三小时 + 今天前三小时

孤立森林

- 可以从多维特征中寻找异常点

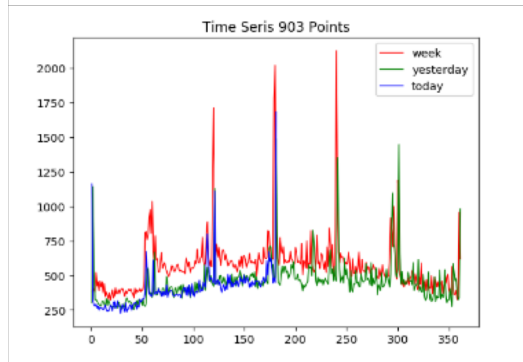
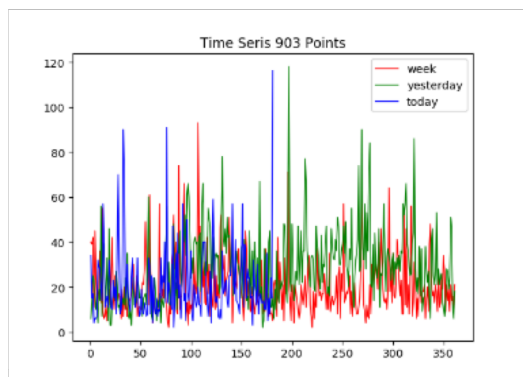
SVM

- 使用超平面的思想来进行异常/正常的区分



监控指标异常检测-算法

◆ 有监督算法



数据提取

- 以当前时刻为标准
- 七天前三小时 + 昨天前后三小时 + 今天前三小时

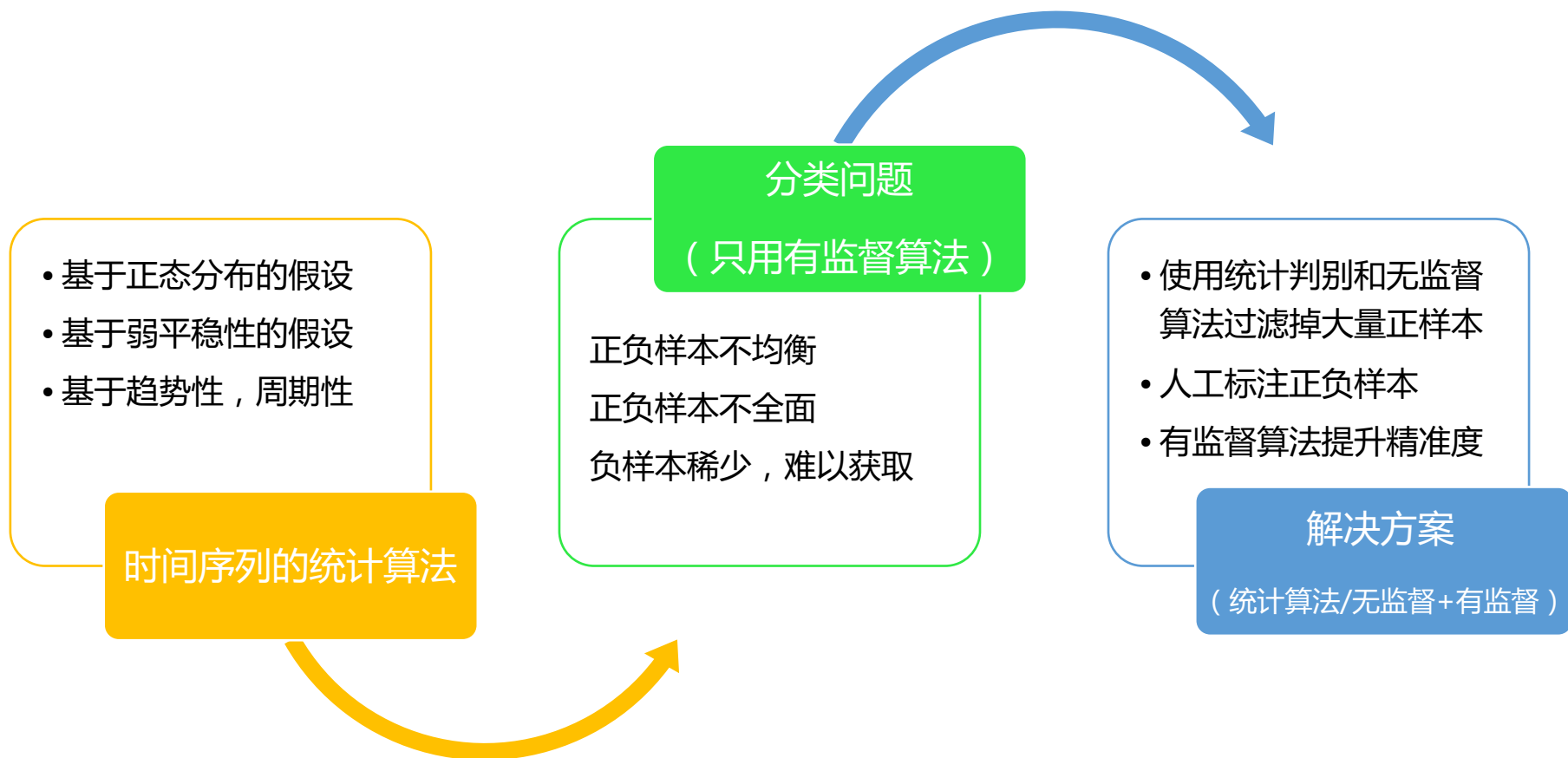
有监督算法

- Linear Regression/Logistic Regression
- Decision Tree/Naïve Bayes
- Random Forest/GBDT/xgboost

有监督算法是持续优化效果的基础，
虽然会因此带来较多样本方面的投入



监控指标异常检测-算法方案

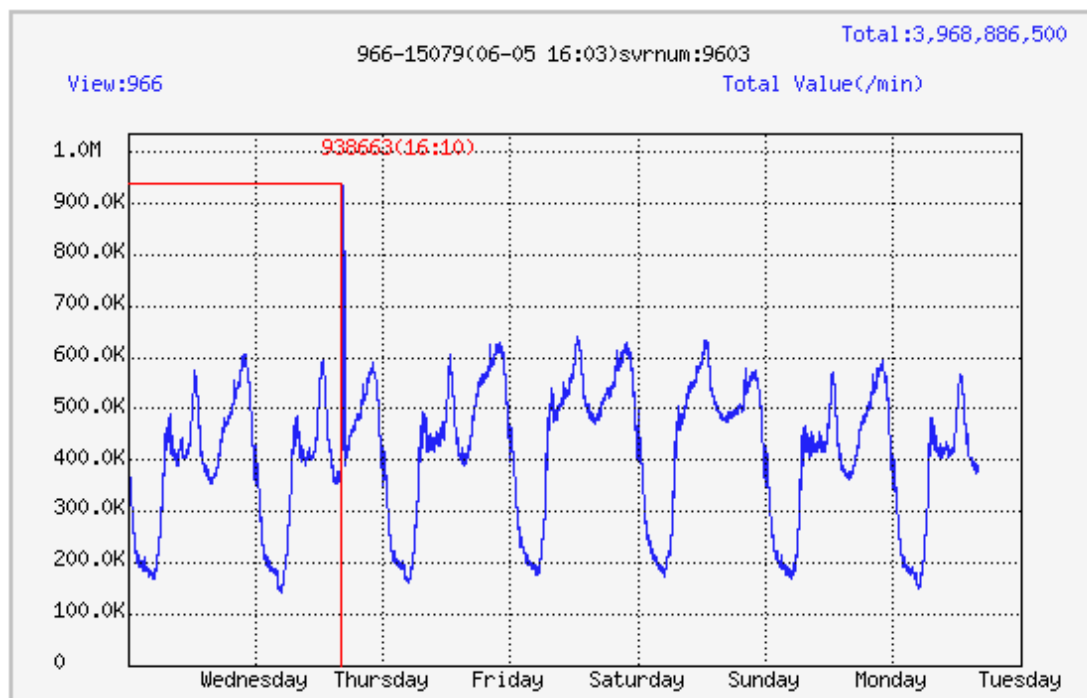




监控指标异常检测-特征工程

统计特征

- 最大值, 最小值, 值域
- 最小值位置、最大值位置
- 均值, 中位数
- 平方和, 重复值
- 方差, 偏度, 峰度
- 同比, 环比, 周期性
- 自相关系数, 变异系数

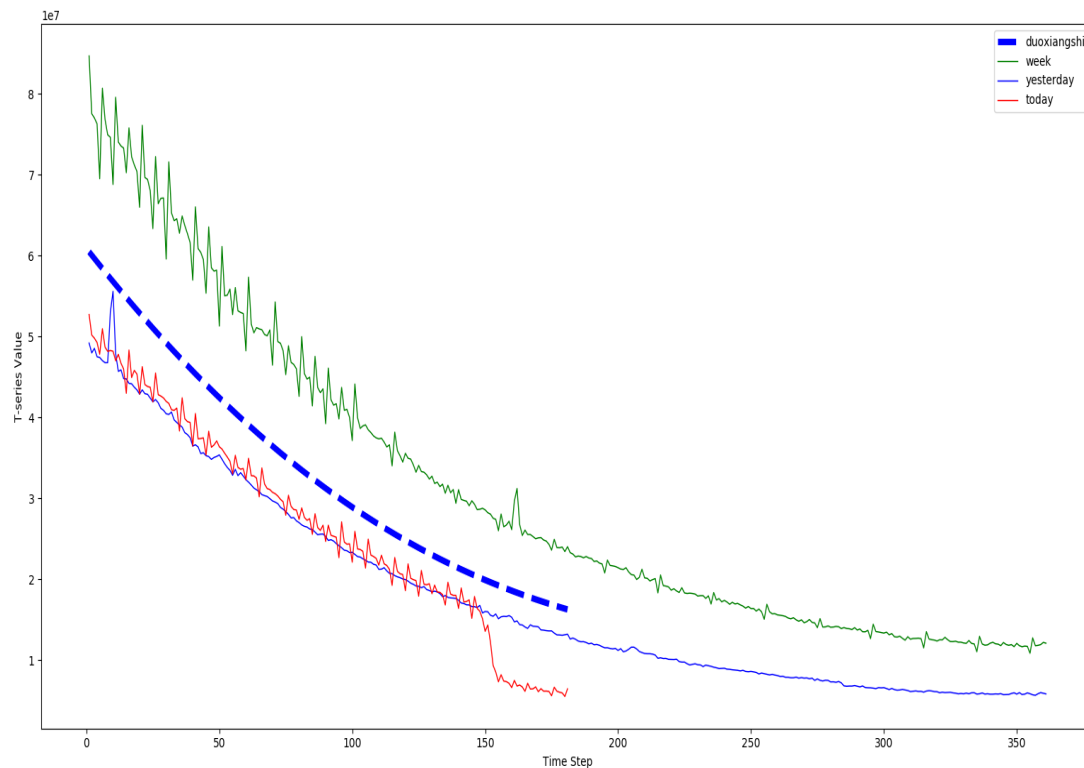




监控指标异常检测-特征工程

拟合特征

- 移动平均算法
- 带权重的移动平均算法
- 指数移动平均算法
- 二次指数移动平均算法
- 三次指数移动平均算法
- 奇异值分解算法
- 自回归算法
- 深度学习算法



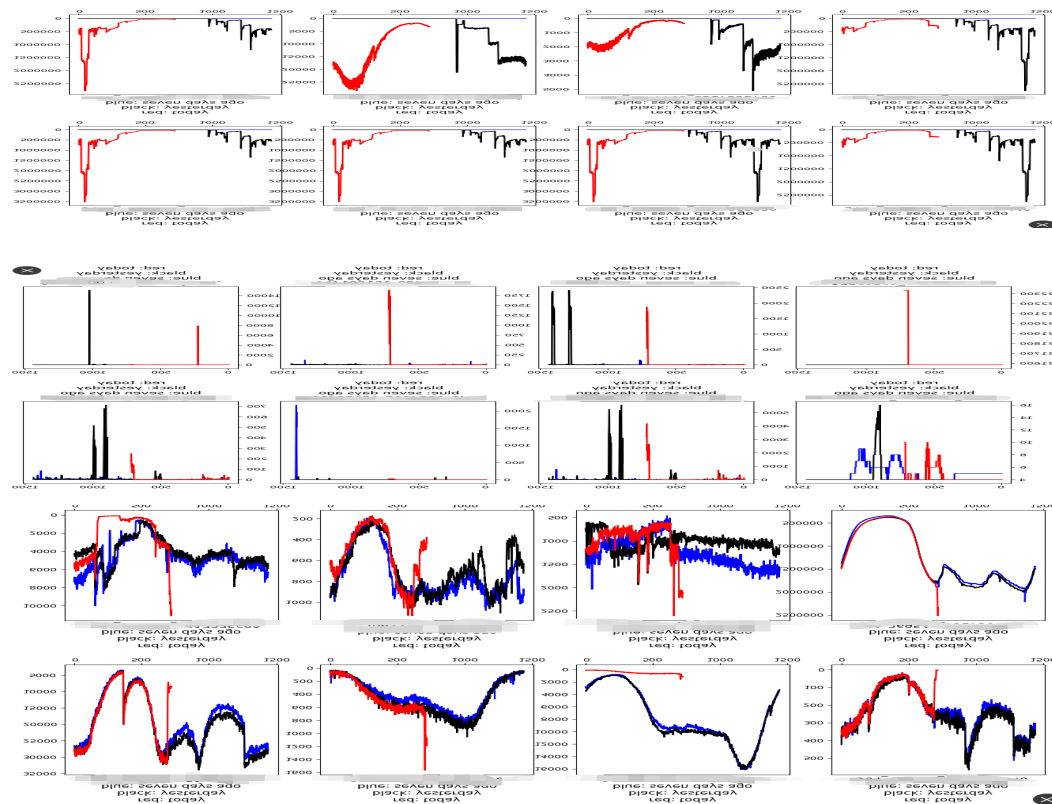
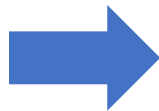


监控指标异常检测-特征工程

分类特征

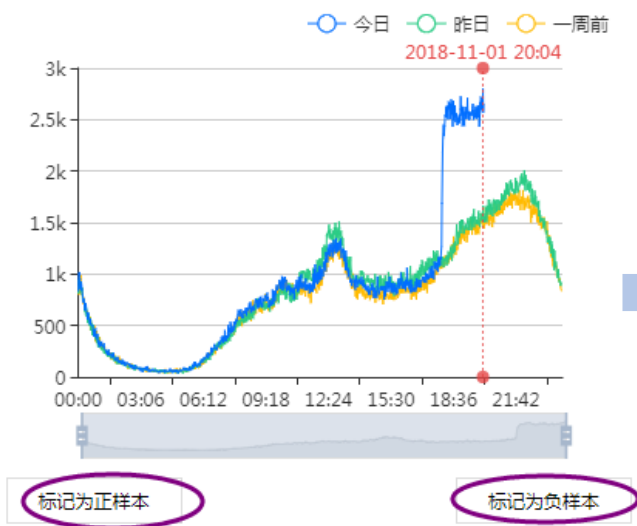
- 熵特征
- 值分布特征
- 小波分析特征

聚类
Kmeans
分类器





监控指标异常检测-打标与样本管理



人工标记

view_id	attr_id	beg_time	end_time	sender	alarm_continue	is_check
7335	2151648	2017-11-07 19:02:03	2017-11-07 19:10:02	0	0	1
1405	142637	2017-11-07 19:59:02	2017-11-07 20:06:02	0	0	1
15669	222132	2017-11-08 00:00:02	2017-11-08 10:39:02	0	0	1
6327	390161	2017-11-08 00:04:02	2017-11-08 00:11:02	0	0	1
6368	390161	2017-11-08 00:04:02	2017-11-08 00:12:03	0	0	1
6327	390161	2017-11-08 00:04:03	2017-11-08 00:12:02	0	0	1
6327	390163	2017-11-08 00:04:02	2017-11-08 00:11:02	0	0	1
6327	390163	2017-11-08 00:04:03	2017-11-08 00:12:02	0	0	1
6241	297212	2017-11-08 09:48:02	2017-11-08 09:53:02	0	0	1
6368	390163	2017-11-08 00:04:02	2017-11-08 00:12:03	0	0	1
16397	2389117	2017-11-08 09:03:03	2017-11-08 09:08:02	0	0	1
4392	20	2017-11-08 10:04:02	2017-11-08 10:09:02	0	0	1
4651	54039	2017-11-08 00:04:03	2017-11-08 00:09:03	0	0	1
11395	129854	2017-11-08 08:53:02	2017-11-08 08:58:02	0	0	1
15013	2389117	2017-11-08 09:04:03	2017-11-08 09:09:02	0	0	1
18018	629583	2017-11-08 09:20:02	2017-11-08 09:25:02	0	0	1
69141	10788	2017-10-25 11:04:03	2017-10-25 11:05:03	1	0	0
69141	10788	2017-10-25 11:04:03	2017-10-25 11:05:03	1	0	0
69141	10788	2017-10-25 11:04:03	2017-10-25 11:05:03	1	0	0
69141	10788	2017-10-25 11:04:03	2017-10-25 11:05:03	1	0	0
69141	10788	2017-10-25 11:04:03	2017-10-25 11:05:03	0	0	0
69141	10788	2017-10-25 11:04:03	2017-10-25 11:05:03	0	0	0

- ◆ 样本的质量，直接决定了模型训练的准确性
- ◆ 样本较为复杂，需要指定规范加以分类管理
 - 正负样本
 - 来源
 - 负样本分类



监控指标异常检测-技术框架

● 离线模块



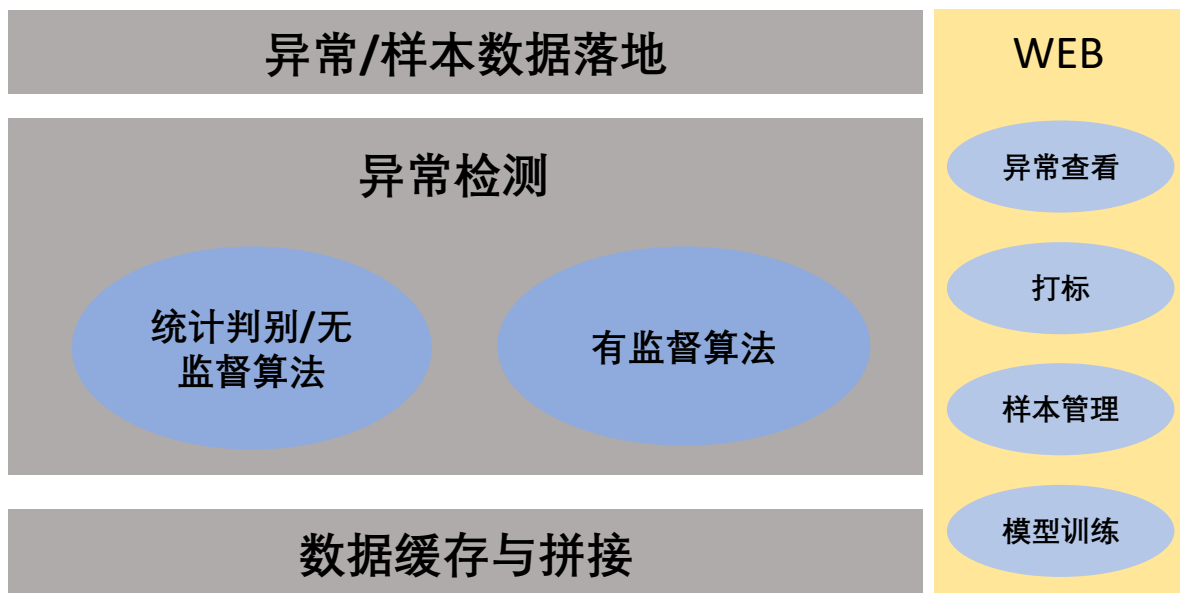
● 在线模块



● ABTest 模块



监控指标异常检测-工程实现



- ◆ 数据缓存的实现避免了从监控系统频繁获取重复数据
- ◆ 数据先经过统计判别算法，过滤大量正样本
- ◆ 结合web，在查看异常的同时方便标注
- ◆ 积累样本，适时重新训练更新模型
- ◆ 问题：如何提高鲁棒性？



其他应用场景实践

内存智能降冷

- 生成数据特征--制作数据画像（抽样扫描业务数据）
- 专家经验打标分类
- 生成分离策略--机器学习预测是否能够下沉
- 根据预测结果自动执行流程

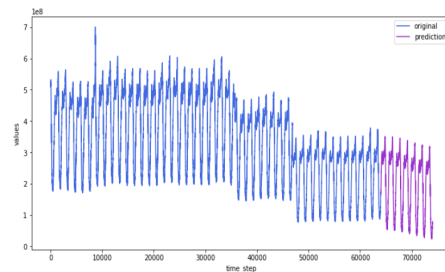
KEY长度	记录大小	最后读时间	最后写时间	过期删除时间	周期热度变化	访问量	业务数据量	空查率
bid	over_3days	expire_in_7_rate	diff_rate	avg_density	used_g	empty_query_rate	class	
0.24	1.00	0.00	82	172	1.00	0		
0.62	1.00	0.29	11	168	0.57	0		
0.00	0.94	0.00	154	198	0.84	0		
0.83	0.94	0.53	27	2	0.87	0		
0.02	0.93	0.05	866	221	0.75	0		

逻辑回归 随机森林

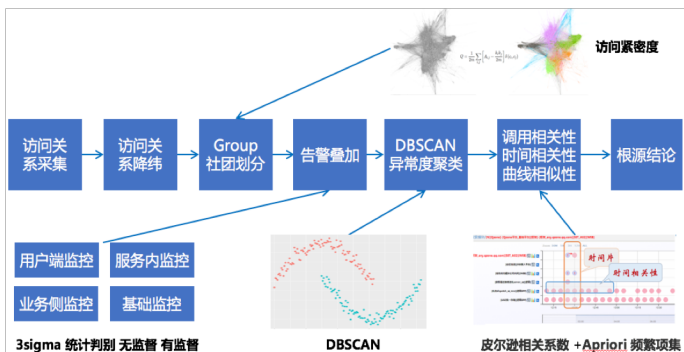
时间序列指标预测

时间序列指标预测的效果评估方式：

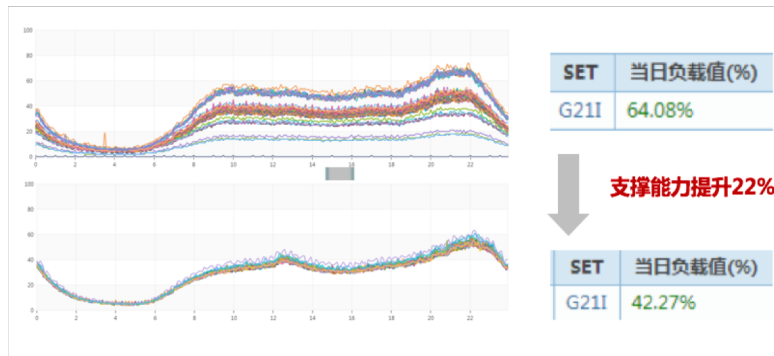
- 针对测试数据，计算均方根误差（RMSD）或者平均绝对误差（MAE）
- 选择误差最小的模型进行预测即可



根源故障分析



平衡树系统



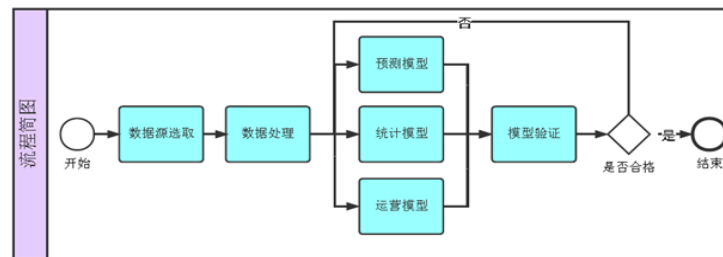


其他应用场景实践

主机异常智能分析



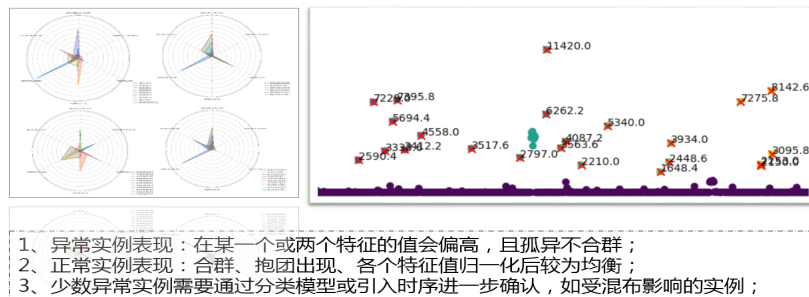
硬盘生命周期预测



智能运维机器人



MySQL异常智能分析





目录

1 AIOps入门

2 腾讯AIOps实践案例



3 学件与Metis



什么是学件

- ◆ 学件 (Learnware) = 模型 (Model) + 规约 (Specification)
- ◆ 概念提出者：南京大学教授，国内机器学习领域领军人物周志华老师。
- ◆ 学件可以认为是带有数据记忆功能的API
- ◆ 学件的特点
 - ◆ 可重用(reusable): 在预训练模型基础上进行更新或者增强即可用于其他任务
 - ◆ 可演进(evolvable): 可以根据环境变化调整
 - ◆ 可了解(comprehensible): 应在一定程度上能被用户了解



AIOps应用的难点



没有数据或者数据量级太少



缺乏算法专家



希望可以自主可控



供需方均担心自身数据泄漏

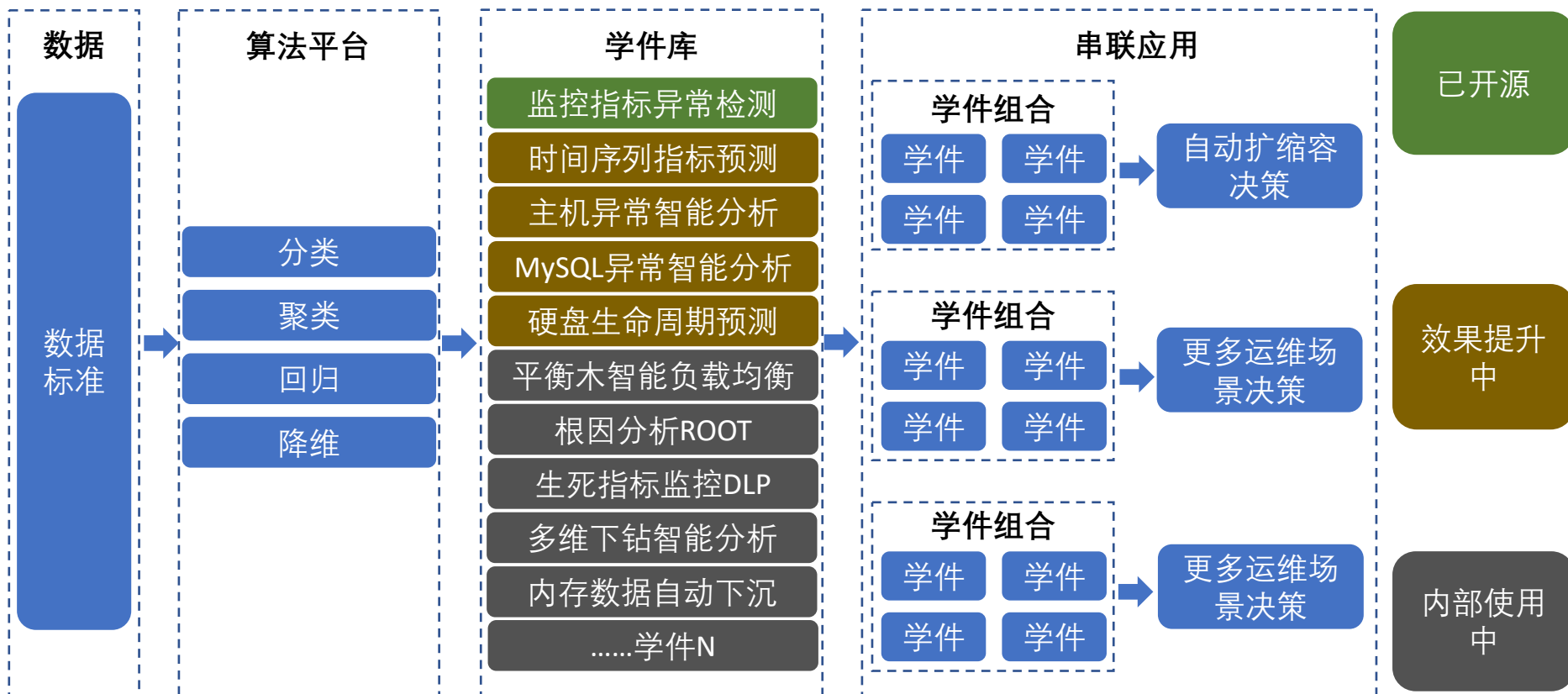


需要大量计算资源

AIOps非常适宜使用学件的方式运作建设



从Metis的诞生到开源





Metis监控指标异常检测学件已经开源

GitHub地址 : <https://github.com/tencent/metis>



异常检测

01

提供率值检测和量值检测API接口，对时间序列进行检测，HTTP协议

特征提取

02

提供三类特征（统计特征、拟合特征、分类特征）的提取功能，特征维度90+；支持增加自定义特征

异常查询

03

经API检测后的时间序列（仅异常）入库存储，提供管理功能，分页查询、检索、放缩等

标注打标

04

提供标注打标的功能，标记/取消标记为正负样本，标记后样本入样本库

样本管理

05

提供样本管理功能，检索、图示、编辑、删除，导入等功能

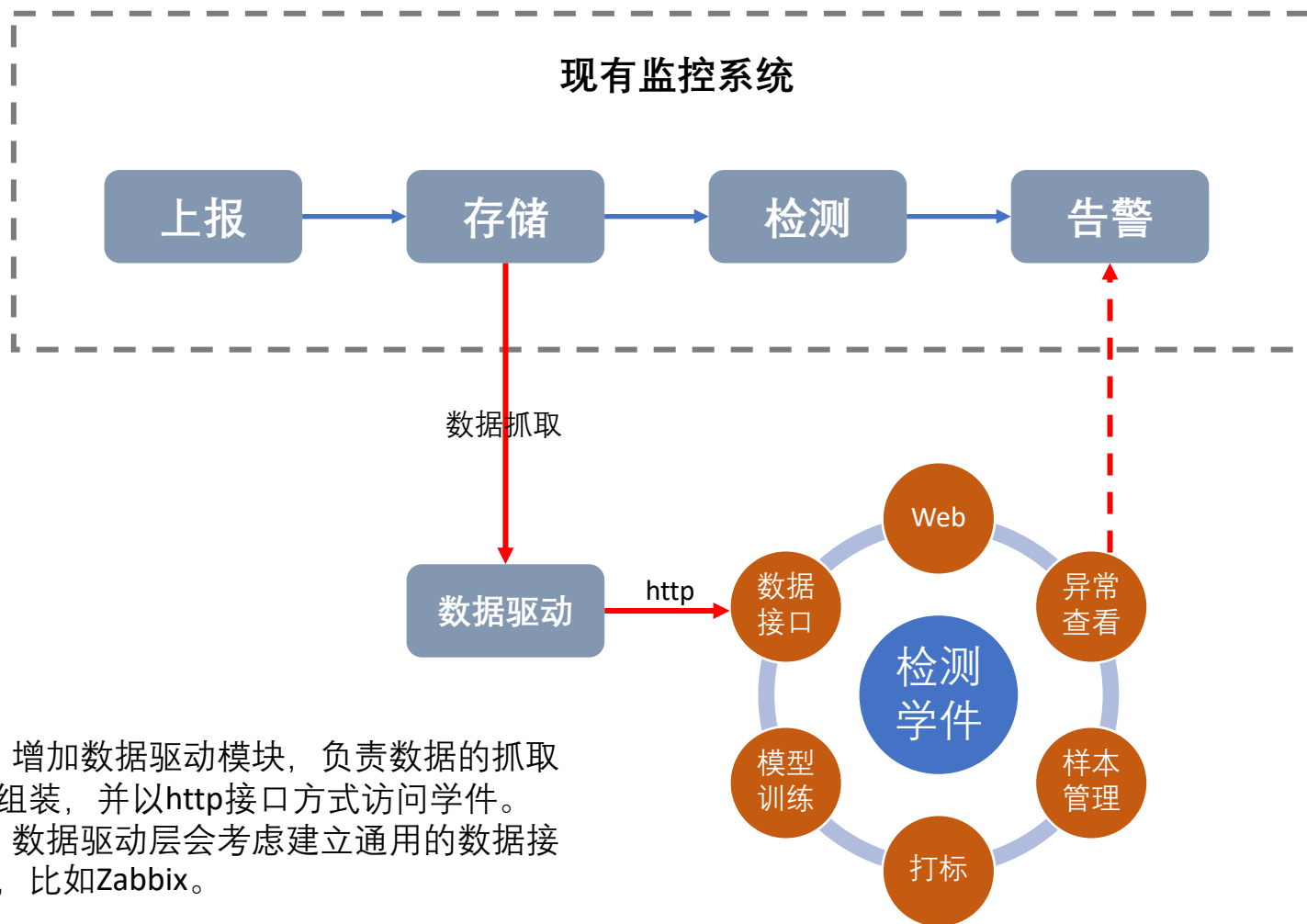
模型管理

06

提供模型管理功能，支持自定义模型训练

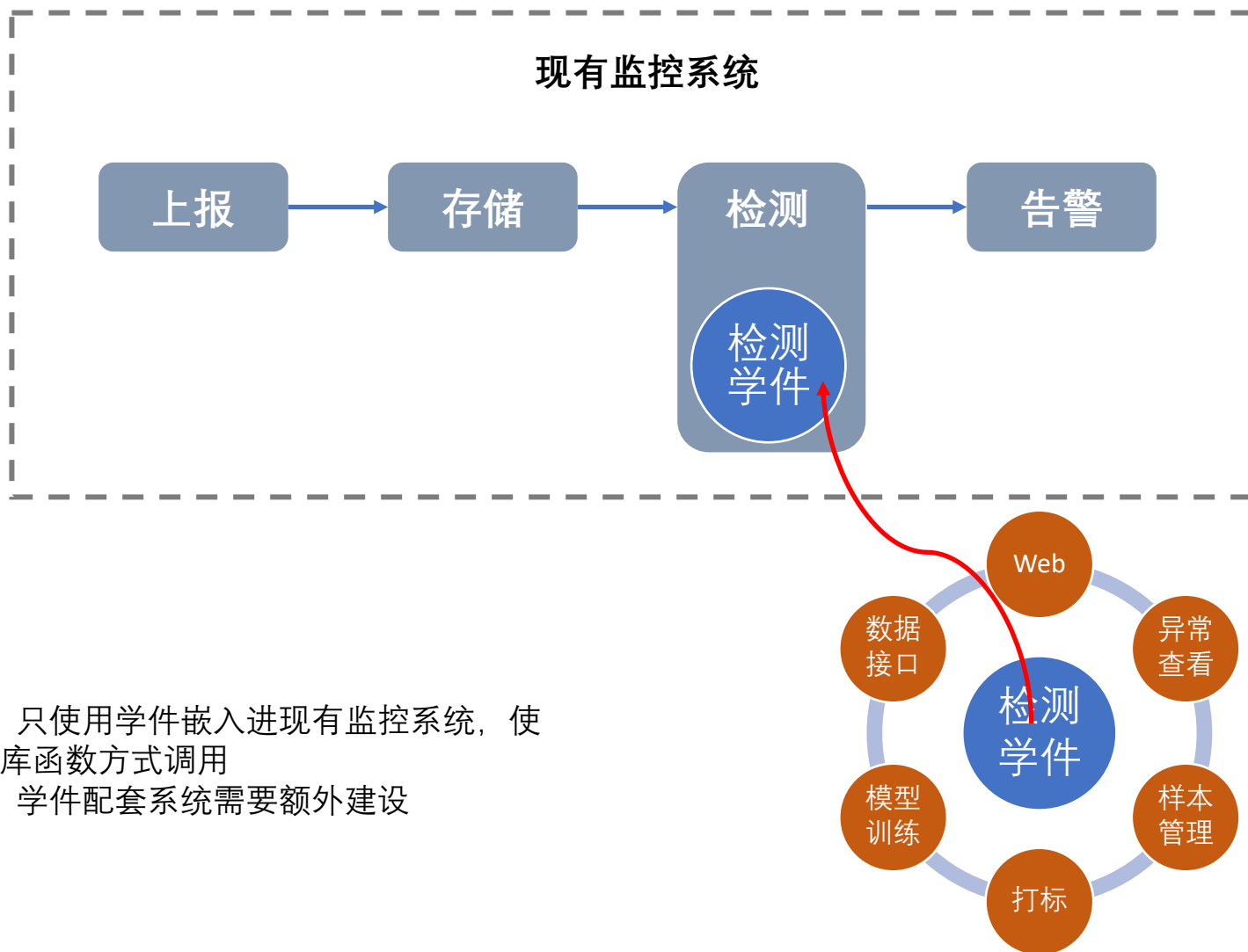


如何使用Metis异常检测学件①





如何使用Metis异常检测学件②





Metis异常检测学件使用场景

没有免费的午餐（NFL）定理

- ◆ 机器学习并非无所不能
- ◆ 没有通用的算法
- ◆ 没有普适的模型

★ ★ ★ 业务指标：收入量、PV量、用户在线量等

★ ★ 系统指标：链接数、流量等

★ 硬件指标：CPU、内存

GIAC

全球互联网架构大会

GLOBAL INTERNET ARCHITECTURE CONFERENCE



关注公众号获得
更多案例实践