

智能运维中的科研问题

清华大学 裴丹
微信: peidanwechat

智能运维落地的核心挑战：

工业界：有数据、有应用，但是欠缺算法经验

学术界：有理论算法，但没数据、不熟悉智能运维场景

工业界-学术界合作：一对一交流效率低、见效慢、不开源开放

智能运维落地的核心挑战：

工业界：有数据、有应用，但是欠缺算法经验

学术界：有理论算法，但没数据、不熟悉智能运维场景

工业界-学术界合作：一对一交流效率低、见效慢、不开源开放

解决思路：科研问题为导向

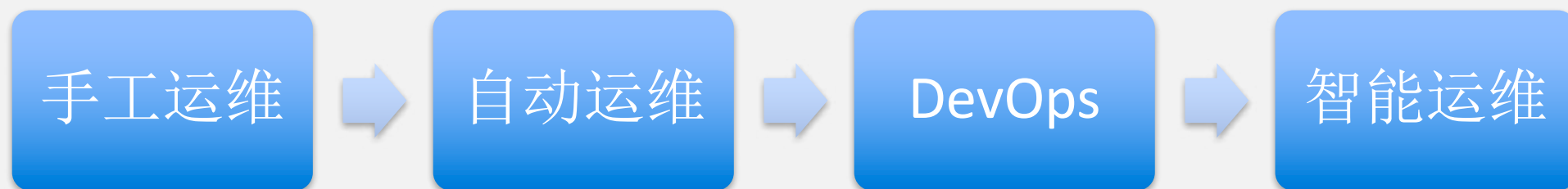
把应用难题分解定义成切实可行的科研问题

企业提供脱敏数据作为benchmark

学术界贡献算法

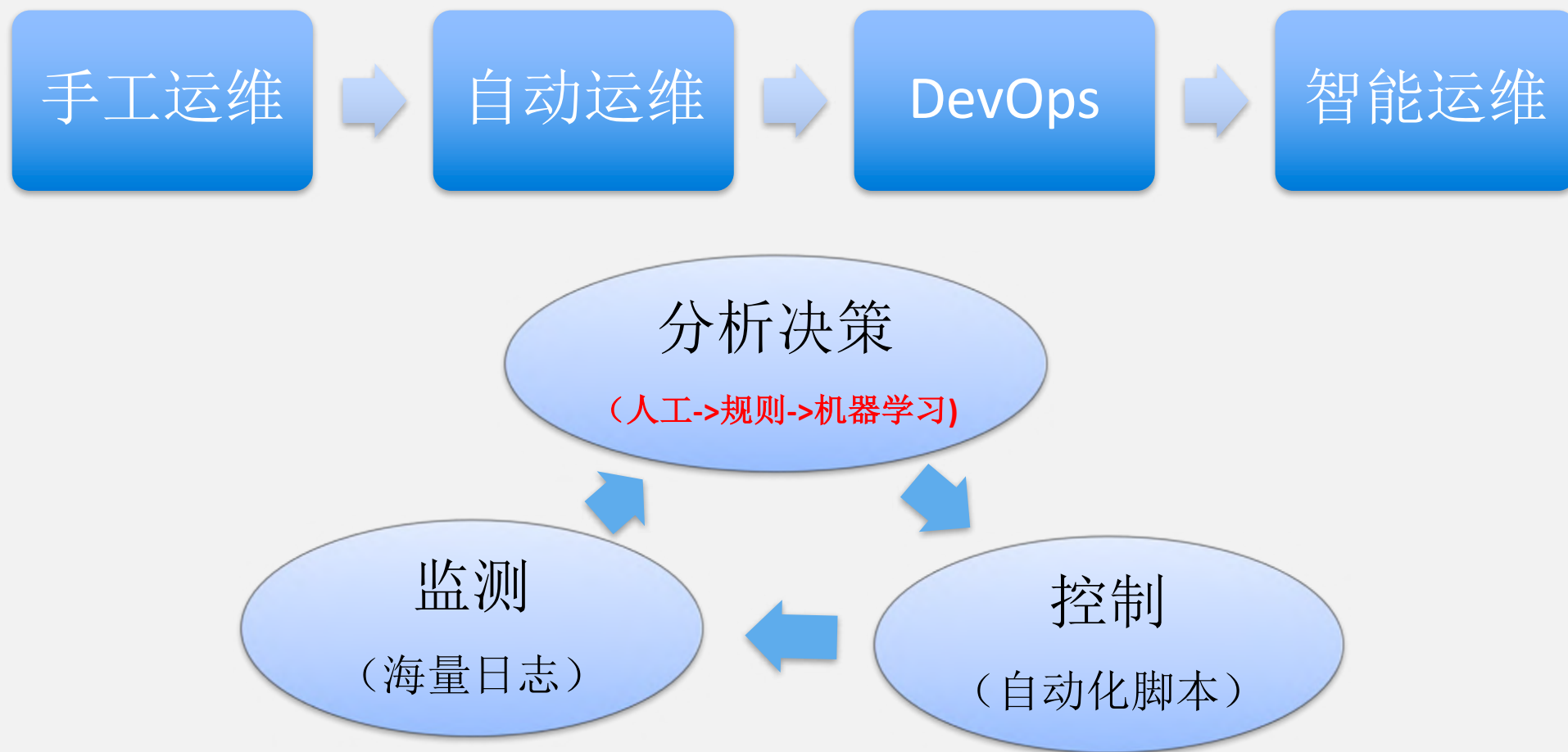
智能运维发展历程

APMCon

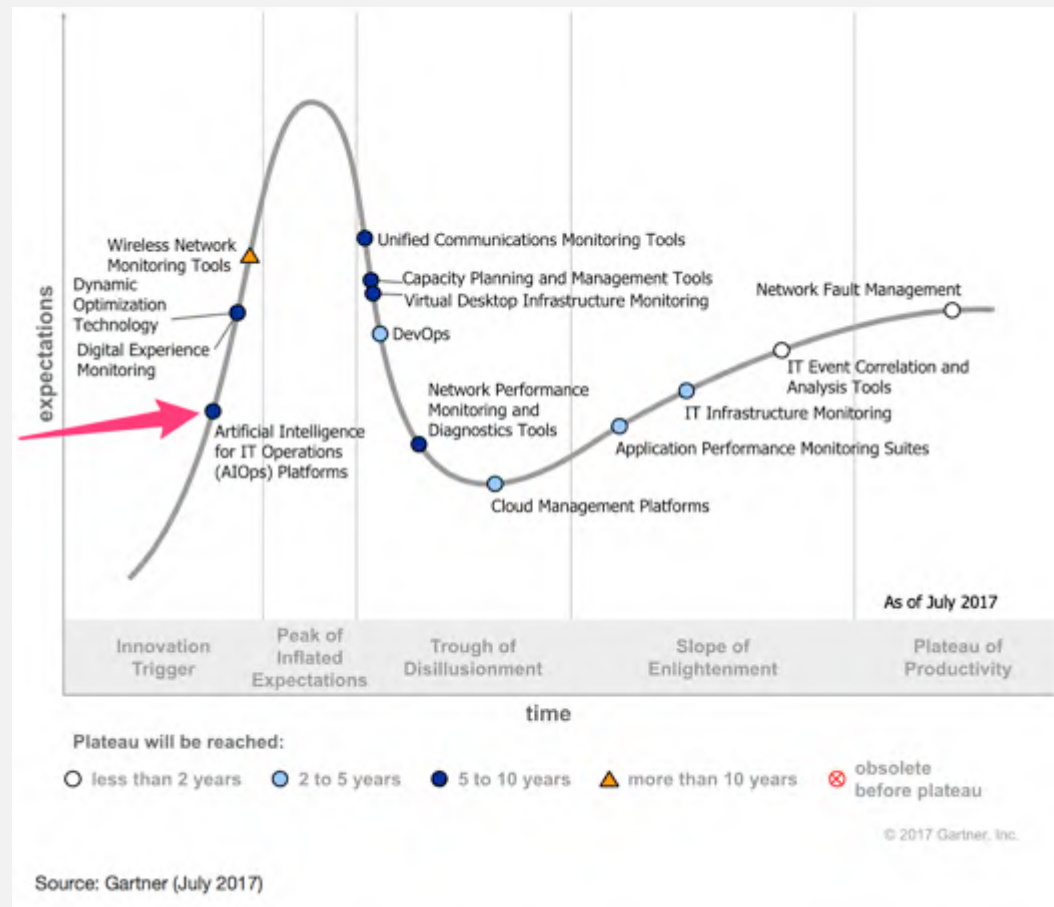


智能运维发展历程

APMCon

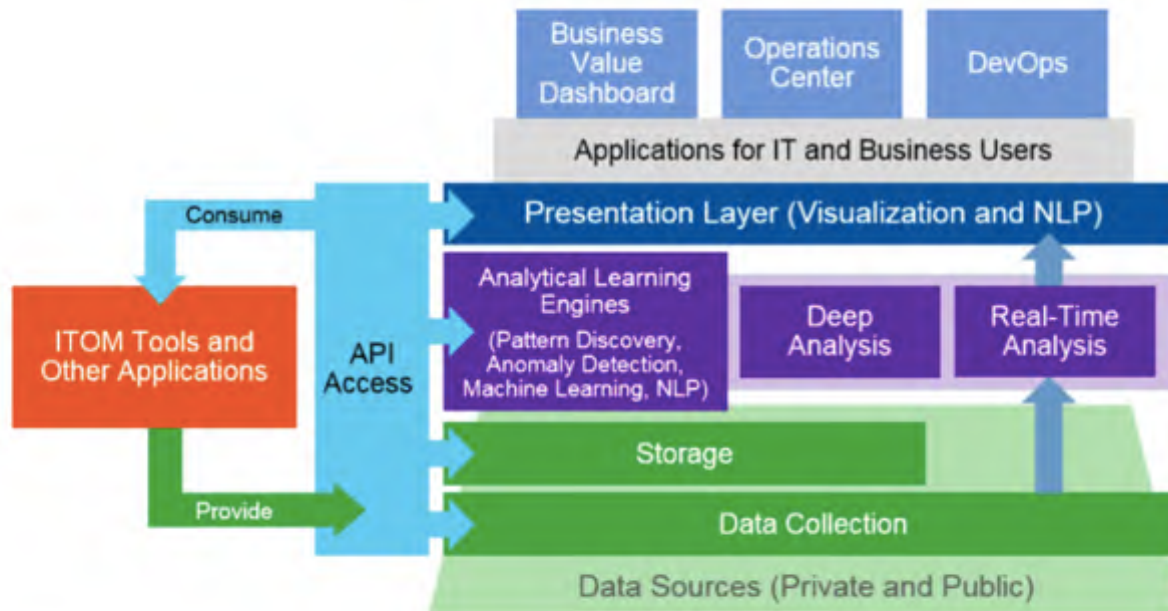


AIOps in Gartner Report



工业界：AIOps

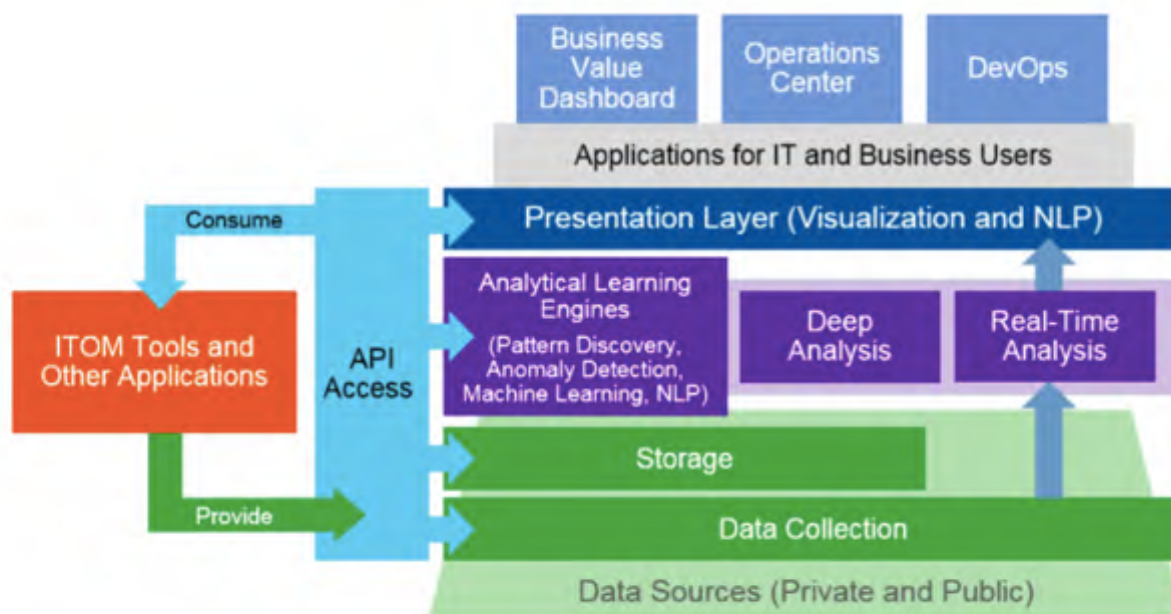
AIOps: Artificial Intelligence for IT Operations



Source: Gartner (March 2016)

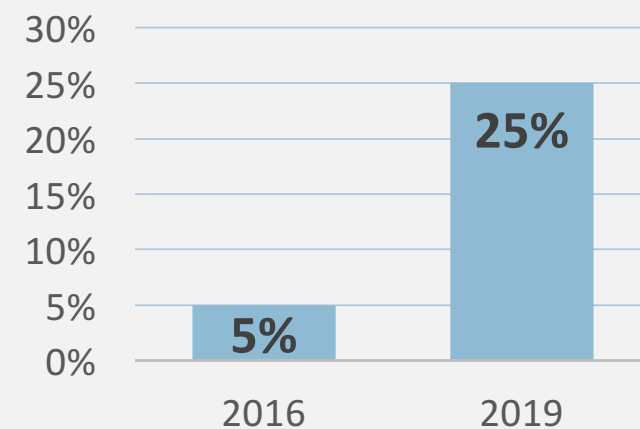
工业界：AIOps

AIOps: Artificial Intelligence for IT Operations Platforms



Source: Gartner (March 2016)

AIOps全球部署率



智能运维前景光明

机器：

- 基础性和重复性的运维工作
- 为复杂问题给出决策建议
- 向运维专家学习解决复杂问题

运维专家：

- 处理运维难题
- 基于机器建议给出决策
- 训练机器徒弟

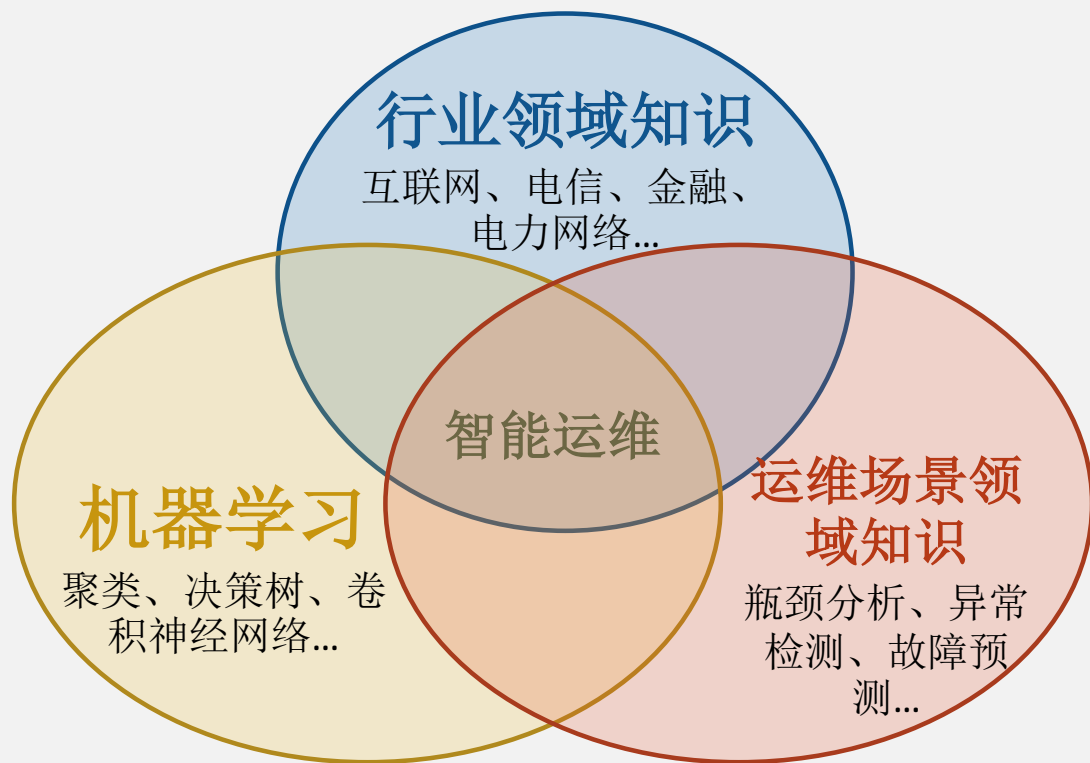
运维工程师：

- 逐渐转型为大数据工程师
- 开发数据采集程序和自动化执行脚本
- 搭建大数据基础架构
- 高效实现基于机器学习的算法

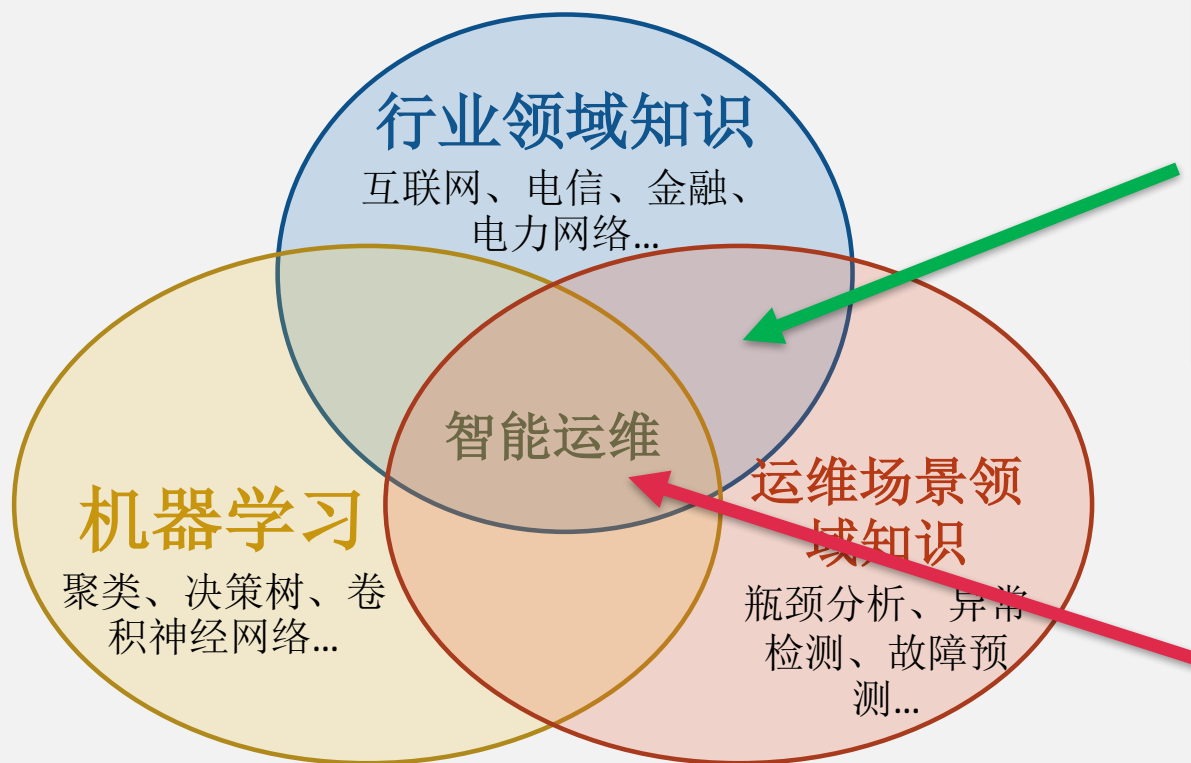
机器学习科学家：

- AI的一个落地应用
- 尚未开采的金矿和低垂的果实

智能运维科研门槛较高 – 工业界



智能运维科研门槛较高 – 工业界



- 熟悉行业和运维场景
- 熟悉生产实践中的难题
- 有数据

- 不熟悉如何把实际问题转化为算法问题
有时一个实践难题需要分解为多个算法问题一个个来解决
- 不熟悉科研参考文献
特别是跨行业的文献

降低工业界门槛的努力：“智能运维前沿” 公众号

APMCon

科普世界范围内智能运维的前沿进展；推动智能运维算法在实践中的落地



《智能运维前沿》课程课件（英文）：

<http://netman.cs.tsinghua.edu.cn/courses/advanced-network-management-spring2017/>

智能运维文献中较为常见的算法:

逻辑回归、关联关系挖掘（事件-事件、事件-时序数据、时序数据-时序数据）、聚类、决策树、随机森林、支持向量机、蒙特卡洛树搜索、隐式马尔科夫、多示例学习、迁移学习、卷积神经网络，递归神经网络（RNN），变分自动编码（VAE）。

发表于如下学术顶会:

ACM SIGCOMM, ACM IMC, ACM/USENIX NSDI, ACM MobiSys, ACM CoNEXT, ACM MobiCom, ACM SIGMETRICS, IEEE INFOCOM, ACM KDD, SIGMOD, VLDB, ICSE

来自Conviva/CMU 的一系列案例

APMCon

通过机器学习，提升视频流媒体用户体验和观看时长

相关分析
信息熵增益
线性回归
SIGCOMM'11



决策树
SIGCOMM'13

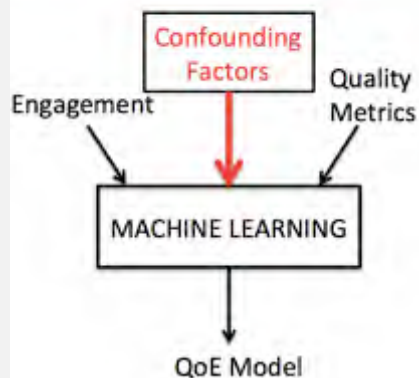


强化学习
NSDI'17

High-level questions & Analysis Techniques

Which metrics matter most? → (Binned) Kendall correlation
Are metrics independent? → Information gain
How do we quantify the impact? → Linear regression

Adding as a feature



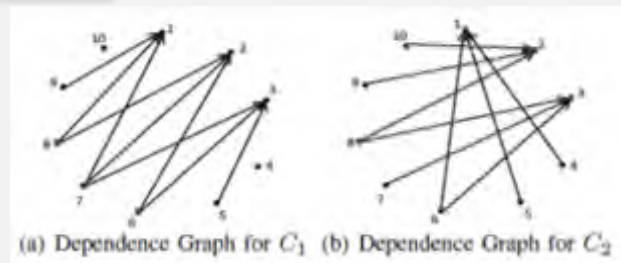
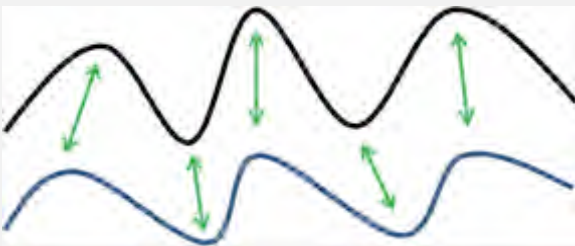
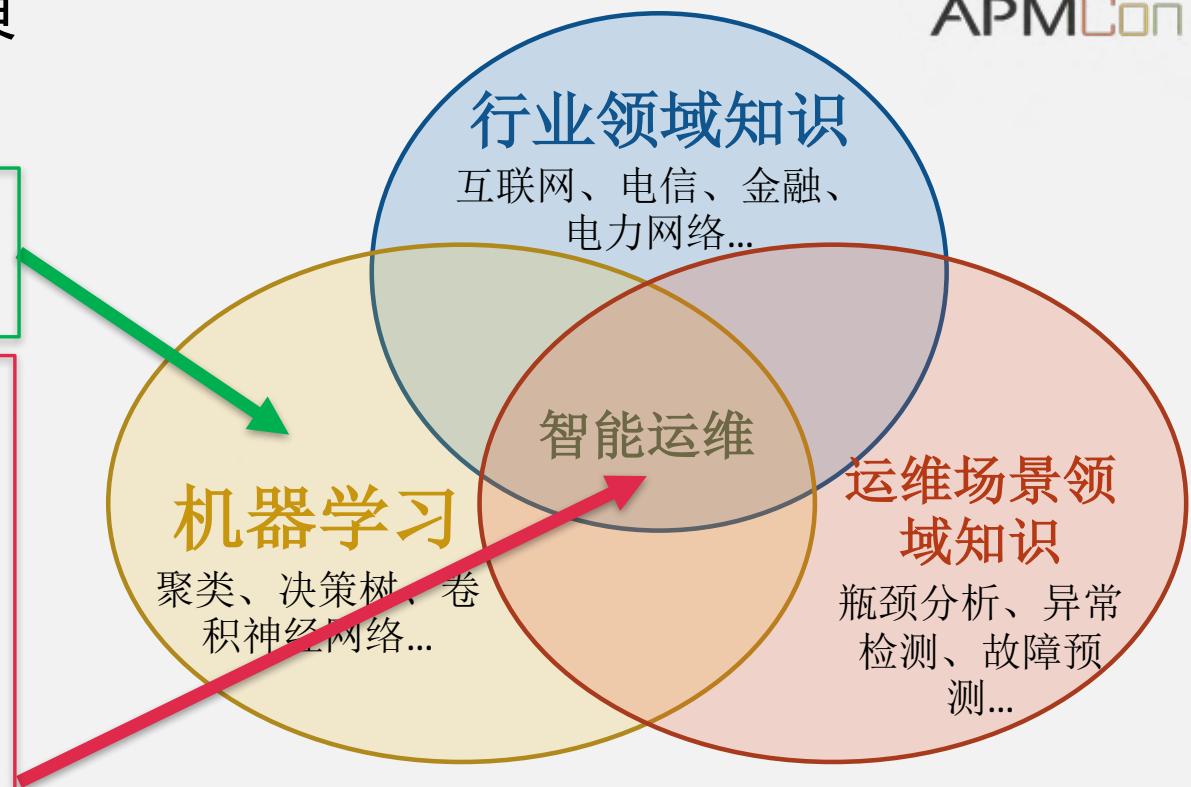
Goal: Optimize mean quality given a limited set of sessions



智能运维科研门槛较高 – 学术界

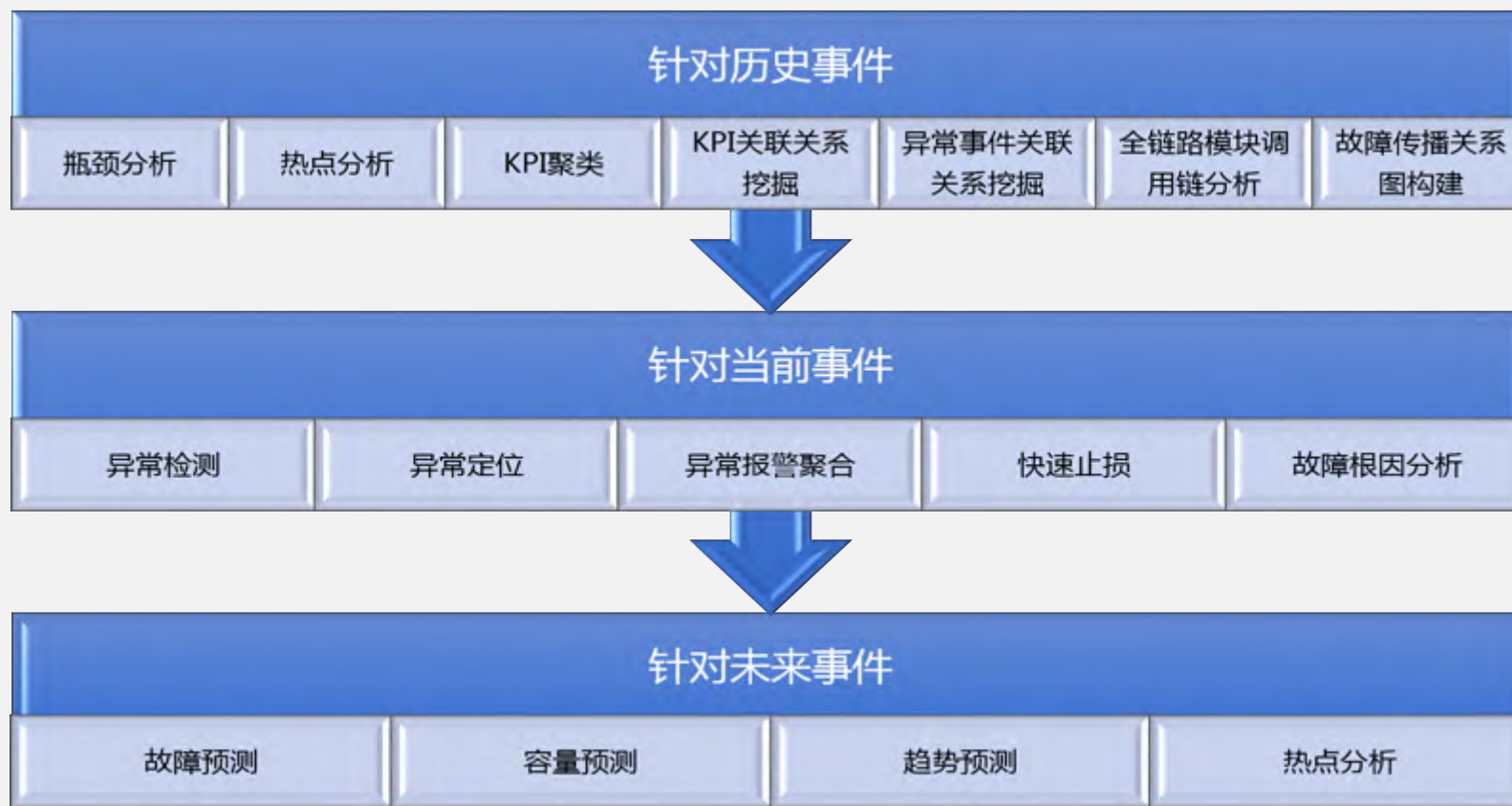
- 算法能力强
- 不熟悉行业和运维的领域知识
- 领域知识门槛高
- 没有数据
- 虽然有相关算法，但是不了解其在智能运维领域的应用

APMCon



降低学术界门槛的努力

应邀在CCF（中国计算机学会）会刊发表专栏文章, 向学术界大同行介绍智能运维科研问题



如何落地：从去年开始号召工业界学术界密切合作

APMCon

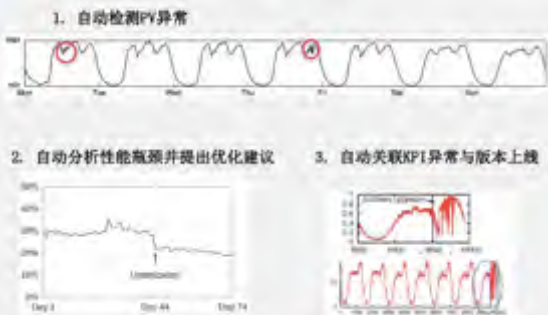
- **工业界与学术界应该在运维领域密切合作**

- 工业界获得算法层面的深度支持
- 学术界获得现实世界的前沿问题及数据，有利发表论文和申请国家项目



Baidu 百度

APMCon



清华副教授从科研角度看运维：
基于机器学习的智能运维

微信公众号文章累计1w+阅读

新的合作

滴滴出行
滴滴一下美好出行

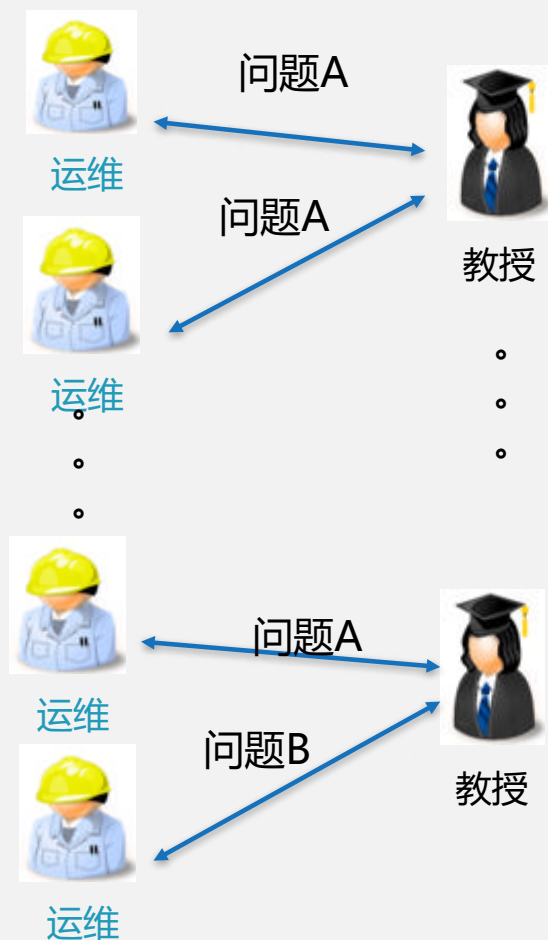
Sogou 搜狗

Alibaba
阿里巴巴

Tencent 腾讯



工业界-学术界合作 1.0： 一对一交流合作



- 交流合作效率低、见效慢
- 智能运维算法不幸成了特权：
 - 仅限于少数大公司与部分合作紧密教授之间
 - 国外：Google, Microsoft, LinkedIn, Facebook, Yahoo!
- 涉及知识产权，不符合开源大趋势
 - 数据不公开
 - 代码不公开

■ 工业界-学术界合作 2.0：开源开放

工业界学术界合作开源开放大趋势：

代码： Hadoop EcoSystem（工业界）

TensorFlow（工业界）

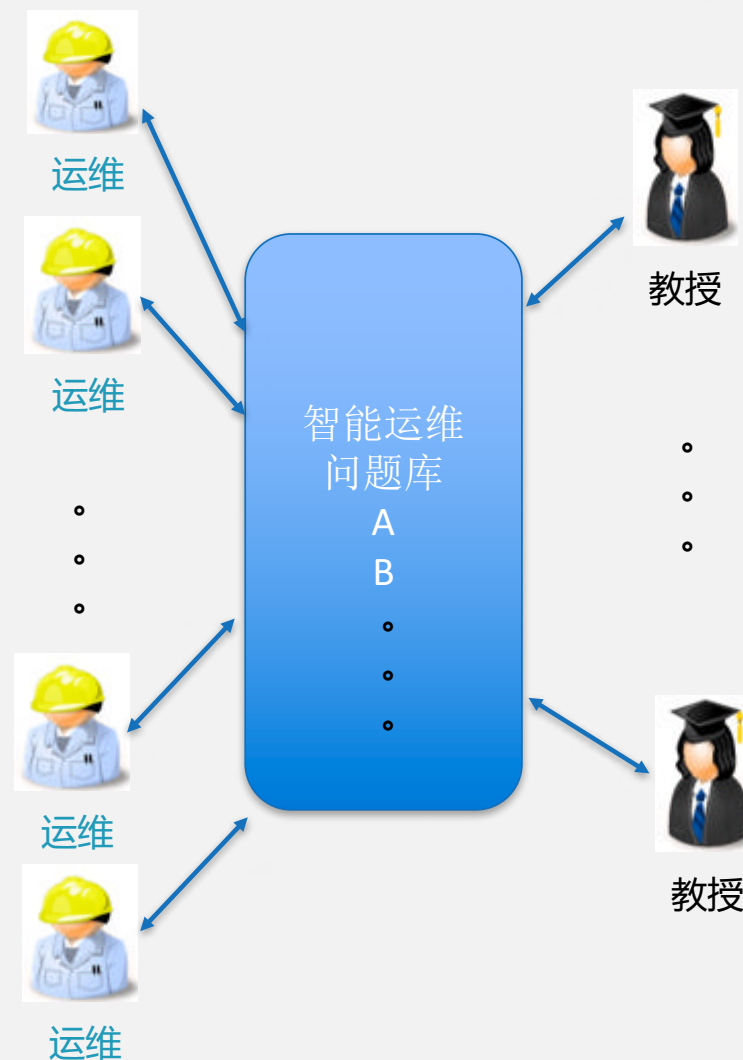
Spark（学术界）

算法： arXiv.org

数据： ImageNet

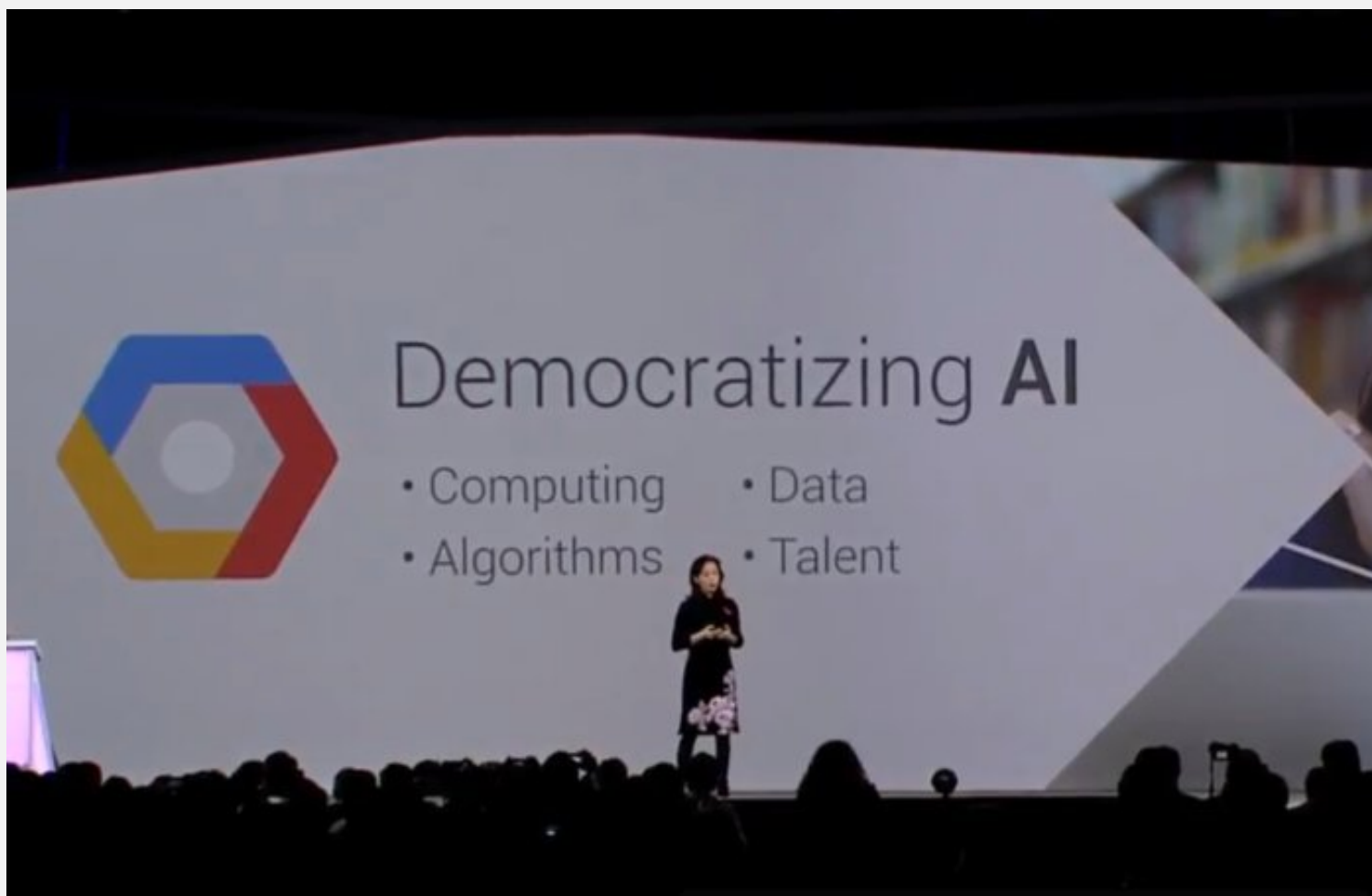
算力： 各大公司的AI云

人才： 美国学术界批量向工业界流动



■ 受 “普世化AI” 启发

APMCon



李飞飞

斯坦福大学副教授、人工智能实验室与视觉实验室主任

ImageNet 创始人

谷歌机器学习部门负责人

普世化智能运维算法



目标：让所有公司都能用上最好的智能运维算法

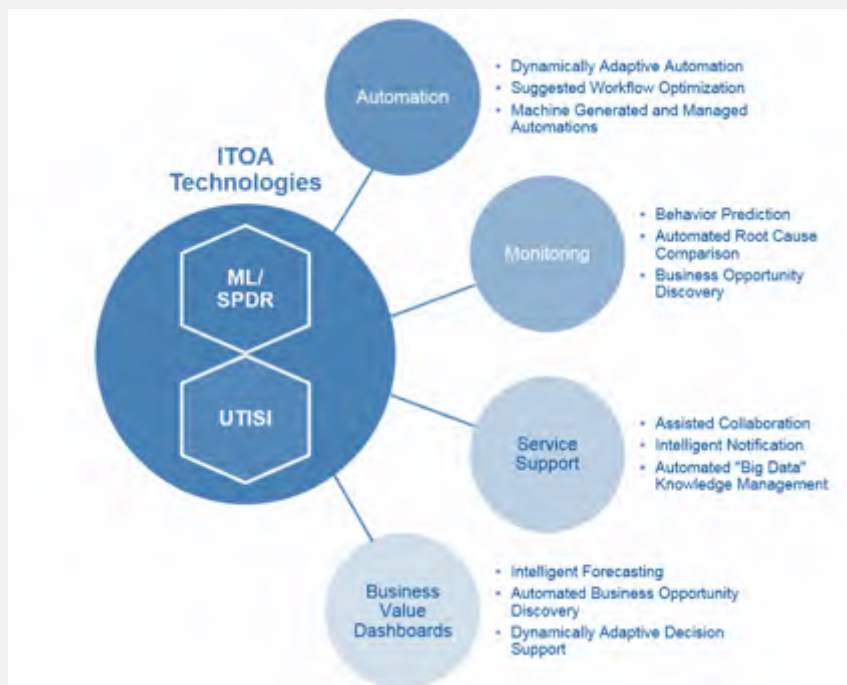
解决智能运维普世化的如下问题：

- 数据
- 算法
- 算力
- 人才

分解定义智能运维中的科研问题

■ 分解定义科研问题

Gartner报告中的问题描述太宽泛



科研问题要求：

- 清晰输入；数据可获得
- 清晰输出；输出目标切实可行
- 有high-level 的技术路线图
- 有参考文献
- 非智能运维领域的学术界能理解能解决

已经定义出的科研问题

(即将公开发布在一个网站上)



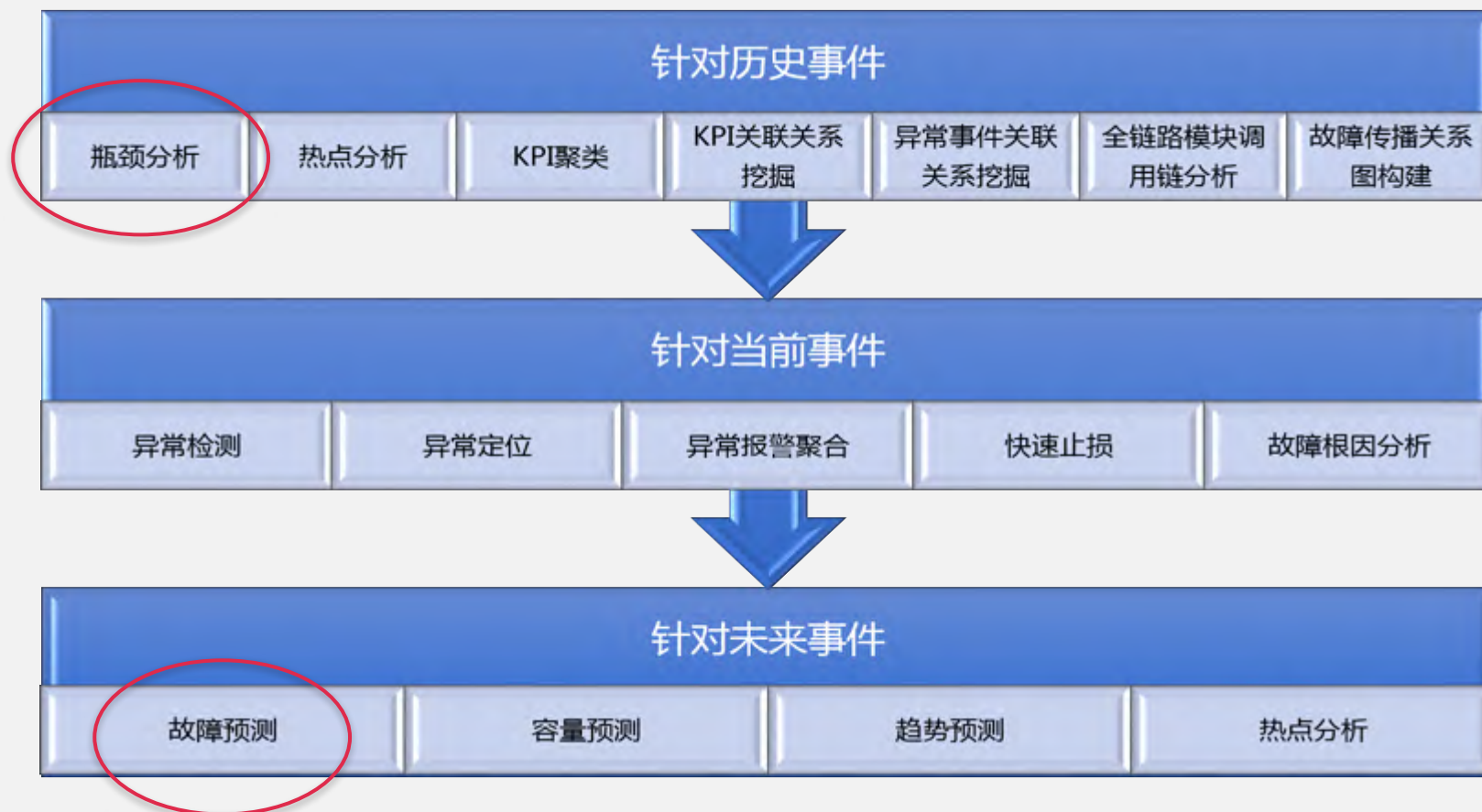
落地智能运维科研算法



- 相对独立算法 -> 直接可落地
- 依赖其它算法 -> “庖丁解牛”
- 数据等条件不成熟 -> “退而求其次”

■ 科研问题定义之 “基础模块”

(即将公开发布在一个网站上)



KPI瓶颈分析算法

面向问题

- 从多维属性数据中挖掘引发KPI瓶颈的条件

输入

- KPI数据及瓶颈阈值
- 可能影响KPI的属性测量数据

输出

- 导致KPI瓶颈的属性（组合）

销售额、利润、订单数、PV、
转化率、用户数、用户增速、
留存率、首屏时间、
闪退率、投诉率
.....

时间戳	关键指标	属性1	属性2	...	属性n
-----	------	-----	-----	-----	-----

运营商、省份、城市、移动设备类型、软件版本号、
移动端模块、浏览器版本、无线网络参数、服务器
端模块、后台负载、用户年龄、用户性别、...

KPI瓶颈分析算法

典型应用场景

- Web 应用首屏时间
- 移动应用加载时间
- 软件报错数
- 视频传输质量

常见算法

- 决策树
- 聚类树 (CLTree)
- 层次聚类

应用挑战

- 瓶颈可能为多种属性和数值的组合
- 不同属性之间可能存在依赖关系
- 避免重叠表示
- KPI可为单、双、多类别

故障预测算法

面向问题

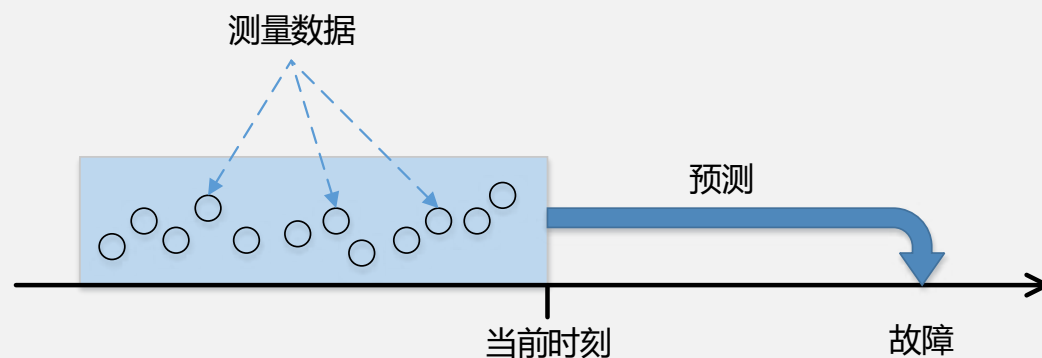
- 在互联网服务运行时，使用多种模型或方法分析服务当前的状态，并基于历史的经验判断在近期是否会发生故障

输入

- 当前服务的运行状态（Syslog日志、SNMP数据）
- 历史故障案例

输出

- 近期是否会发生故障/发生故障概率



典型应用场景

- 硬盘故障预测
- 服务器故障预测
- 交换机故障预测

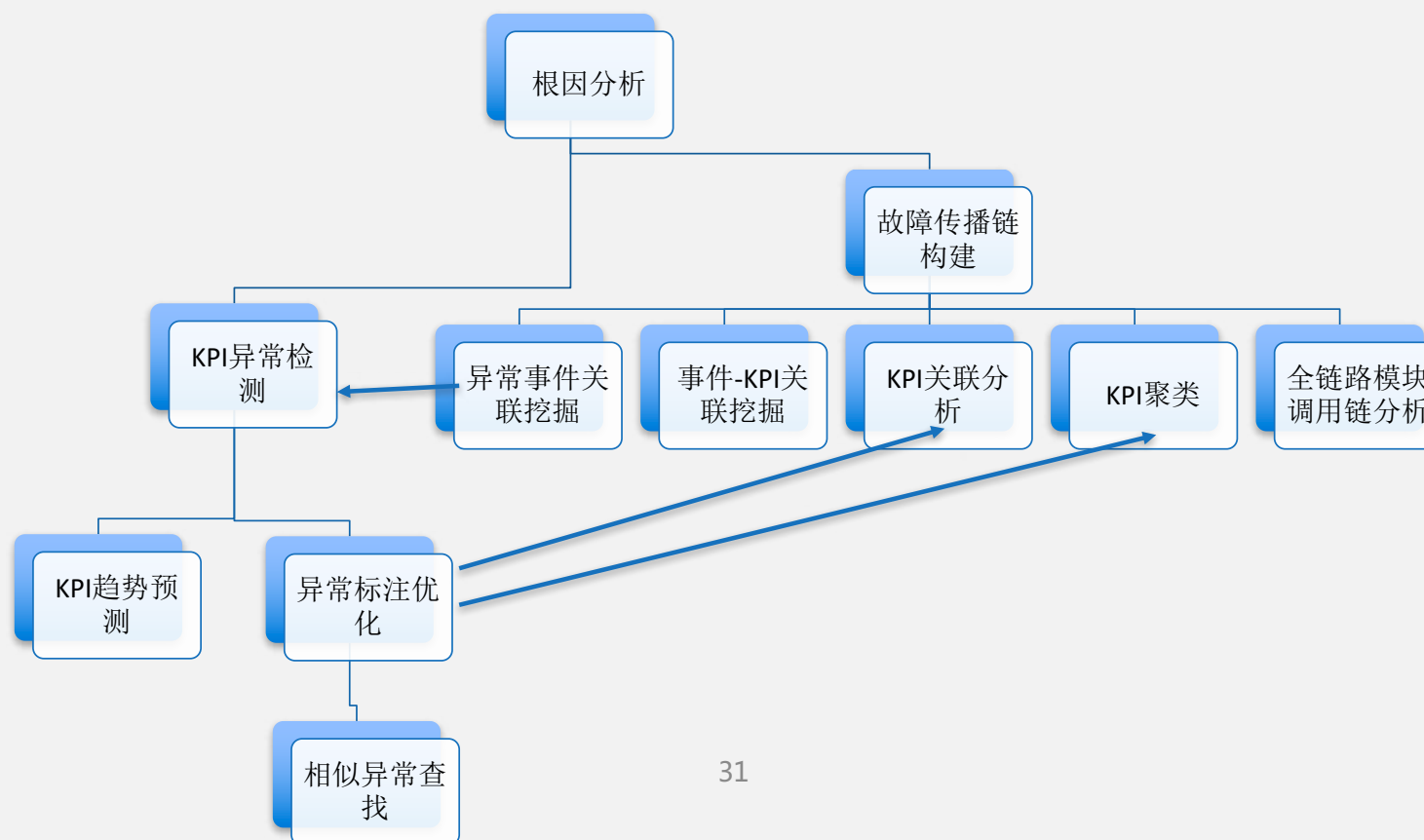
常见算法

- HSMM
- 随机森林
- SVM

应用挑战

- 故障案例少
- 日志量大
- 有益信息少

科研问题定义之 “庖丁解牛”



故障根因分析算法

面向问题

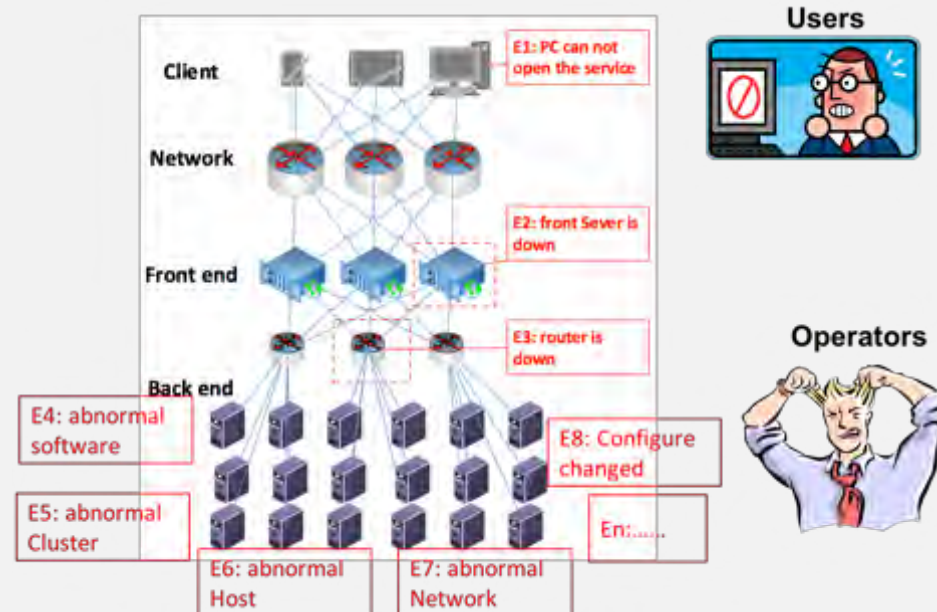
- 当前应用服务发生异常时，分析导致服务异常的根本触发原因

输入

- 服务相关的指标异常状况（包括客户端，网络，服务端等）
- 故障传播关系图

输出

- 根因（排序列表）



故障根因分析算法

典型应用场景

- 应用服务发生异常时，快速诊断根因，快速止损。

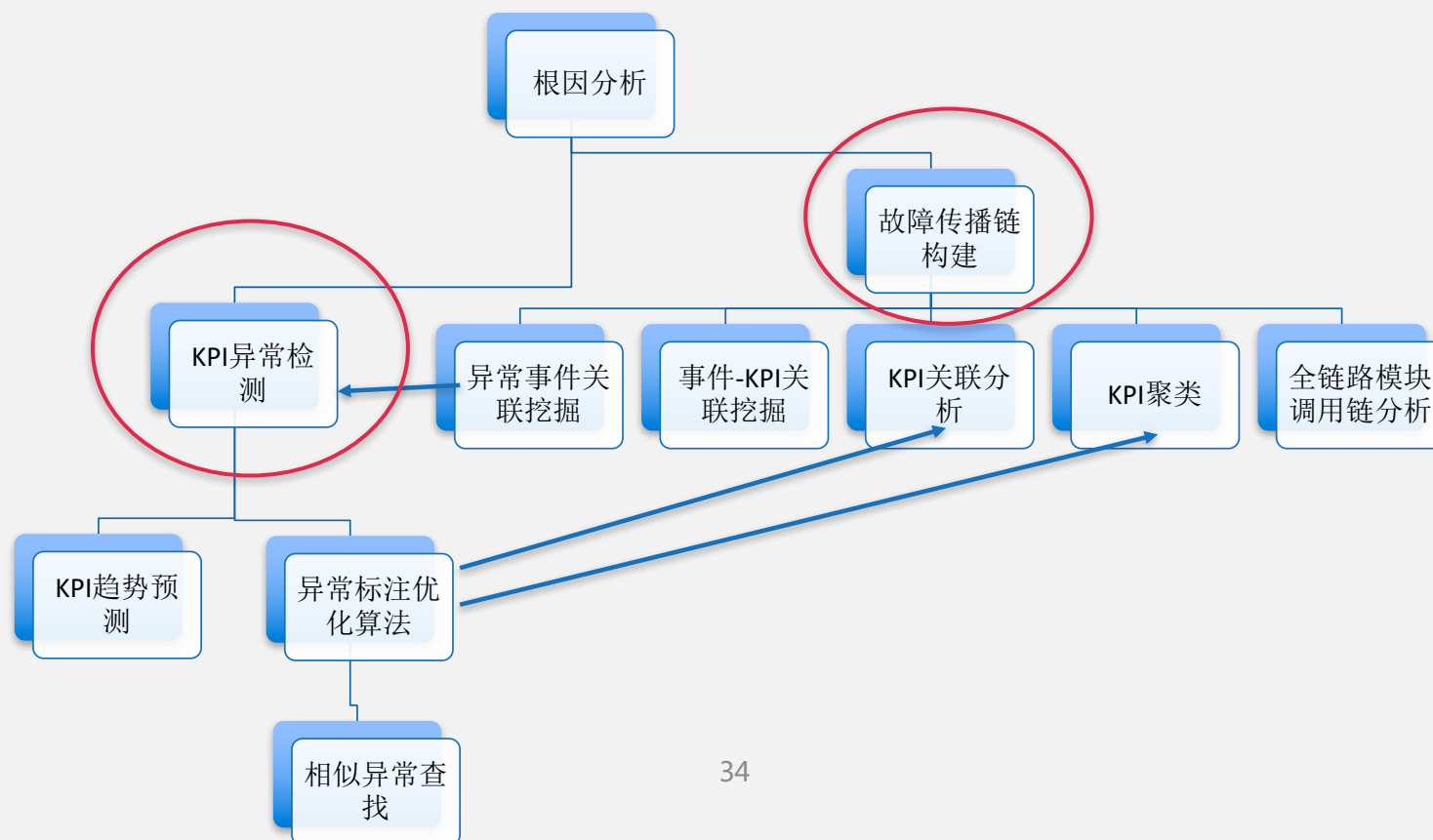
常见算法

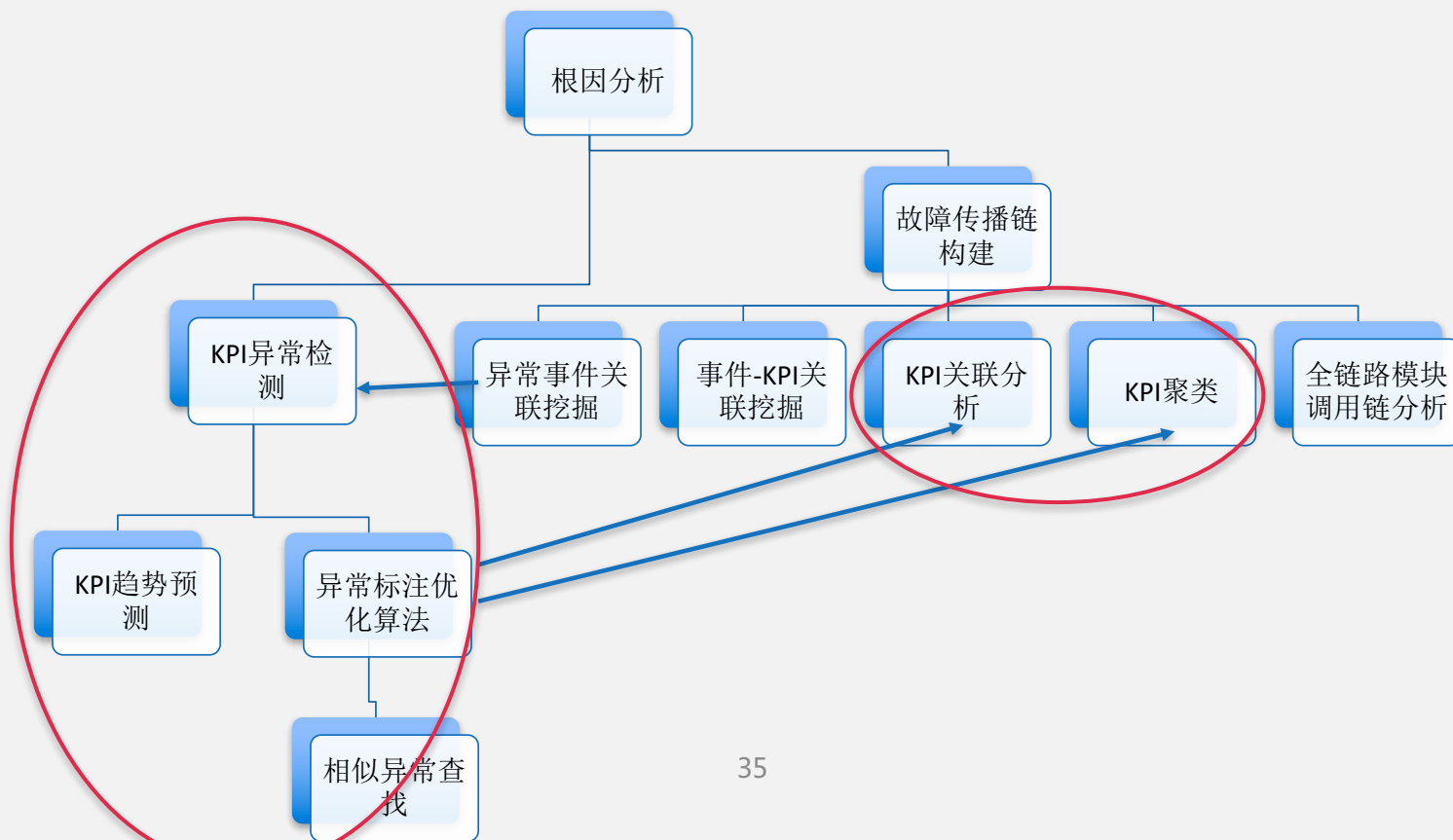
- 基于故障传播链
- 概率图模型

应用挑战

- 数据收集不全
- 故障案例少
- 依赖故障相关的先验知识
- 异常检测存在漏报误报

科研问题分解之 “庖丁解牛”





KPI异常检测算法

面向问题

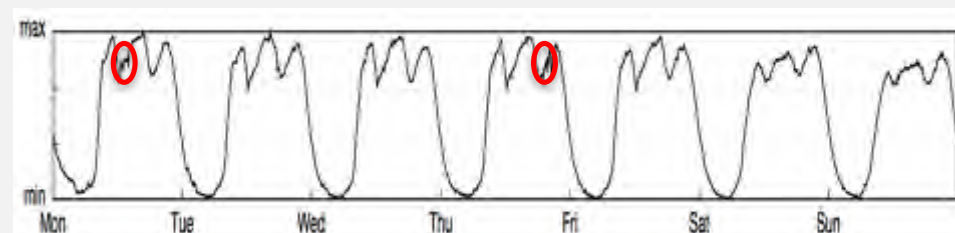
- 检测KPI的异常行为

输入

- KPI时序测量数据
- 异常区间标注

输出

- KPI是否发生了异常



KPI异常检测算法

典型应用场景

- 网络故障
- 服务器故障
- 配置错误
- 缺陷版本上线
- 网络过载

常见算法

- 基于窗口
- 基于近似性
- 基于预测
- 基于隐式马尔科夫模型
- 基于机器学习
- 基于集成学习
- 基于迁移学习
- 基于深度生成模型
- ...

应用挑战

- KPI种类各异
- KPI异常行为难以定义
- 调整算法、参数费时费力
- 需要人工标注
- 人工标注不准确

KPI趋势预测算法

面向问题

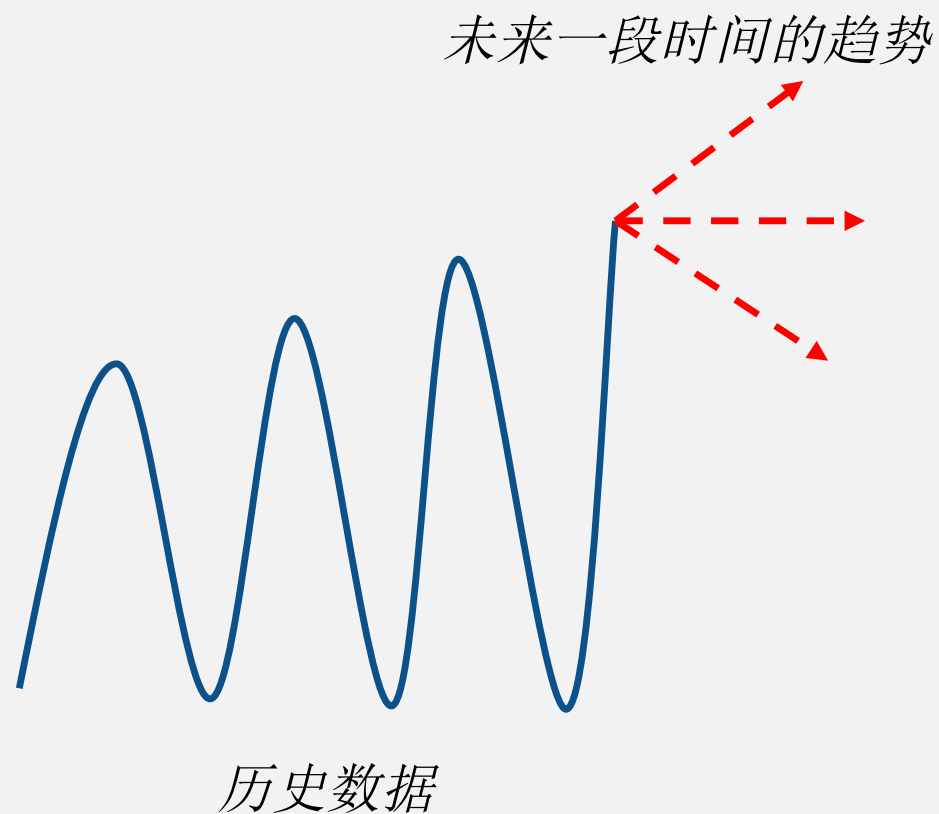
- 通过分析历史数据，判断未来一段时间KPI的趋势

输入

- KPI的历史数据

输出

- 未来一段时间KPI预测值



典型应用场景

- 机器资源需求预测
- 订单量预测
- 作为异常检测、异常定位、容量预测等科研问题的输入

常见算法

- ARIMA
- EWMA
- Holt-Winters
- 时序数据分解
- RNN

应用挑战

- 突发事件的影响
- 节假日，天气等因素的影响
- 数据存在不规则的变动



KPI相似异常查找

APMCon

面向问题

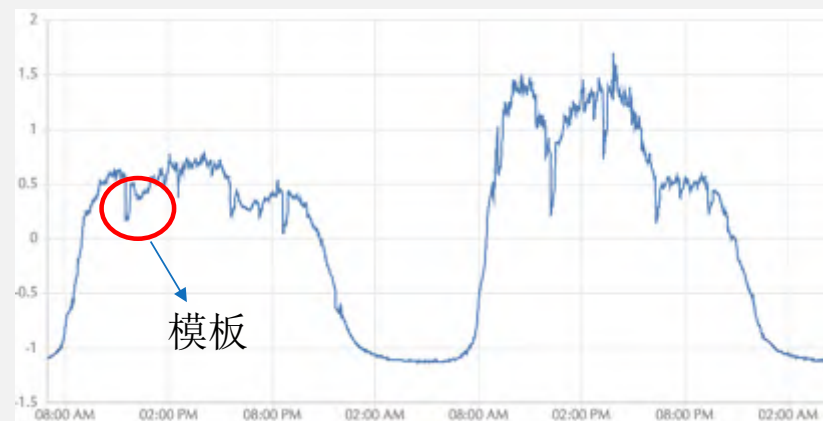
- 面对一根较长KPI曲线的标注，根据已经标出的片段作为模板，找到该KPI曲线上其它的相似异常，减少重复标注的工作量。

输入

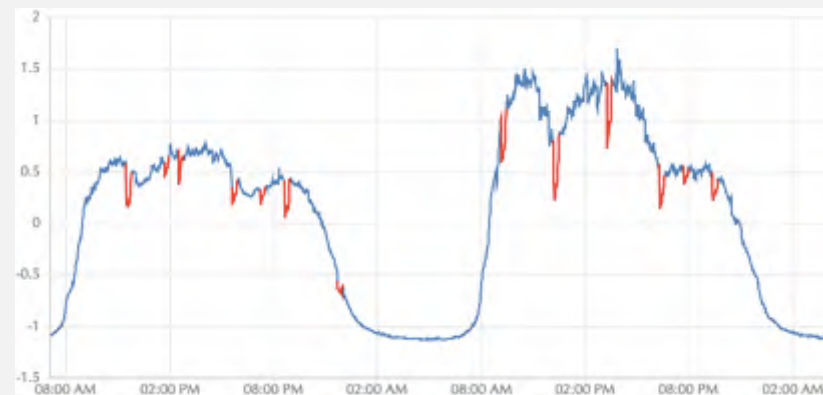
- 一根待标注的KPI曲线和一段已经标注出的异常片段（模板）

输出

- KPI曲线上与模板相似的异常片段



输入



输出

典型应用场景

- 减少异常标注量
- KPI时间序列信息挖掘

常见算法

- Matrix Profile
similarity: DTW,
Euclidean distance
- Mueen-Keogh (MK)
Best-matching Pair

应用挑战

- 实时性要求高
- 异常定义复杂

KPI聚类算法

面向问题

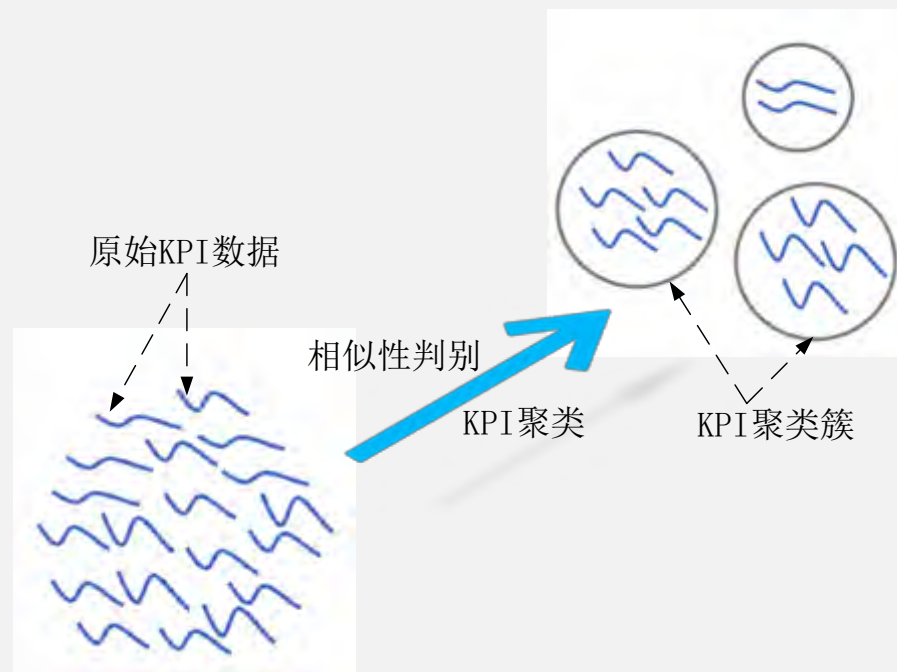
- 面对大规模KPI时序数据曲线，选取合适的度量刻画曲线间的相似性，采用聚类与分派算法快速确定曲线类别。

输入

- 大量KPI时序数据曲线

输出

- 每条曲线所属的类别



典型应用场景

- KPI异常检测中的迁移学习
- 相关异常查找，以减少标注开销

常见算法

- DBSCAN
- K-medoids
- CLARANS

应用挑战

- 数据量大
- 曲线模式复杂
- 对类别的定义不同
- 缺乏ground truth

KPI关联关系挖掘算法

面向问题

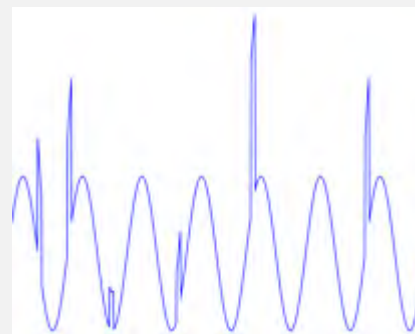
- 互联网公司存在大量的各式各样的时序KPI数据。KPI波动的相关性对于根因分析、故障定位等可以提供很好的线索

输入

- 两条时序KPI数据

输出

- 两条曲线波动是否相关



典型应用场景

- 根因分析
- 故障定位
- 异常预测
- 跨KPI寻找相关异常，减少标注开销

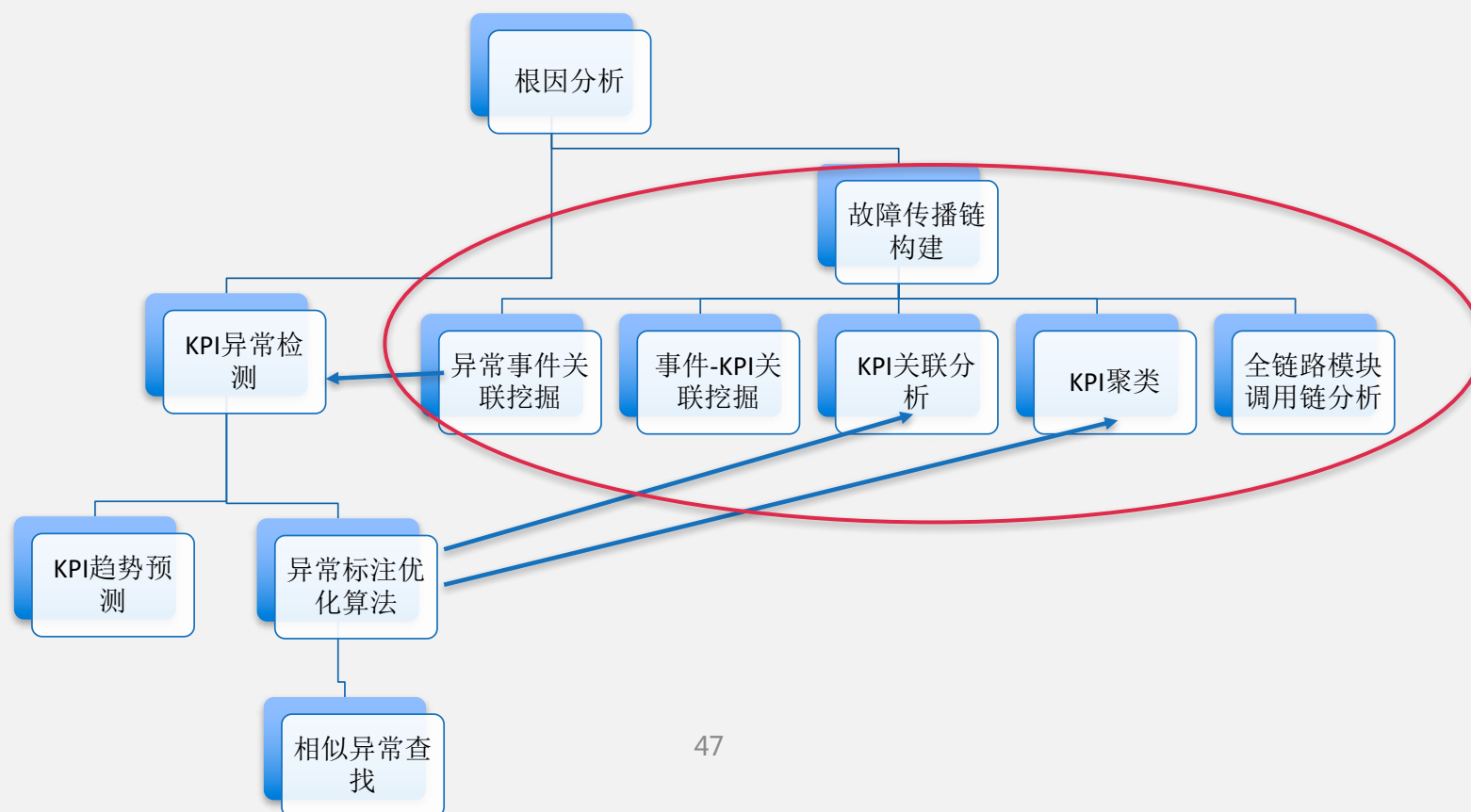
常见算法

- Pearson correlation
- Spearman correlation
- Kendall correlation
- Information gain
- Granger causality

应用挑战

- KPI种类繁多
- 关联关系复杂
- 无标注无监督

科研问题定义之 “庖丁解牛”



故障传播关系图构建算法

面向问题

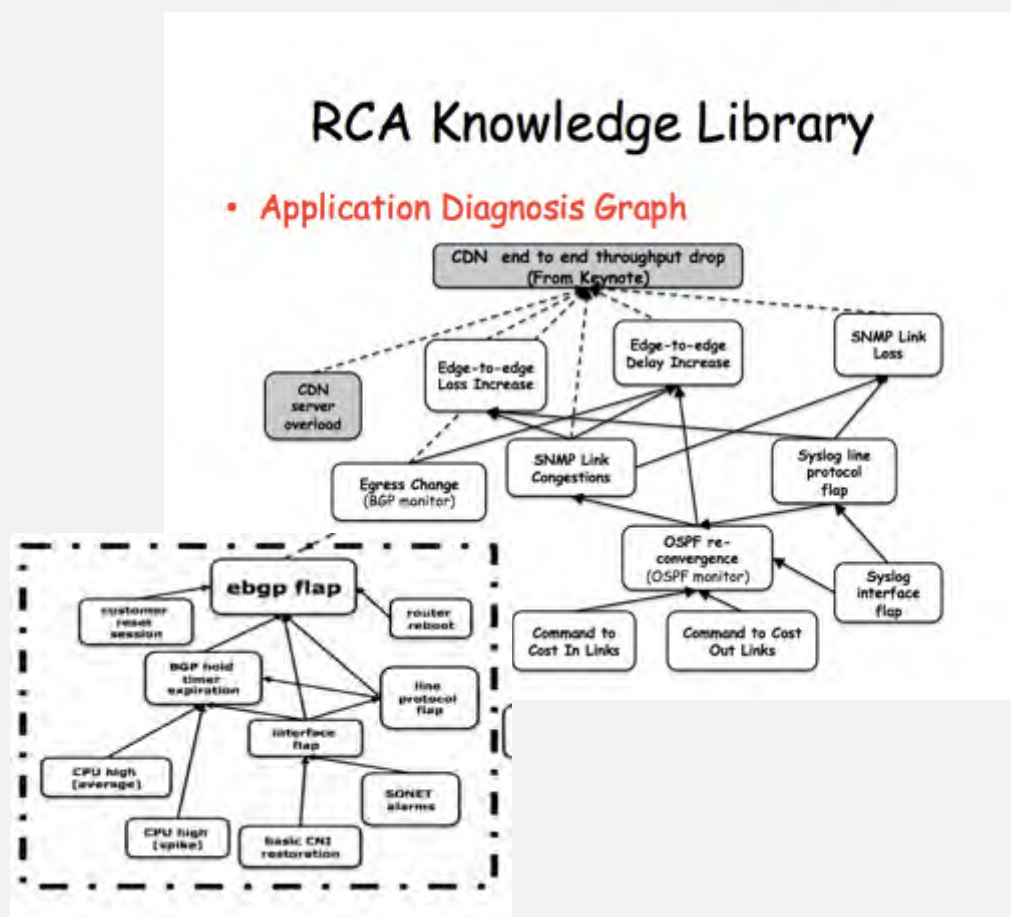
- 系统故障发生时，异常事件众多且具有相互导致关系。借助精准故障传播关系图，可以快速进行根因定位。

输入

- 历史异常事件，全链路调用链，异常关联，异常-KPI关联，KPI关联，KPI聚类

输出

- 故障传播关系图, 作为根因分析的输入



故障传播关系图构建算法

典型应用场景

- 根因分析

常见算法

- Dapper: call graph
- KPI 聚类算法
- KPI关联算法
- 事件关联算法：FP-Growth, Apriori
- 事件-KPI关联算法

应用挑战

- 异常检测需要准确可靠
- Ground Truth 难以获取
- Call graph 不一定有

异常事件关联规则挖掘算法

面向问题

- 分析异常事件两两之间的关联关系

输入

- 近段时间发生的异常事件

输出

- 异常事件的关联规则

time	异常事件
2014-10-29 06:09:10	http port unreachable
2014-10-29 06:09:10	high cpu usage
2014-10-29 06:10:10	page view number< 500
2014-10-29 06:11:10	mem usage
...	...

挖掘算法

关联规则
high cpu usage – mem usage
high cpu usage– page view number< 500
high cpu usage err – http port unreachable
http port unreachable – mem usage
...

典型应用场景

- 故障传播链构建
- 报警压缩

常见算法

- FP-Growth
- Apriori
- 随机森林

应用挑战

- 异常检测结果需要准确
- 关联的异常事件需要在历史数据中一起出现多次

■事件-KPI关联关系挖掘算法

面向问题

- 对事件与KPI的关联关系挖掘

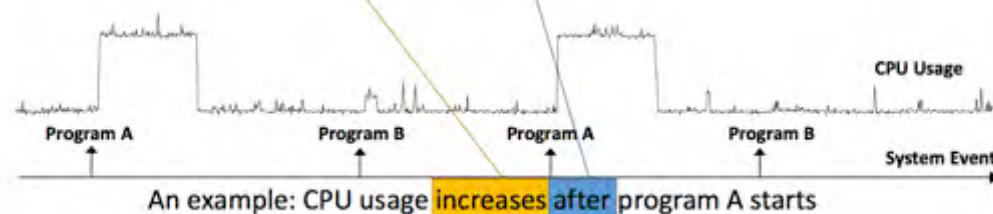
输入

- 一条KPI曲线，一条事件数据源

输出

- 是否相关，先后顺序，变化方向关系

- Detecting the existence of correlation
- Finding out temporal relationships
- Identify monotonic effects



事件-KPI关联关系挖掘算法

典型应用场景

- 故障传播链构建
- 根因分析
- 服务事件诊断
- 系统故障定位

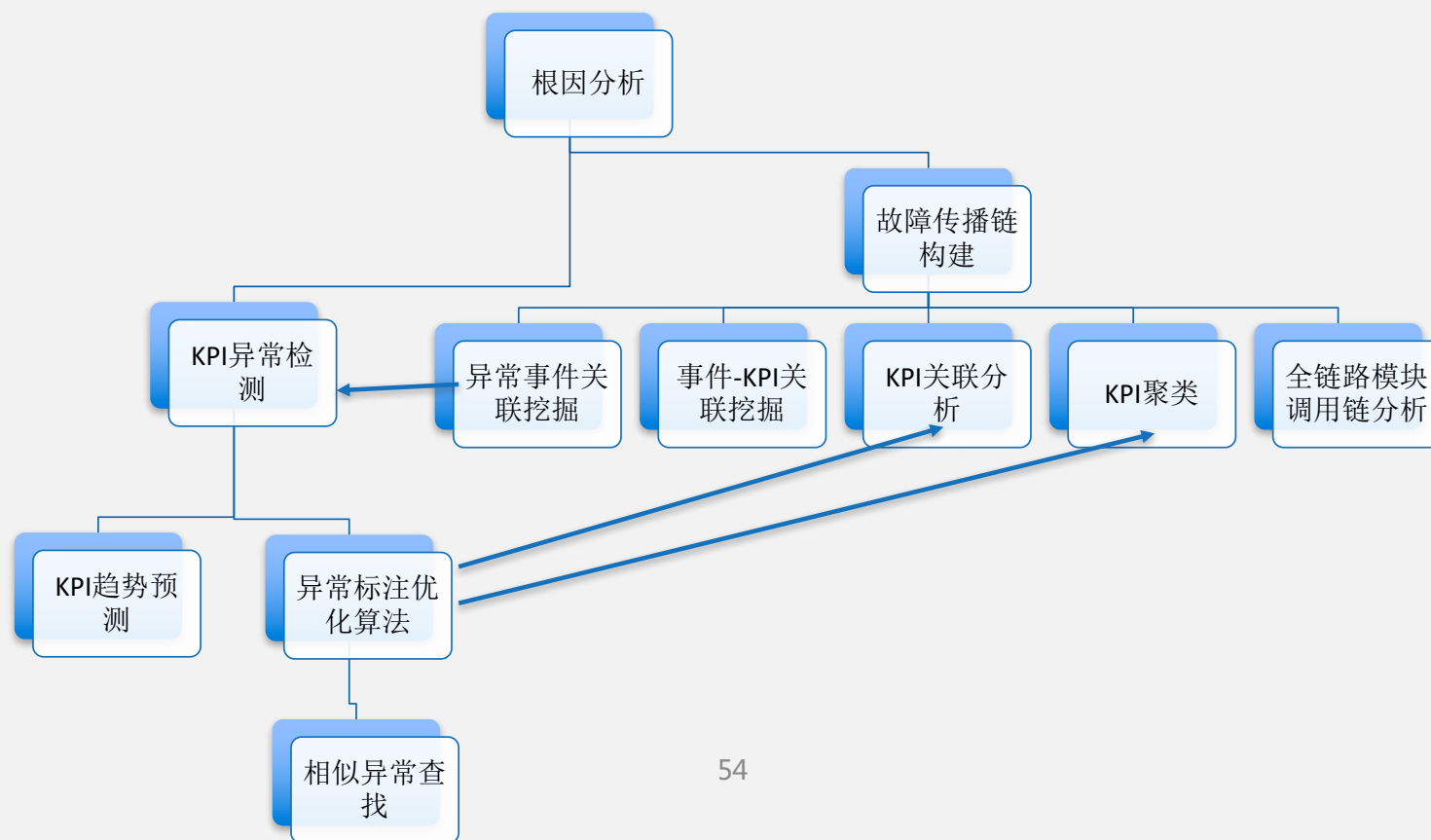
常见算法

- Pearson Correlation
- J-Measure
- Two-sample test

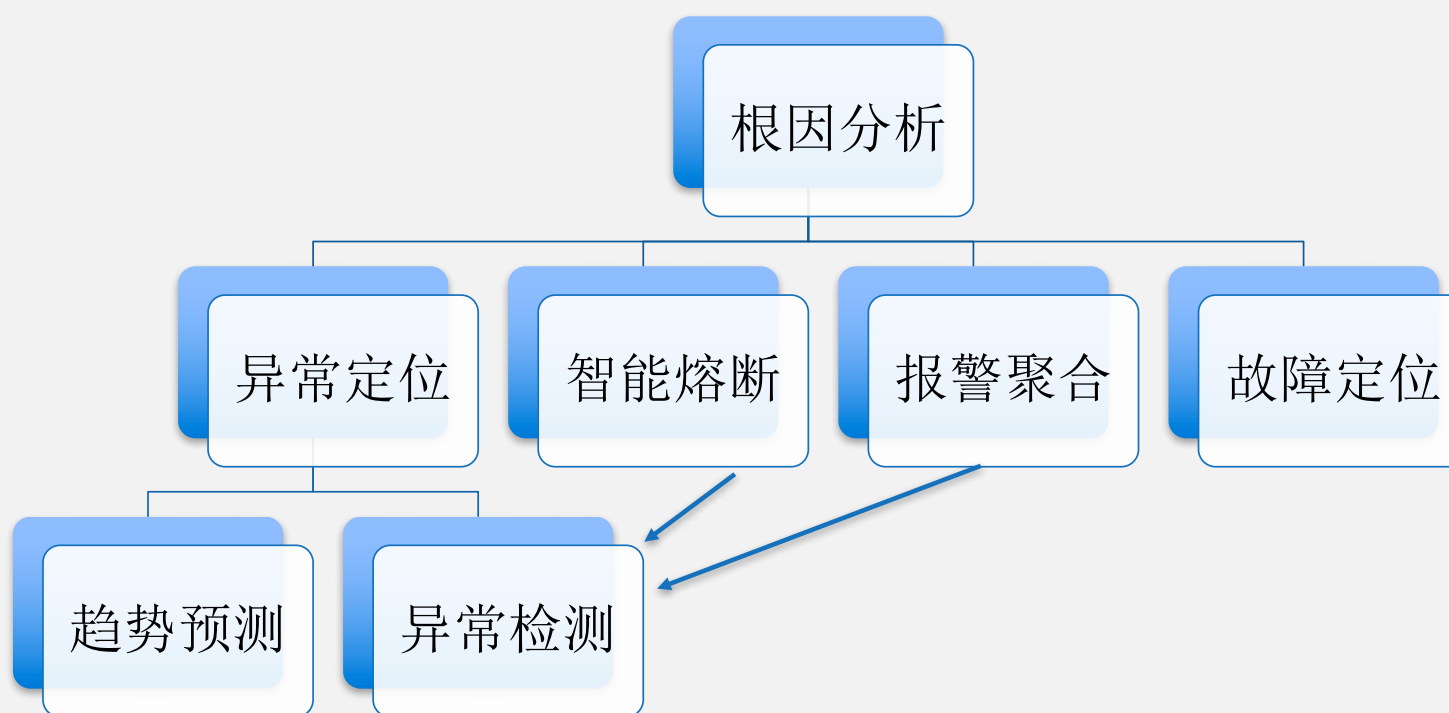
应用挑战

- 事件和KPI种类繁多
- KPI测量时间粒度过粗，导致判断相关、先后、单调关系困难
- 算法参数调整

科研问题分解之 “庖丁解牛”



科研问题定义之 “退而求其次”



异常定位算法

APMCon

面向问题

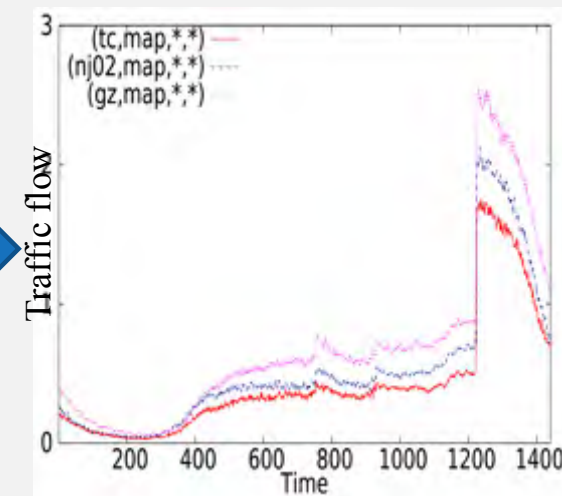
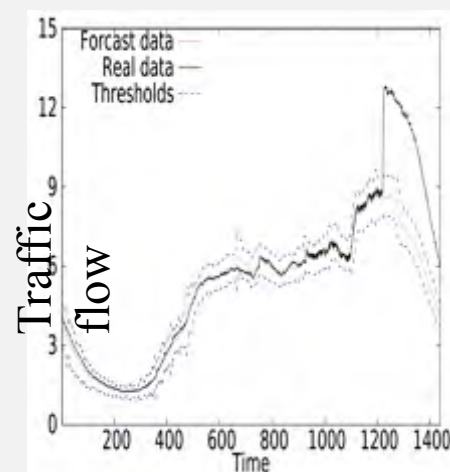
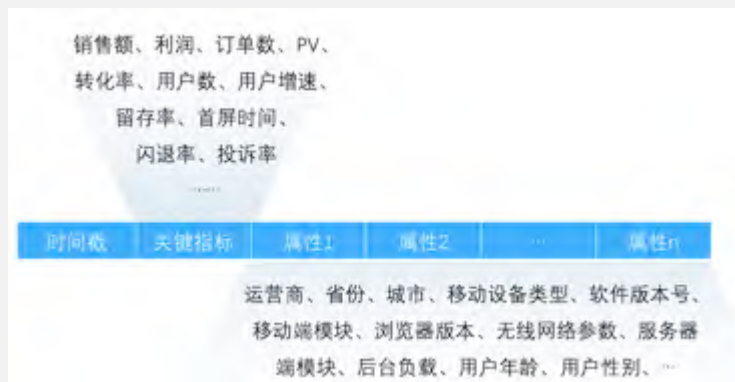
- 多维属性KPI指标的总量发生异常时，需定位到根因所在的具体的属性维度（或维度组合）

输入

- 多维属性KPI的粗细粒度的KPI监控数据

输出

- 导致异常的原因所在位置



典型应用场景

- 各类多维属性KPI (PV , 销售额 , 在线用户数 , 登录请求数...)的异常定位

常见算法

- Adtributor
(explanatory power, surprise)
- iDice(Isolation Power)

应用挑战

- 多维度指标搜索空间巨大
- 关键指标受多指标、多因素共同影响
- 历史异常信息数量不足
- 受异常检测准确度影响

智能熔断提示算法

面向问题

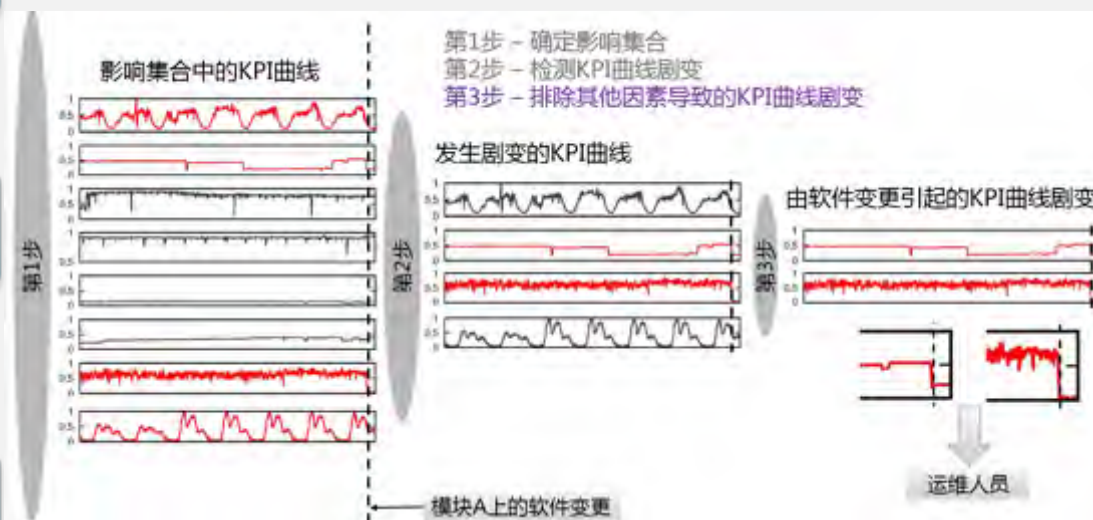
- 软件升级或配置变更后智能地提示是否应该回滚此次升级（变更）

输入

- 软件升级或配置变更
- 关联KPI时序数据

输出

- 是否应该回滚此次升级（变更）的建议



典型应用场景

- Web应用软件升级
- Web应用配置变更
- 数据中心网络设备软件升级
- 数据中心网络配置变更

常见算法

- CUSUM
- 奇异谱变换 (SST)
- Difference in difference

应用挑战

- KPI数量巨大
- KPI类型多样
- 要求低检测时延且高鲁棒性
- 其他因素干扰

■异常报警聚合算法

面向问题

- 监控KPI太多，粒度细，导致异常报警冗余度大

输入

- 异常检测的原始异常信息

输出

- 把相关报警聚合精简后的报警信息

报警聚合示意图：



典型应用场景

- 报警压缩

常见算法

- 基于拓扑层级关系
(如：服务-机房-集群-主机)
- 频繁项集挖掘 FP-Growth , Apriori
- 基于故障传播链

应用挑战

- 拓扑层级关系自动挖掘
- 要求异常检测准确性较高、历史异常信息数量较多

故障定位算法

面向问题

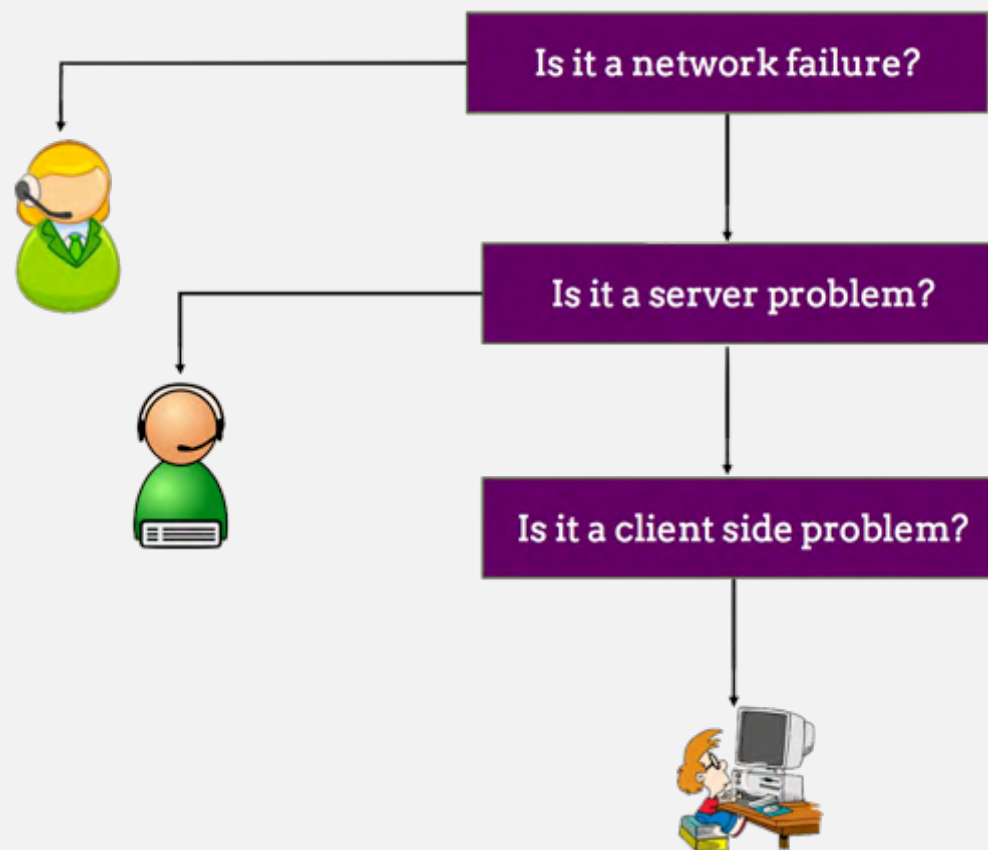
- 当前应用服务发生故障时，确定根因所在的大致位置(主机、数据库、网络、前端、客户端等)。

输入

- 客户端，网络，服务端等性能指标

输出

- 根因所发生的位置



故障定位算法

典型应用场景

- 服务发生故障，需要定位根因大致位置，加速止损

常见算法

- 故障模拟
- 随机森林
- 故障指纹构建
- 逻辑回归
- 马尔可夫链
- 狄利克雷过程

应用挑战

- 历史故障案例少
- 需要对案例进行人工标注
- 无法有效定位历史上没发生过的故障

落地智能运维科研算法



- 相对独立算法 -> 直接可落地
- 依赖其它算法 -> “庖丁解牛”
- 数据等条件不成熟 -> “退而求其次”

总结与前瞻



智能运维算法竞赛网站

APMCon

iOps

运维场景数据集竞赛科研问题

请输入你想要搜索的内容

知识库论坛

注册登录

中文

让您用上最好的智能运维算法

历史事件

瓶颈分析

热点分析

KPI聚类

KPI关联关系挖掘

异常事件关联关系挖掘

全链路模块调用链分析

故障传播关系图构建

当前事件

KPI异常检测

KPI异常定位

报警聚合

快速止损

故障根因分析

未来事件

故障预测

容量预测

热点预测

KPI趋势预测

■诚邀在座各位共同参与！

- 付出：

- 参照科研问题提供脱敏数据
- 资金赞助感兴趣的算法竞赛
- 建议新的科研问题
- 参与社区讨论

- 回报：

- 根据本公司实际问题，查询试用相关算法
- 根据网站建议，优化本公司数据采集和清洗工作
- 寻找潜在合作教授
- 在竞赛参与学生中招聘

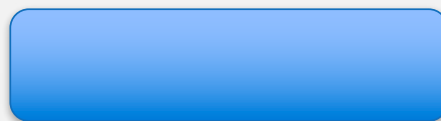
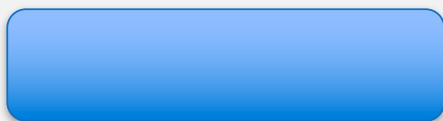
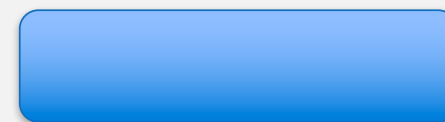
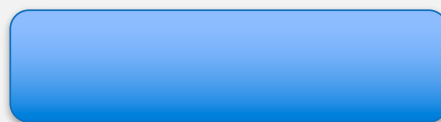
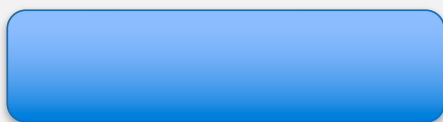
正在确认首批数据赞助商

APMCon

官方已经审批通过



口头意向：五家大厂



欢迎贵司参与！

感谢工业界合作伙伴

APMCon



• • • • •

感谢清华NetMan团队

APMCon



总结

智能运维前景光明

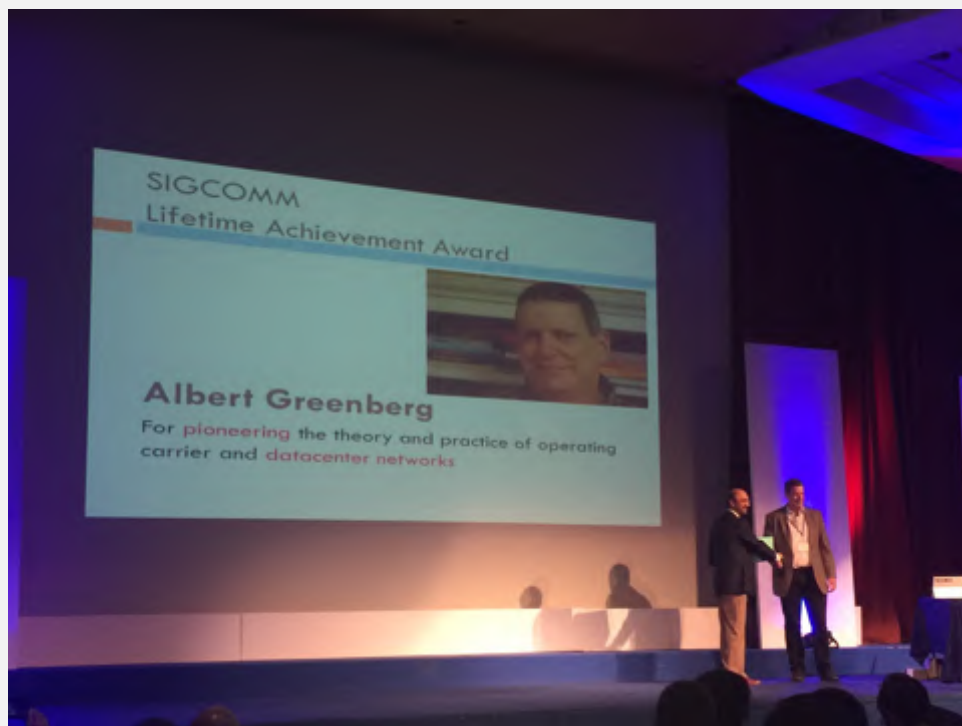
- 具有丰富的数据和应用场景
- 将极大提高运维领域的生产力
- 是AI领域尚未充分开采的金库和低垂果实

智能运维科研需要工业界-学术界密切合作，但是目前仅限于一对一合作：

- 合作效率低、见效慢
- 还是少数大公司和教授的特权
- 涉及知识产权，不符合开源大趋势

解决思路：科研问题为导向, 促进工业界-学术界合作 2.0

- 把应用难题分解定义成切实可行的科研问题
- 企业提供脱敏数据作为benchmark
- 学术界贡献算法



关于科研成果落地，我最推崇的 Albert Greenberg 的两句名句：

- “如何赢得学术顶会 **Test of time** 奖？论文发表后再花五年时间把论文里的算法变成产品。”
- “人们往往高估两年内能完成的成果；同时又往往低估五年内能完成的成果。”

谢谢！



清华大学 裴丹
微信: peidanwechat

《智能运维前沿》课程课件（英文）：
<http://netman.cs.tsinghua.edu.cn/courses/advanced-network-management-spring2017/>