智能运维系列（五）| 浅析基于知识图谱的根因分析系统

智能运维系列专题简介：智能运维（AIOps），根据 Gartner 的最新阐释，意指整合大数据和机器学习能力，通过松耦合、可扩展方式去提取和分析数据量（volume）、种类（variety）和速度（velocity）这三个维度不断增长的 IT 数据，进而为 IT 运维管理产品提供支撑。在此，微众银行智能运维团队根据一线工作的实践经验与心得体会，特别撰写了《智能运维系列》文章，敬请持续关注。

点击回顾：[智能运维系列（一） | AIOps 的崛起与实践](https://www.infoq.cn/article/fqUfkjhecOla1zKUKycN" \t "_blank)

点击回顾：[智能运维系列（二）| 智能化监控领域探索](https://www.infoq.cn/article/Qta6VCyjvHdoiJg5wKze" \t "_blank)

点击回顾：[智能运维系列（三）| 浅析智能异常检测——“慧识图”核心算法](https://www.infoq.cn/article/mryjNLXOlujV7fkQFUaL" \t "_blank)

点击回顾：[智能运维系列（四）| 曝光交易路径](https://www.infoq.cn/article/a72WeEMuM1O63iX1w0ZK" \t "_blank)

作为中国首家开业的民营银行和互联网银行，微众银行近年来业务发展非常迅速，海量客户和交易使IT 系统的轻微抖动就可能影响众多用户金融交易体验。为了保证业务正常运转，全面提升MTBF(平均故障间隔) 和MTTR(平均故障修复时间) 这两个数字化运维的关键指标，除了需要迅速检测出异常，还需要快速、准确、有效地分析出异常的根因，迅速恢复。为了实现这一目标，微众银行运维团队在智能根因分析方面付诸了长期的努力，研发出了智能根因分析系统。目前基于该系统的根因分析准确率稳定维持在80% 左右。下文将具体阐述该系统的设计理念。

**数据基础**

“不积跬步，无以至千里；不积小流，无以成江海”。不论是人类专家或者是计算机，在进行分析、推理、决策时都需要数据的支撑。因此数据的准确性、及时性和完整性在根因分析中非常重要。在实现智能根因分析的道路上没有捷径，基于配置管理系统的IT 运维系统群为其提供了坚实的数据基础。长期投入IT 基础工程的研发，构建了较为完整的运维体系，在此基础上开始了智能化运维实践，接下来简单介绍根因分析主要应用到的数据。

**配置数据**

配置数据主要从CMDB（配置管理）系统获取。CMDB 系统是众多运维工程师较为熟悉的系统，它包含了配置项全生命周期的信息以及配置项之间的关系(包括物理关系、逻辑关系和依赖关系)。从图1 可以看出，从业务层到基础架构层，配置项与配置项间的关联关系都完整保存在配置管理系统内。智能根因分析可以从中获取到关联配置数据进行相关性分析。



图1 配置项分层

**日志**

日志主要包括WEMQ 日志、业务流水日志和应用日志。WEMQ 日志是微众银行研发的消息总线系统所产生的日志，微众银行的系统间调用基本都要通过WEMQ 系统来完成。业务流水日志是业务模块输出的日志（格式化的业务交易日志），它的内容跟产品和场景息息相关，记录了业务相关的信息。应用日志是应用程序输出的日志，包括一些异常堆栈信息。通过WEMQ 消息的日志，我们可以分析出每一笔交易所经过的子系统及其调用情况，如下图所示：

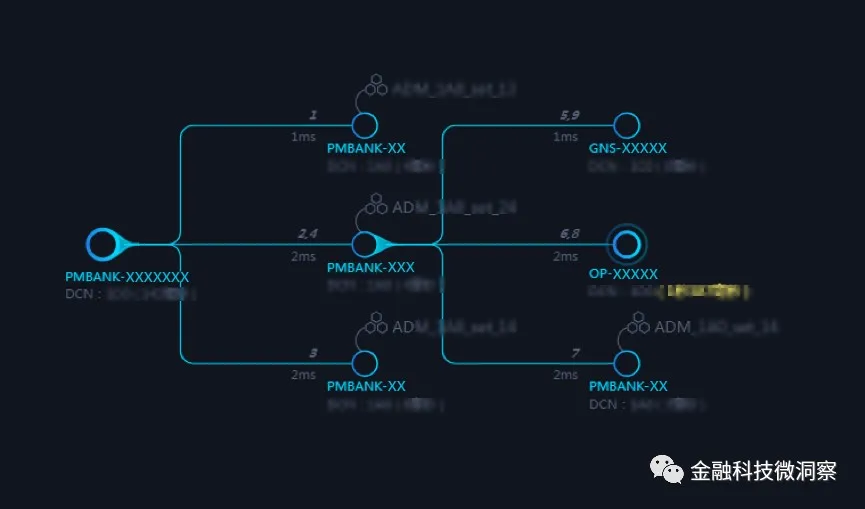


图2 交易调用树

**告警**

监控系统可以说是IT 运维的生命线，它集中采集了业务和基础架构相关的指标数据，并支持指标的实时计算，通过监控策略可以从这些指标或用户上报的数据中发现异常并生成告警，为IT 运维的故障诊断提供了完整的数据支持。

监控系统提供了两部分的数据用于智能根因分析：一部分是实时采集的时间序列数据，即指标数据；另一部分是基于指标计算或者其他第三方系统上报的告警信息。

**变更**

变更系统提供了数据库、系统版本发布、主机、网络等一系列的变更操作和记录，所有的运维操作都要求必须在该系统上完成，因此该系统记录了内部全量的变更数据。基于这些变更数据，智能根因分析系统能够获取到跟异常相关的运维操作数据，并结合其他数据来进行根因定位。



图3 变更视图

**技术选择**

在根因分析技术的选择上，我们初期进行了探讨和调研。在异常检测方面，我们采用了深度学习、机器学习等技术，取得了不错的效果。但在根因分析方面，我们决定先采用专家系统的技术来实现，主要有以下原因：

首先，“业务异常”的数据是“小数据”。正常的公司运营过程中，真正影响业务的异常事件数据本身就会很少，数据累积的速度会很慢。在“小数据”的基础上，机器学习在根因分析方面的应用相对有限。

其次，“根因分析”是需要有较强的解释性的，每次业务异常后，运维工程师都会有完整的异常事件分析报告，机器学习在可解释性上相对较弱，而专家系统能更好的解释根因是如何分析出来的，更符合人类的思考逻辑。

最后，利用人类专家“举一反三”的能力，可以短期内构建出根因分析系统。

因此我们先选择了专家系统的解决方案，将IT 专家的经验总结起来形成推导规则。

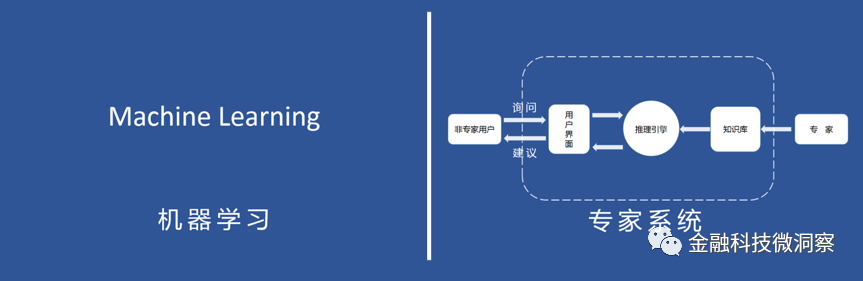


图4 机器学习和专家系统

**专家系统和知识图谱**

早期我们采用Drools 规则引擎实现了基于规则的根因分析系统，通过不断丰富和完善推导规则，很快就使其具备了根因分析的能力。但在应用一段时间后，发现这个方案还存在不足的地方。主要有两个方面：

首先是数据不透明。每一次异常发生过后，我们都需要检视根因分析是否正确。如果根因正确，那么就需要向团队内同步是基于什么样的数据和推理逻辑推导出根因的；如果根因错误，则需要检视是否存在数据缺失、推导规则错误等问题。团队把这一类工作叫做案例复盘。复盘就需要依赖当时获取到的异常数据来开展，原先我们把数据都放在TDSQL 内，彼此之间相互不关联，所以复盘时这些数据都是碎片化的，数据透明程度差，复盘也相对困难。后来引入了图数据库，将异常数据以知识图谱的方式来存储，即方便查询，也方便展示。

最后是规则维护困难。在根因分析系统早期版本的时候，推理模块是采用了Drools 来实现的规则引擎，虽然它解决了知识和代码的耦合问题，但规则越来越多以后，从单个规则来看也很难看出整体的推导逻辑，维护起来相对困难。经过调整，我们在图数据库的基础上，按照不同的异常类型编写了推导模型，通过模型就可以在图数据库中找到根因。这样只需要维护好模型即可，降低了规则的维护难度。

**根因分析设计**

根因分析的总体思路是在异常事件发生时，系统收集信息并生成该异常事件的知识图谱，在图的基础上运用演绎推理和归纳推理等方法来对事件根因进行分析。简单理解就是图 + 统计 + 规则。根因分析将针对四种指标类型的异常：时延、交易量、业务成功率、系统成功率，再经过三个步骤的处理和分析：信息收集、根因定位、根因补充，最终分析出根因。从这个思路来看，异常事件的知识图谱如何设计，是我们根因分析设计的关键。接下来将详细介绍该设计，进而阐述根因分析设计的细节。

**异常事件的知识图谱设计**

异常事件的知识图谱是结合“动”态和“静”态数据来设计，“动”态数据包括业务流水相关的日志、证据数据，“静”态数据则来自于CMDB 等配置系统。两类数据共同构建起一个完整的异常事件图谱。如下图所示，从图上可以看出，图谱是有方向的，从左到右进行根因的分析和推导，最终分析出根因。

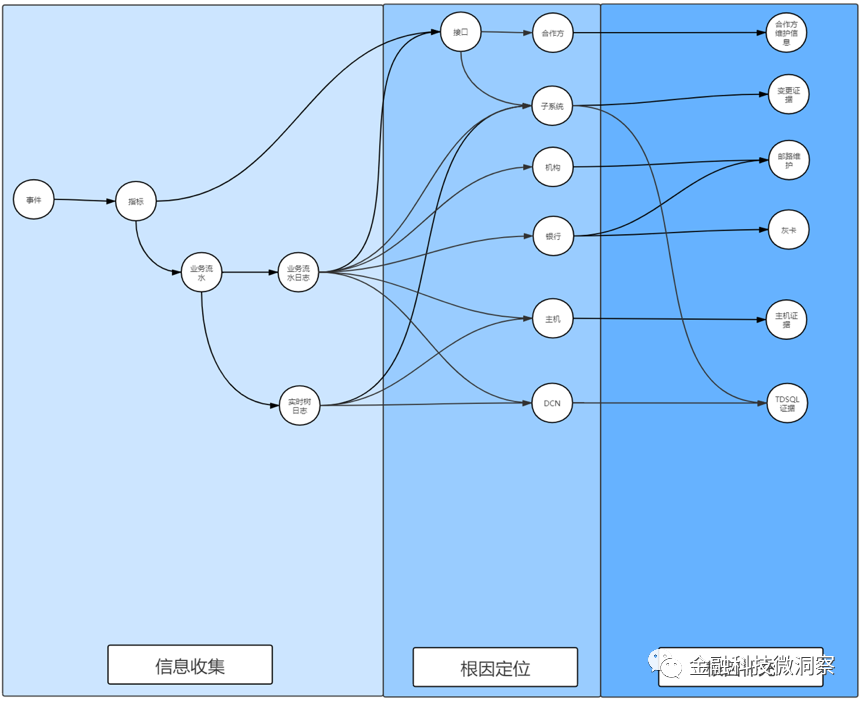


图5 异常事件的知识图谱设计

一般来说，知识图谱设计及根因分析一般包括信息收集、根因定位、根因补充三个阶段。

首先是信息收集阶段，该阶段将收集用于构建知识图谱的完整信息，主要收集以下几个维度的信息：

1、事件：异常事件的起点，包含了异常事件的相关信息，例如事件的开始时间等。

2、指标：产品的关键指标，我们选择了四种类型的指标作为黄金指标，用于检测产品业务是否异常，每个场景都有其对应的黄金指标，其中包括：

（1）交易量：单位时间内的交易笔数。

（2）业务成功率：单位时间内的业务成功率，业务失败时该指标会下降。业务失败是指符合业务逻辑的失败，例如验证码未通过。

（3）系统成功率：单位时间内的系统成功率，系统失败时该指标会下降。系统失败时指系统内部的失败，例如连接数据库异常。

（4）时延：单位时间内完成交易的总耗时时间。

3、业务流水：用户在产品上操作所产生的编号，每一次操作都会产生一个唯一的编号，也称为一笔交易。这个编号可以关联出业务日志和实时树日志（WEMQ 日志）。

4、业务流水日志：每个系统按照规范打印的业务相关日志，可以从中查到具体的业务参数和相关调用信息。除了业务相关的信息外，日志中还有交易发生的子系统、主机、DCN 等信息。

5、实时树日志：之前提到的 WEMQ 日志，可以将交易的完整调用路径分析出来，其中还有经过的主机、耗时等明细数据。

其次是根因定位阶段，该阶段根据收集到的日志数据，对交易进行统计分析，并定位到哪个子系统、主机才是本次事件的根因。以系统成功率为例，在异常发生时，有多笔交易产生了错误日志，这时对异常时间点的交易信息进行提取，找到当时存在 n 笔交易是异常交易。对这 n 笔交易进行统计分析后，发现这些交易都在某个相同的子系统报错了，说明该子系统就是根因子系统。

最后是根因补充阶段。这个阶段会采用告警、变更等数据来进行分析，并对根因进行补充。如果定位到一个子系统存在异常，该子系统如果还存在告警、变更等数据的话，就会进一步的补充到根因内，使得根因更加具体、明确。

**实际案例**

整个系统的根因分析工作分三个阶段。接下来将以实际案例来简单介绍智能根因分析系统是如何工作的。

**阶段一：信息收集**

信息收集阶段以事件中心，关联查询异常相关的日志、告警、变更、配置等信息，用以构建该异常事件的完整知识图谱。

1、以事件为起点，关联查询本次异常事件相关的指标信息。

2、通过获取到异常时间点的业务流水信息，连带查询出对应业务流水号可以关联出来的业务流水日志和实时树日志。

3、获取当前存在的证据。

4、将所有数据写入到图数据库，生成知识图谱。

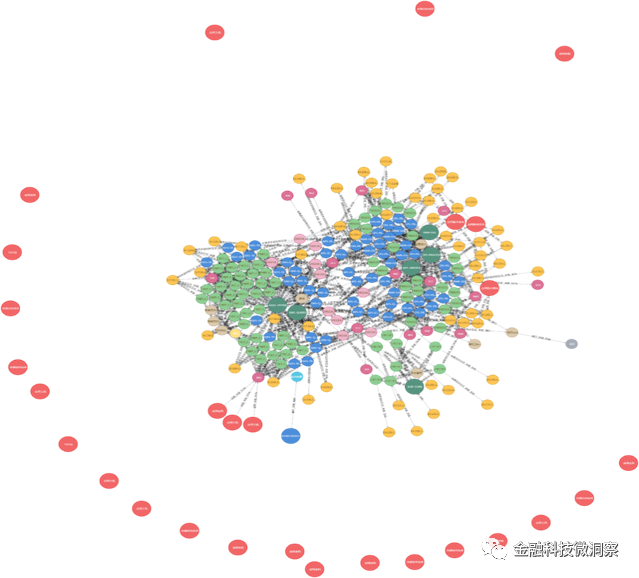


图 6 异常事件的知识图谱

**阶段二：根因定位**

根因定位阶段是在异常事件知识图谱的基础上，应用推导模型，对存在异常的子系统，及其相关的、IP、DCN、服务信息进行提取，起到定位的作用，也对异常事件的知识图谱进行了裁剪。

如图 7 所示，应用图分析的推导模型后，异常的子系统及其相关 IP、DCN 及其证据从知识图谱中提取出来。

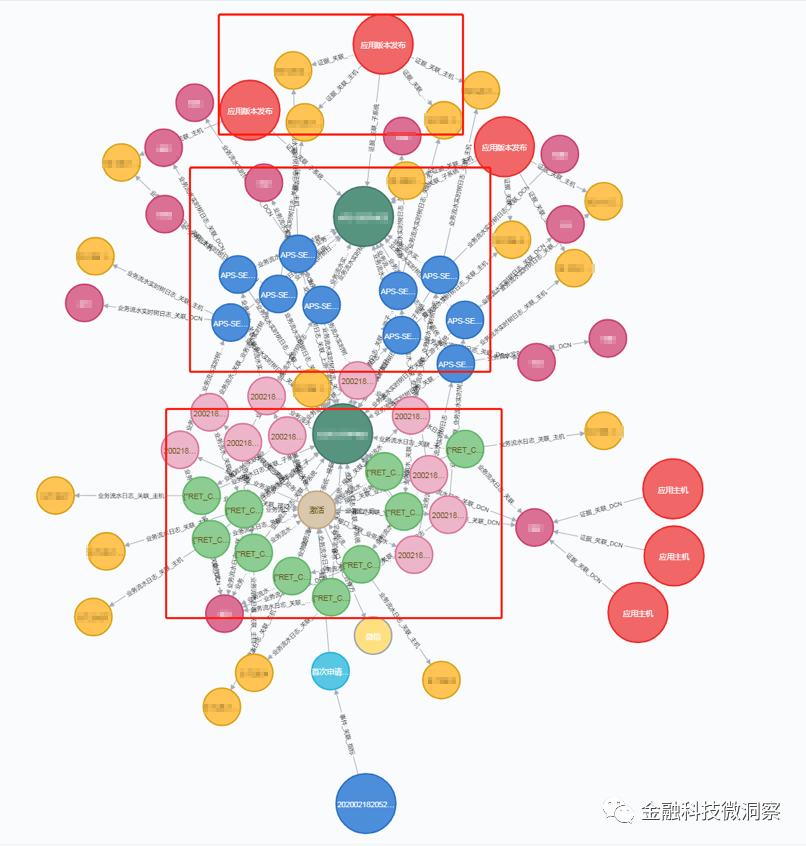


图 7 应用了推导模型后的知识图谱

**阶段三：根因补充**

根因补充是为异常事件进行最终根因定性的阶段，在阶段二的数据基础上，应用规则引擎最终推导出根因结论。

从阶段二中已经可以清晰发现该事件的根因。所有高耗时实时树日志都指向的 APS 子系统，而此时高耗时实时树日志经过的主机和 APS 子系统都能关联到应用版本发布信息。通过该图获取到的信息，经过话术的加工，最终智能根因分析系统给的异常事件根因是：[应用版本发布] 子系统 6009 AOMP 应用发布。影响: 上游 DSFS 接口 激活, 返回信息: 激活成功 (交易耗时异常升高)。

上述是时延异常的一个案例，通过三个阶段的分析，最终定位到是应用版本正在发布，导致服务的耗时异常升高。整体思路是通过收集系统异常时的相关信息，构建异常事件的知识图谱，再应用推导模型对图谱进行信息提取，最后应用规则引擎对根因进行推导。

**结语**

回顾这一年多的系统建设历程，我们最初梳理了根因分析的数据基础，在这些数据基础上确定了根因分析的方法，即以业务流水日志为抓手，将各纬度数据进行整合和分析；选择专家系统的方案来快速构建根因分析系统；还应用了图数据库和知识图谱的技术来解决数据透明和根因推导的问题。根因分析系统后继还会持续优化，历史上异常事件的知识图谱对我们也是宝贵的财富，记录了所有异常当时的全貌，为我们挖掘更多知识和优化推导逻辑提供了数据支持。

后续我们将持续推出关于智能根因分析更加详细和深入的分享，欢迎大家持续关注。