# Graph-based root cause analysis for service-oriented and microservice architectures

作者：Álvaro Brandóna,∗ , Marc Solé

### 大意

对于困境的描述：

Service-oriented architectures and [microservices](https://www.sciencedirect.com/topics/computer-science/microservices) define two ways of designing software with the aim of dividing an application into loosely-coupled services that communicate among each other. This translates into rapid development, where each service is developed and deployed by small teams, enabling continuous shipping of new features and fast-evolving applications. However, the underlying complexity of this type of architecture can hinder observability and maintenance by the user. In particular, identifying the root cause of an anomaly detected in the application can be a difficult and time-consuming task, considering the numerous services and connections to be examined.

实验比较的结果是：基于图的比较比基于机器学习的方法更为有效。

基于对架构的图表示，建立了一个根因分析框架。该框架收集RCA(root cause analysis)过程真正重要的信息，如架构的拓扑图或者系统不同element之间的connections。

建立框架的目的是：1. 让用户更好的理解当前系统的行为。 2. 将现在系统的异常情况(anomalous situation)和以前被诊断的situation进行匹配.

### 相关工作

#### RCA

相关综述：

* Wong, W.E., Gao, R., Li, Y., Abreu, R., Wotawa, F., 2016. A survey on software fault localization. IEEE Trans. Softw. Eng. PP (99), 1. doi:10.1109/TSE.2016.2521368.
* Solé, M., Muntés-Mulero, V., Rana, A. I., Estrada, G., 2017. Survey on models and  
  techniques for root-cause analysis. arXiv:1701.08546.

基于模型的诊断框架：

* Joshi, K.R., Hiltunen, M.A., Sanders, W.H., Schlichting, R.D., 2005. Automatic model–  
  driven recovery in distributed systems. In: Reliable Distributed Systems, 2005.  
  SRDS 2005. 24th IEEE Symposium on. IEEE, pp. 25–36.
* 应用于web服务当中：Mayer, W., Friedrich, G., Stumptner, M., 2010. Diagnosis of service failures by trace analysis with partial knowledge. In: International Conference on Service-Oriented Computing. Springer, pp. 334–349.

基于分类的方法：  
（使用ML的方法进行诊断）

* Demetgul, M., 2013. Fault diagnosis on production systems with support vector  
  machine and decision trees algorithms. Int. J. Adv. Manuf. Technol. 67 (9-12),  
  2183–2194.

使用图的方法：

* 基于Probabilistic Graphical Models (PGMs)：Schoenfisch, J., Meilicke, C., von Stülpnagel, J., Ortmann, J., Stuckenschmidt, H.,2017. Root cause analysis in it infrastructures using ontologies and abduction in Markov logic networks. Inf. Syst..
* Akoglu, L., Chau, D.H., Vreeken, J., Tatti, N., Tong, H., Faloutsos, C., 2013. Mining connection pathways for marked nodes in large graphs. In: Proceedings of the 2013  
  SIAM International Conference on Data Mining. SIAM, pp. 37–45.

# Survey on Models and Techniques for Root-Cause Analysis

### 基本信息

##### 作者：

Marc Sol ́e…

##### 研究问题和特点：

* 对RCA用到的技术和模型进行总结，
* 在性能和可扩展性上对不同需求的的技术进行分析。
* 引导，如何针对于特定的系统选择合适的RCA策略。

##### 文章结构

* Introduction
* 概念和术语介绍
* 主要的models for RCA
* 用于模型的inference algorithm
* 总结

### 重点内容

##### 核心概念

causality（因果关系 ）和 explanation（ 解释，说明，理由，原因，）通常被认为是RCA的output

##### 术语Terminology

Event: an exceptional condition

Fault/problems/root causes: 是指源头的Events，只有它造成别的event，而非被别event导致。分为三类—permanent/ intermittent(间歇的) /transient

Error：由一个或者多个Fault造成

Failure：从系统外可以观测的Error

Symptom

Root Cause Analysis：fault localization、fault isolation、alarm/event correlation、

**指标**:

intent目标、time时间、Complexity复杂度（系统size、数据的size、**Inference Length**、Effect propagation time、Evolution rate–被诊断的系统的变化速度和程度）、Domain knowledge required、System knowledge required

**RCA workflow**:

##### RCA模型

许多的推断模型都来自AI，大致分为两类：1. Deterministic确定式 2. Probabilistic概率式

图表简要展示不同类型的RCA模型

…

为什么一些ML中的分类器没有占领RCA领域？

1. 它们通常只返回一个预测的root cause，无法获得解释
2. 很难和domain knowledge结合起来，不能产生逻辑规则
3. 它们往往被简化成了single-label classifier，当需要进行多种故障的检测时，需要和多种classifier协同工作

什么因素对诊断的性能影响最大？

* Size：可以理解为是基本的分析元素的数量，通常是组件component的数目，但是也可以有多个Size属性。
* Inference structure：它定义了不同的elemental analysis element是如何相互关联到一起的。通常利用**Inference Length**指标来进行判别。

##### 如何获得一个RCA模型？

1. Manual generation
2. Assisted generation：根据相当数量的以子模型库的形式出现的domain knowledge 和向下的system information ，产生最后的诊断模型。
3. Automated generation

##### learning model for RCA

当没有domian knowledge可用时，只能采用learning algorithm，学习raw data of the system。

learning algorithm 往往能够学习structure 和 parameter。

##### RCA模型的推断Inference for RCA model

对分类器的模型而言，推断的过程只涉及了使用模型对symptom进行分类。

对于其他模型而已，存在不同的算法来进行实现。

分为两种

# Microservices The Journey So Far and Challenges Ahead

未来挑战：

* 服务的模块化与重构 Service Modularization and Refactoring
* 服务粒度 Service Granularity
* 前端集成 Front-end Integration
* Resource Monitoring and Management
* Failure, Recovery, Self-repair
* Organizational Culture and Coordination
* Addressing the Chanllenges