中国地质大学自动化学院
School of Automation . China University of Geosciences

PCIC2021: Causal Discovery

—— Causality analysis of complex network system based on TTPM

Team Member: Xiangxiang Zhang

Pan Zhang

Xin Cheng

Supvisior: Wenkai Hu

中國地質大學 自动化学院
School of Automation . China University of Geosciences

Member Profile

- Ph. D. student, School of Automation, China University of Geosciences
- B. Eng., School of Automation, China University of Geosciences (2020)
- Research interests: Causal reasoning and root cause analysis

Xiangxiang Zhang

- Master student, School of Automation, China University of Geosciences
- B. Eng., School of Automation, China University of Geosciences (2021）
- Research interests: Causal analysis of complex network system

Pan Zhang

- Master student, School of Automation, China University of Geosciences
- B. Eng., School of Electrical and Electronic Engineering, Hubei University of Technology (2020)
- Research interests: Correlation analysis of industrial alarms

Xin Cheng

中国地质大学 自动化学院
School of Automation . China University of Geosciences

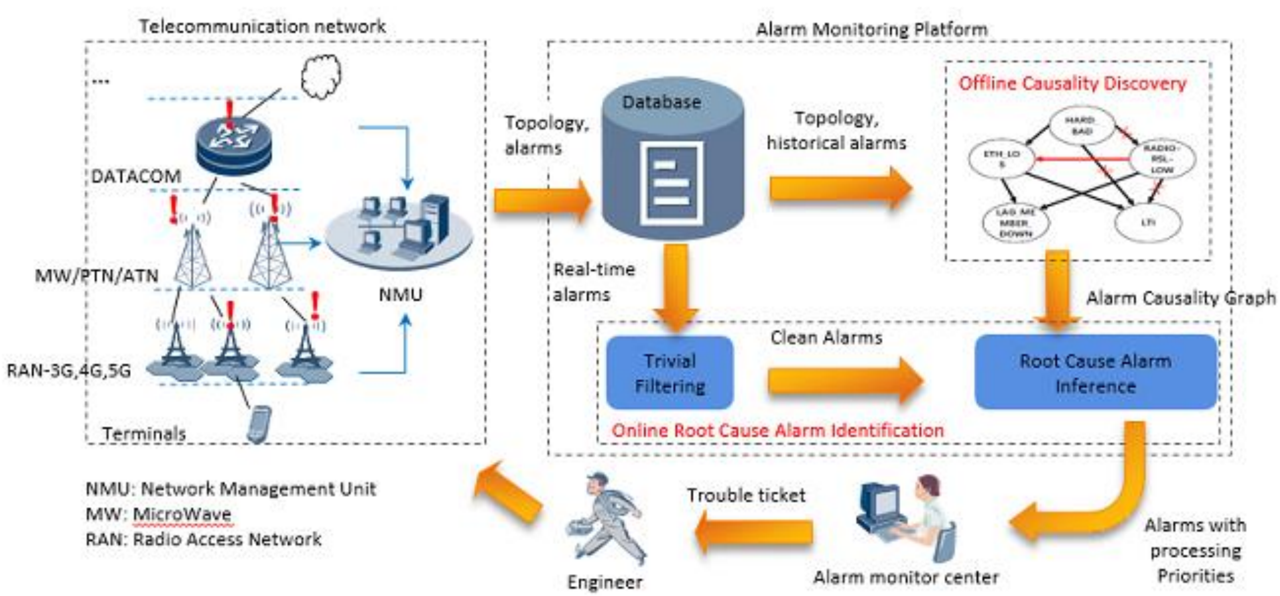
Outline

School of Automation . China University of Geosciences

1. Problem Description

Telecommunication networks root cause analysis

- The graph of alarm causality is given by using data
- Quickly and effectively locate the root cause alarm of a device fault
- Identify more true causal relations and less false causal relations



中国地质大学 自动化学院
School of Automation . China University of Geosciences

2. Framework

According to the above problems, we will introduce the main framework

Data Processing

- Data cleaning
- Data transformation

Ground-truth datasets analysis

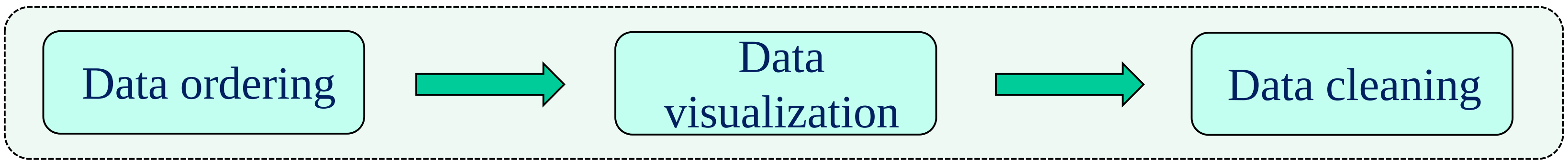
Form a basic model(TTPM)

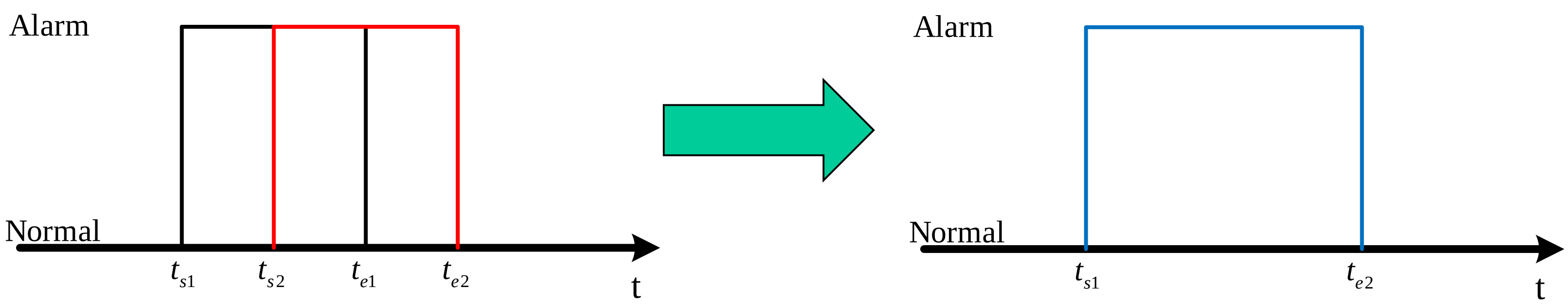
Training other datasets

- Program optimization
- Important parameter setting

3. Data Processing

- Data cleaning





Alarm-id	Device-id	Start_time	End_time
6	0	362167	362190
6	0	362178	362182
6	0	362178	362182

Alarm-id	Device-id	Start_time	End_time
6	0	362167	362190

3. Data Processing

- Data transformation
  - Dataset 1 and 2 in phase2
    - ✓ Data missing
    - ✓ Start_time is the same as end_time
  - Data deletion
    - ✓ Reduce interference with alarm data
    - ✓ Less false causal relations

Alarm-id	Device-id	Start_time	End_time
0	7	317669	317699
0	7	319502	319502
0	9	103271	103271
0	9	103880	103880
0	9	266701	266701

Alarm-id	Device-id	Start_time	End_time
0	9	103271	103271
0	9	103880	103880
0	9	266701	266701

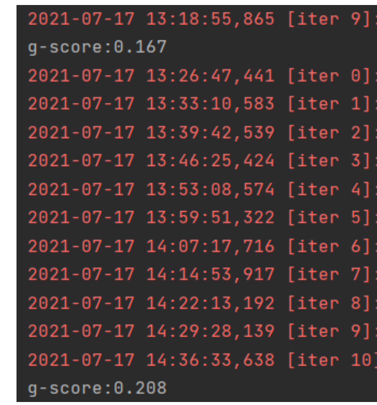
中国地质大学 自动化学院
School of Automation . China University of Geosciences

4. Causal Inference

➢ Program optimization

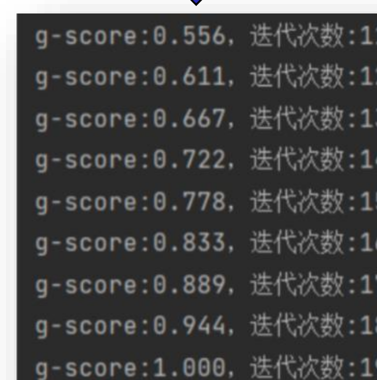
Select the number of iterations, only output the final result.

- the debugging speed is slow
- time-consuming



Each iteration will output a result without exiting.

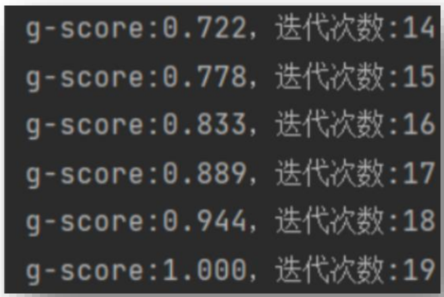
- reduced running time
- convenient to observe the law of output results

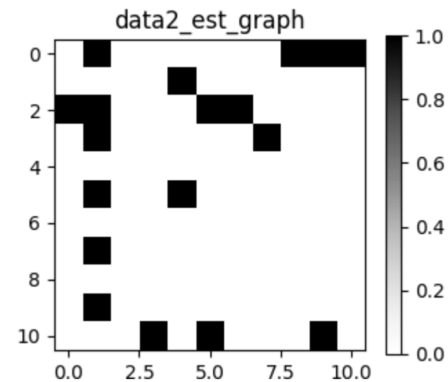


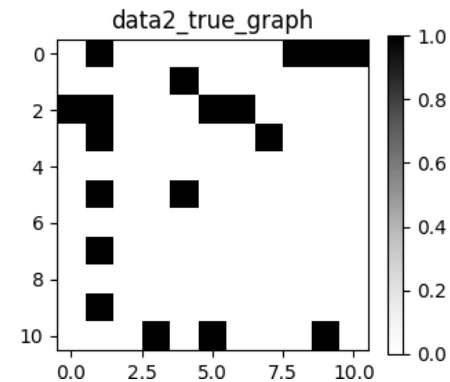
4. Causal Inference

➢ Training results

In phase 1, the use of a given eight contains ground-truth we found TTPM parameter regulation







Training data	iteration	g-score	Δg-score
Data 1	max_iter=11	1.0	0.091
Data 2	max_iter=18	1.0	0.056
Data 3	max_iter=20	1.0	0.05
Data 4	max_iter=20	1.0	0.05

中国地质大学自动化学院
School of Automation . China University of Geosciences

4. Causal Inference

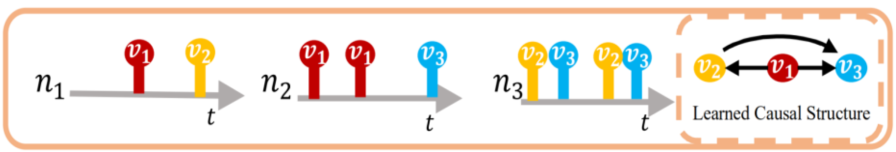
- Parameter: iteration

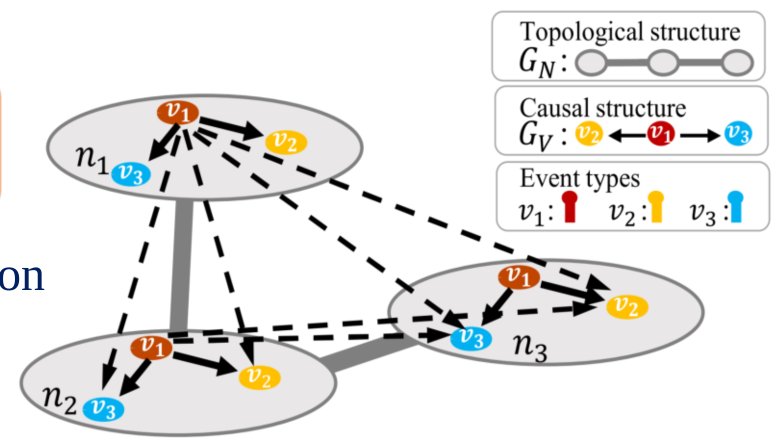
1. Each iteration: get a directed edge (“1”)
2. The optimal number of iterations b is equal to the number of all directed edges (“1”) n. That is $b = n$
3. g-score of x iterations, where $x < n$
4. b = [x/g-score], where [] denotes a rounded integer
5. Use the optimal number of iterations b to get the final result

中国地质大学自动化学院
School of Automation . China University of Geosciences

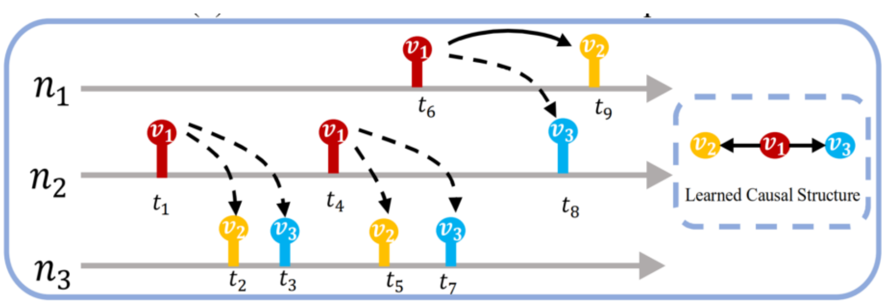
4. Causal Inference

- Parameter: max_hop
  - For data with Topolopy.npy files, the topology needs to be considered[1]
    - ✓ False causality

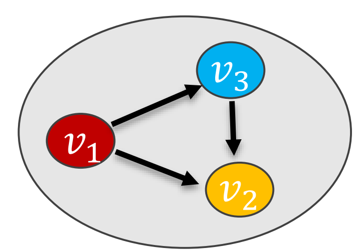




    - ✓ Correct Detection with topology Information



  - For data without Topolopy.npy files, max_hop=0
    - ✓ Think that the nodes are independent of each other



[1] Cai R, Wu S, Qiao J, et al. THP: Topological Hawkes Processes for Learning Granger Causality on Event Sequences[J]. arXiv preprint arXiv:2105.10884, 2021.

5. Lab Introduction

➤ 研究方向简介

“工业数据挖掘与智能安全监控”

面向复杂工业系统对安全保障的需求，研究工业数据挖掘与智能安全监控技术，主要涉及报警系统设计、报警泛滥模式挖掘、故障诊断与预测、因果关系分析、智能感知与决策等

❑ 智能报警监控与故障诊断

- √ 工业智能监控报警系统优化设计
- √ 报警泛滥序列模式挖掘
- √ 融合多源信息的故障检测与诊断
- √ 监控性能评估和报警数据可视化

❑ 复杂工业系统数据挖掘

- √ 工业事件数据模式挖掘技术
- √ 因果关系分析与根源推理技术
- √ 过程工业软测量智能建模方法
- √ 工业数据可视化技术

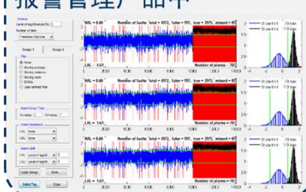
软件：AMTool

报警模式挖掘
报警系统设计
过程关联分析
可视化分析

监控报警数据+连续过程数据+连接图

ENBRIDGE SUNCOR ENERGY SPARTAN Controls EXXON Syncrude BBA Saudi Aramco Honeywell

Honeywell
报警系统设计方法嵌入其报警管理产品中



Suncor Energy
报警率下降44%，报警泛滥减少77%

Syncrude Inc.
制氢厂和油砂萃取厂低优先级扰动报警减少90%

为工业安全运行提供可靠的决策支持，以减少甚至预防工业事故的发生，提高工业制造过程的安全性和生产效率

中国地质大学 自动化学院
School of Automation . China University of Geosciences

导师简介

胡文凯 教授

- 中国地质大学（武汉）教授，博士生导师，智能系统研究所副所长
- 加拿大阿尔伯塔大学电气与计算机专业博士（2016）
- 加拿大阿尔伯塔大学博士后研究员（2016.9—2018.9）
- 湖北省高层次人才计划青年人才、湖北省“楚天学者计划”青年人才、中国地质大学“百人计划”学科骨干人才等
- 期刊Frontiers in Chemical Engineering，Editorial Board
- 期刊Entropy客座编辑、Journal of Beijing Institute of Technology客座编辑
- 神经计算与先进应用国际会议（NCAA）程序委员和征稿主席、中国控制会议CCC2020 Session Co-Chair等

主要学术成果：

- 发表论文58余篇，其中高质量学术论文23篇，获第30届CPCC张钟俊院士优秀论文奖
- 主持国家自然科学基金青年项目1项、湖北省自然科学基金青年项目1项、中国地质大学人才启动经费项目1项，参与加拿大自然与工程研究基金(NSERC)项目2项
- 作为召集人在IFAC World Congress 2020、SICE 2019、CCECE 2019等国际学术会议上召集和组织会前专题研讨会3场

主要项目：

- 基于多源数据融合的工业报警泛滥分析与抑制方法，国家自然科学基金青年项目，26万
- 面向工业报警事件混合数据的报警响应智能决策模型，湖北省自然科学基金青年项目，5万元
- 大规模复杂工业系统智能监控与大数据分析研究，中国地质大学人才引进项目，200万

Thank you！

✓ Some information about us:

- **Team: cug_402**
- **Email: wenkaihu@cug.edu.cn**