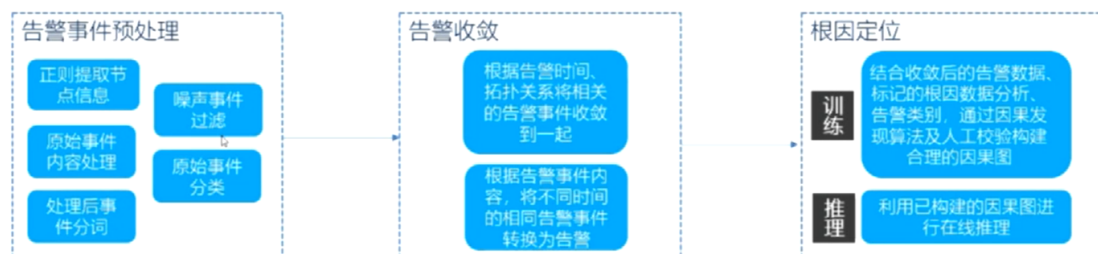


## 常规流程

# 赛题解析

## 根因定位常规流程

根因定位主要涉及数据分析、图论、信息论、自然语言处理等技术，整体流程如下：



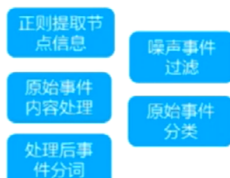
## 预处理

# 赛题解析

## 告警事件预处理

- 1、正则提取节点信息
  - 主要用于将告警事件数据和网络拓扑中的节点关联
- 2、原始事件内容处理
  - 主要用于将原始告警事件内容中的url、数字、路径等对根因分析无影响的部分去除
- 3、事件内容分词
  - 分词主要由于后续事件过滤、分类
- 4、噪声事件过滤
  - 结合标记的根因数据，分析噪声事件的特征，选择合适的算法过滤
- 5、原始事件分类
  - 将内容相似的告警事件分为一类，便于后续构建因果模型

### 告警事件预处理



## 特征工程

# 赛题解析

## 告警收敛

- 1、告警收敛
  - 时间维度
  - 拓扑维度
  - 将相关的告警事件收敛到一起，便于后续因果分析及建模
- 2、事件转换为告警
  - 原始事件数量较多，不便于分析，可以考虑将内容相同，时间不同的事件转换为告警，便于后续因果分析及建模

### 告警收敛

根据告警时间、拓扑关系将相关的告警事件收敛到一起

根据告警事件内容，将不同时间的相同告警事件转换为告警

## 建模--根因定位

这部分可暂时不考虑神经网络、分类算法，只考虑因果图算法。

# 赛题解析

## 根因定位

- 1、训练
  - 训练主要基于收敛后的告警数据
  - 可以通过因果发现算法构建抽象的因果图
  - 标记数据主要用于训练结果的分析验证和迭代
- 2、推理
  - 利用训练得到的因果图进行在线推理，计算出疑似根因告警

### 根因定位

训练

结合收敛后的告警数据、标记的根因数据分析、告警类别，通过因果发现算法及人工校验构建合理的因果图

推理

利用已构建的因果图进行在线推理

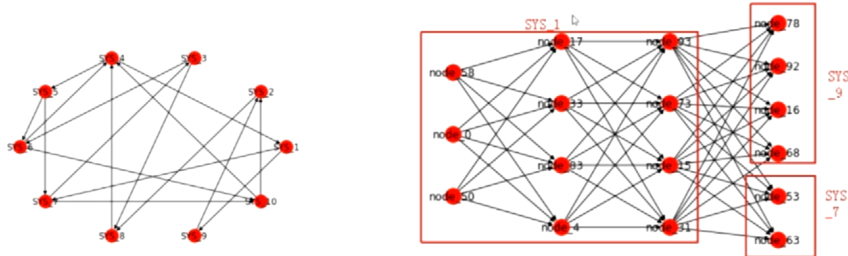
## 数据解析

# 赛题解析

## 训练数据解析

### 2. 网络拓扑

- 网络拓扑共100个节点，node\_0 – node\_99。共10个系统，SYS\_1 – SYS\_10。每个节点都属于一个系统。
- 系统间的关联关系，如下图（左）所示。
- 同一系统内的节点也存在关联关系，以SYS\_1为例，如下图（右）所示。比如SYS\_1与SYS\_7和SYS\_9相连，那么SYS\_1中节点可能指向本系统中节点，也可能指向SYS\_7或者SYS\_9中节点。但是不会指向其他不与SYS\_1相连的系统。



## 评审标准

# 赛题解析

## 评审标准

评审要点	满分分值	实际得分
测试集 F1 值： $F1 = 2 \cdot \text{准确率} \cdot \text{召回率} / (\text{准确率} + \text{召回率})$ ，其中准确率=准确定位根因的组数/定位根因的组数，召回率=准确定位根因的组数/实际存在根因的组数。	50	
系统功能：要求能可视化展现根因信息及根因关联的局部拓扑图；要求系统运行稳定、可靠、流畅；系统设计：要求目录结构设计合理；功能模块划分合理。代码要求整洁，易读；符合程序编码规范，命名规范；注释清楚。	30	
文档：要求提供《算法原理说明文档》，《系统详细设计报告》，《测试报告》等。要求提供的文档要完整（缺少会扣分）。文档结构要合理，表达要清晰。	20	
满分总计	100	