Individual meeting

Advisor: Wen-Chih Peng
Student: Zheng-Ming Lin
Date: 2022/11/16

TSMC JDP

Metric, infra, and log data

⬇

Data preprocessing and log clustering

⬇

Anomaly detection (Metric and infra data)

⬇

Combine three types of data

⬇

Feature selection

⬇

Causal discovery (TCDF)
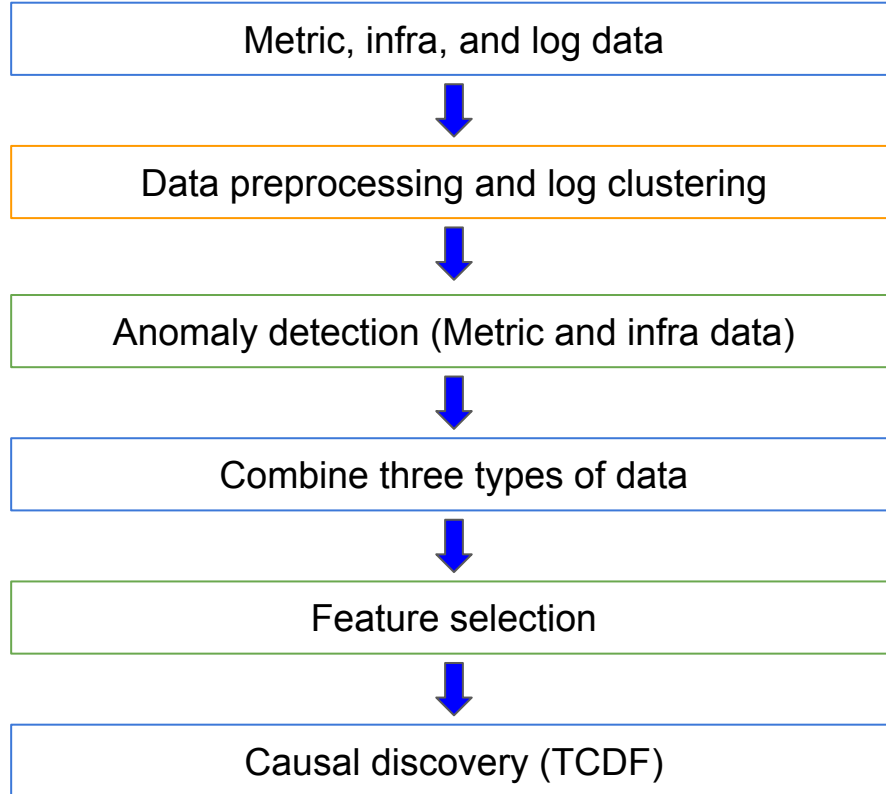
TSMC JDP

- Log clustering
  - TSMC requires the KPI for the latest result
    - Here is an example
      - cluster159, 160, 161, 162, 163, 164, 165, 166, 167 were grouped together in the original data
      - But in the latest result, they are split into three subgroups
        - cluster159, 160
        - cluster161, 162
        - cluster164, 165, 166, 167
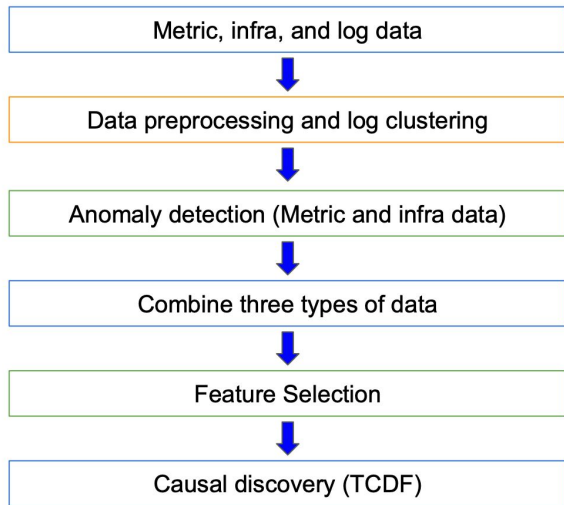   - In my view, I think that the accuracy should be 8/9

TSMC JDP

- Log clustering
  - My thought

$$kpi = \sum_{i=0}^{\# \, of \, subgroup} len(group \, i) \frac{len(group \, i)}{len(the \, longest \, group)} \, if \, len(group \, i) > 1$$

- Case: cluster159, 160, 161, 162, 163, 164, 165, 166, 167
  - cluster159, 160
  - cluster161, 162
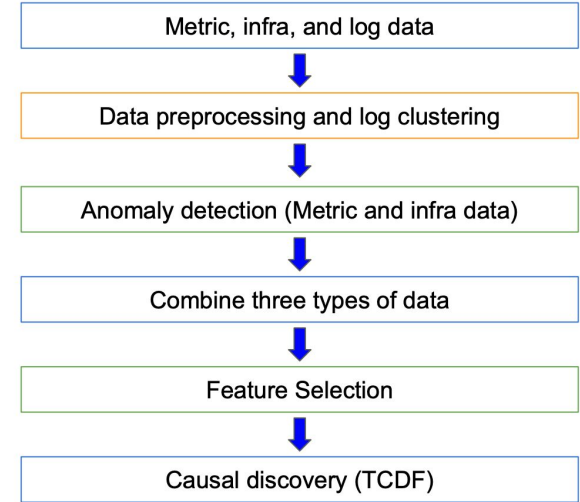 - cluster164, 165, 166, 167

TSMC JDP

- Anomaly detection
  - At first, I want to try GDN for anomaly detection
  - However, the feature in metric and infra data is too much
  - It will consume too much time for learning the graph structure
  - So, I am going to try univariate anomaly detection
    - DEPTS: Deep Expansion Learning for Periodic Time Series Forecasting
    - ICLR 2022, citations = 3

Metric, infra, and log data
Data preprocessing and log clustering
Anomaly detection (Metric and infra data)
Combine three types of data
Feature Selection
Causal discovery (TCDF)

TSMC JDP

- Feature selection
  - At first, I don't know that the data contains the trigger point
  - So now I will use the trigger point to count the pearson relation for feature selection
    - Choose top-k for important feature

Metric, infra, and log data
⬇
Data preprocessing and log clustering
⬇
Anomaly detection (Metric and infra data)
⬇
Combine three types of data
⬇
Feature Selection
⬇
Causal discovery (TCDF)

TSMC JDP

- Causal discovery
  - TCDF utilizes the attention mechanism to predict whether each two time series are similar
  - If I use feature selection, the features I choose will be similar as the trigger point
  - So each time series will be similar and make them be relative each others
  - TSMC uses the same method to detect the root cause

```
Metric, infra, and log data
        ↓
Data preprocessing and log clustering
        ↓
Anomaly detection (Metric and infra data)
        ↓
Combine three types of data
        ↓
Feature Selection
        ↓
Causal discovery (TCDF)
```