Title

Causal Alignment Based Fault Root Causes Localization for Wireless Network

Authors: Yuequn Liu, Wenhui Zhu, Jie Qiao, Zhiyi Huang, Yu Xiang, Xuanzhi Chen, Wei Chen, Ruichu Cai

Contents

Part I

Background

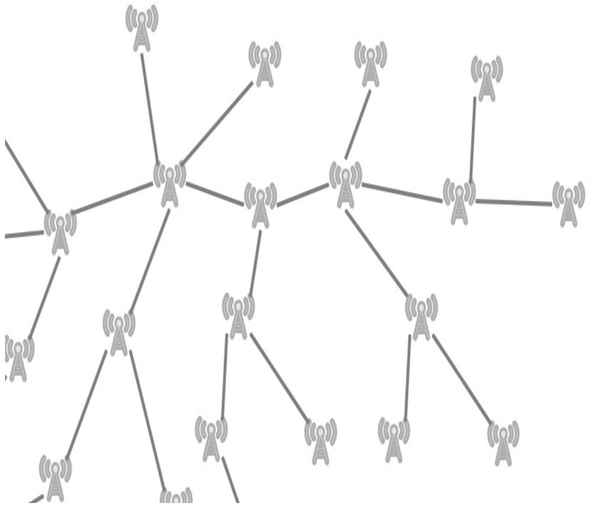
Background

Localizing the root cause of network faults is crucial to operation and maintenance.

Resorting to data analysis and machine learning is promising but remains difficult.

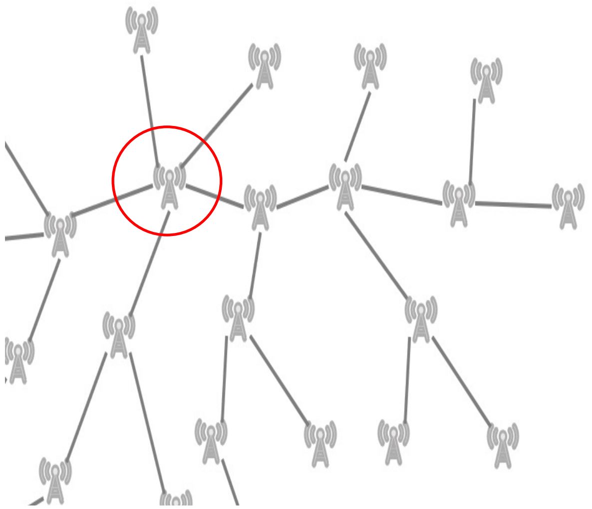
Design a root cause location model to quickly locate faults.

- Learn root causes from data
- Suitable for different domain dataset
- Help AI operation and maintenance



Real-world telecommunication dataset

Offline Model Training



Root causes prediction

Part II

Methodology

Why CARCL Is Needed

- Though supervised methods have shown promising results in training samples, most of the existing approaches assume that the training and the testing samples are independent and identical distributed.
- Such an i.i.d assumption usually does not hold due to network faults that may occur in different devices across different domains (well known as the distribution shift).
- Thus, it is necessary to align distributions between the training and test data set.

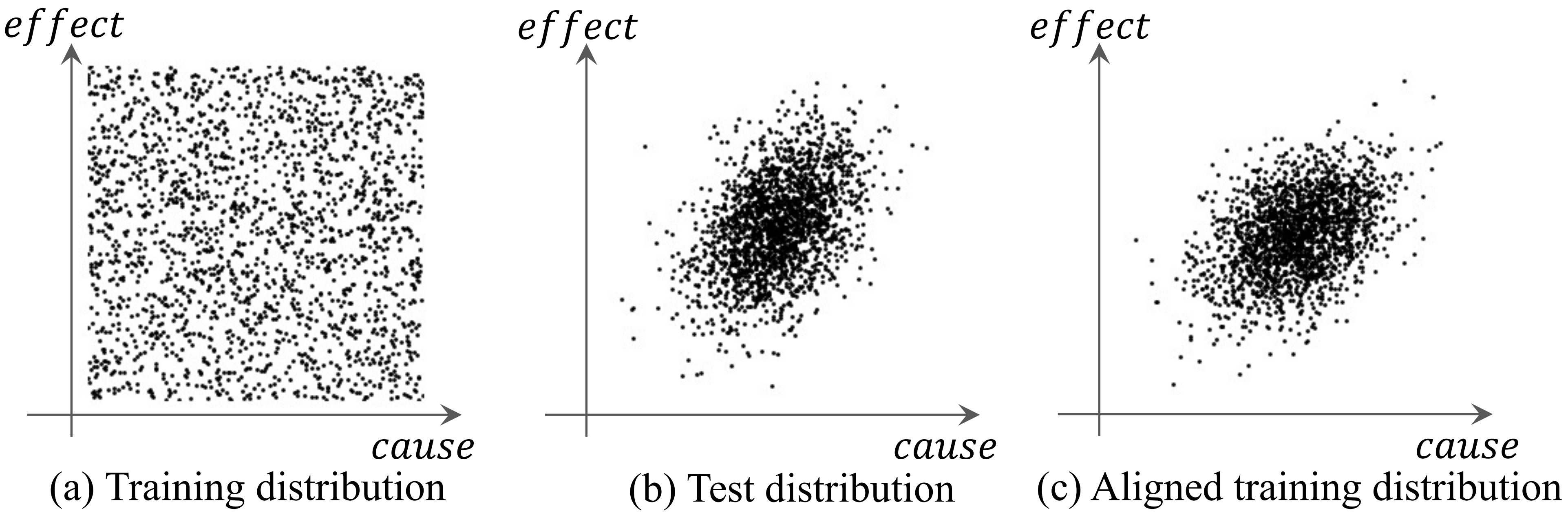
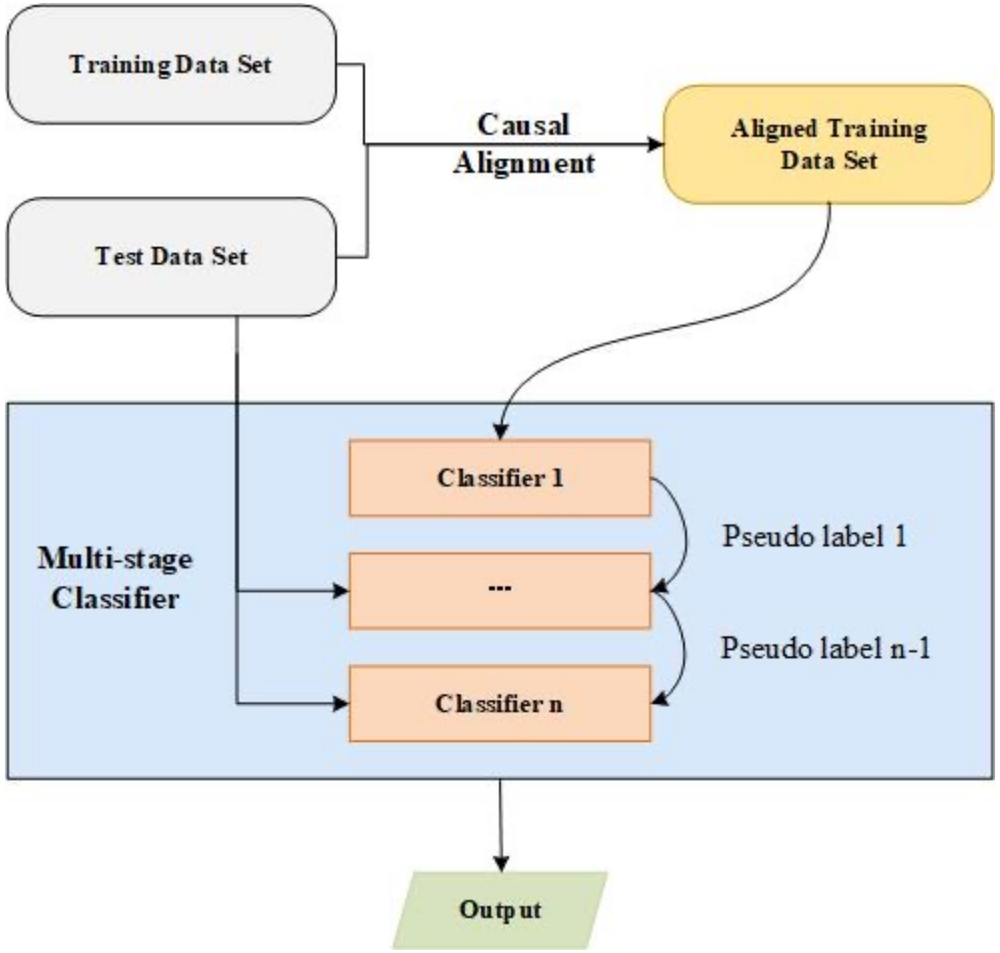


Figure: Illustrative example of the causal alignment.

Introduction of CARCL

Causal Alignment based Root Cause Localization (CARCL) framework:



Causal Alignment

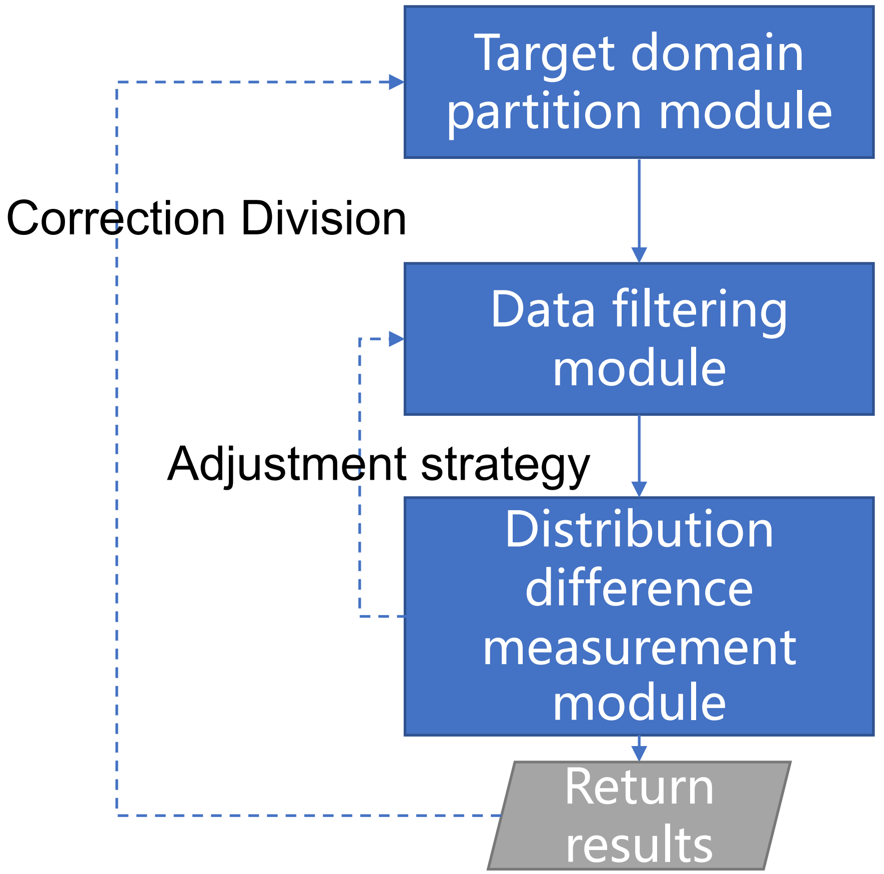
Align the distributions locally for each causal module but not globally on the complete variable set.

Multi-stage Classifier

Determine the root causes with the help of predicted pseudo labels.

Solution Introduction - CARCL

- **Causal alignment module:**
  - **Measure the distribution difference between the training set and the test set**
  - **Filter the training set data so that the causal mechanism reflected by the distribution of the remaining data is similar to the test set data**



use the Bayesian network-based method to classify the test data set into fault and fault-free parts.

Including but not limited to the maximum flow problem solving algorithm represented by Ford Fulkerson labeling method

1. Take Kullback-Leibler divergence as the difference measurement standard

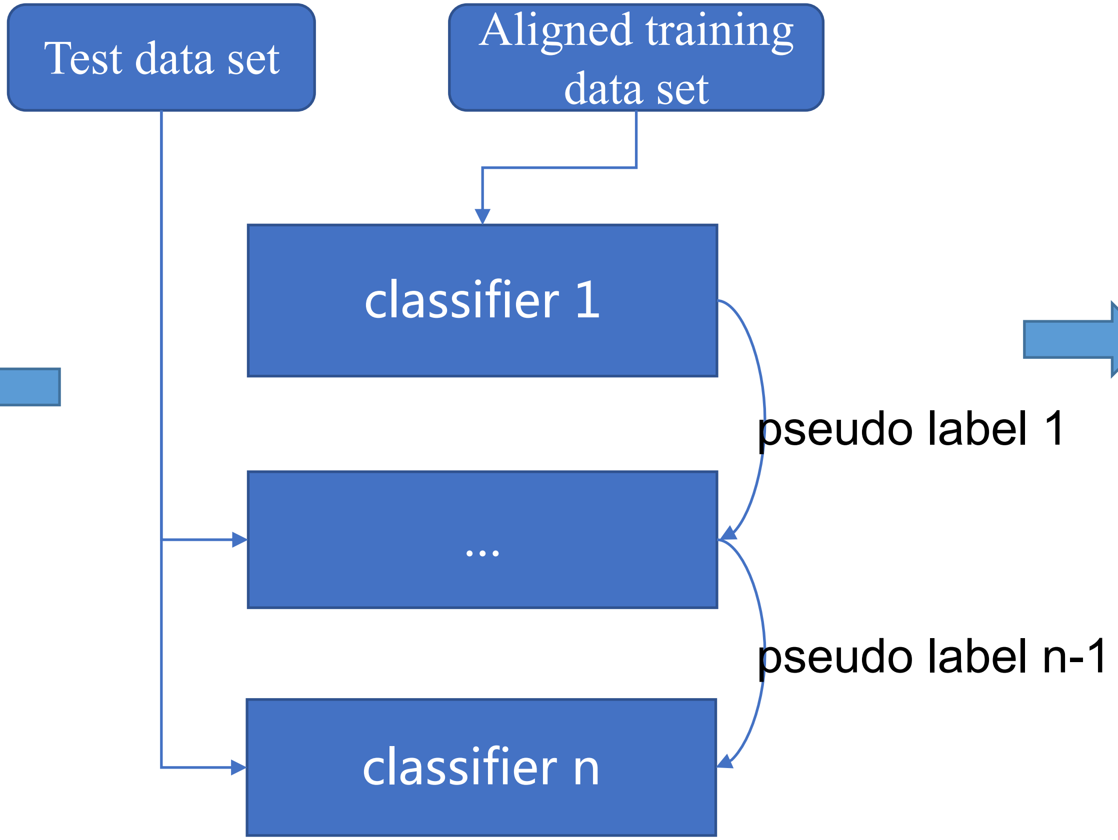
$$\min_F KL\left(\frac{P_S(\text{effect}|\text{cause},F)}{P_S(\text{effect},F)/P_S(\text{cause},F)} \middle\| \frac{P_T(\text{effect}|\text{cause})}{P_T(\text{effect})/P_T(\text{cause})}\right), \quad (1)$$

2. Take maximum mean dispersion as the difference measurement standard
3. Take total variation as the difference measurement standard

Solution Introduction - CARCL

- **Multi-stage classification module:**
  - **multi-layer classifier module based on pseudo label is used to improve the classification accuracy of the model.**

1. The first layer classifier: the input is the output of the causal alignment module and the original label. The output is the prediction label on the test set as the first layer of pseudo label.
2. Start with the second layer: test data and the previous layer of pseudo tags as input. The output is the prediction label of the test data.



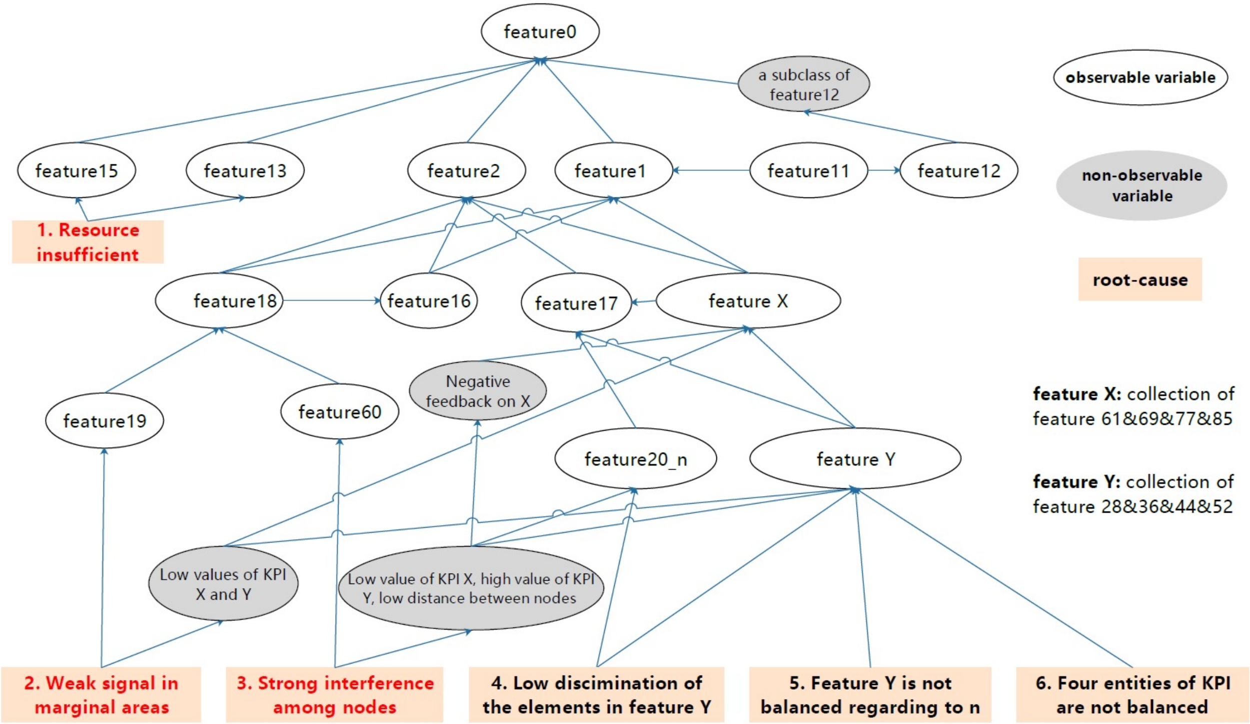
Including but not limited to:
1. XGBoost, LightGBM, CatBoost and other classifiers based on decision tree
2. Classifier with k-nearest neighbor as reference data
3. Classifier based on Bayesian network
4. Mixed use of multiple classifiers

Part III

Practice

Data set

- The experimental data set is a 5G wireless network data set published by ICASSP-SPGC-2022 communication network intelligent operation and maintenance competition. The address is https://www.aiops.sribd.cn/home/data 。 The purpose of the competition is to deepen the understanding, research and application of such complex practical network fault root cause inference. This public data set consists of three parts:
  - **Causal relationship graph**. In this data set, the sponsor provided a causal relationship graph (desensitized) drawn by experts as a priori.
  - **Training data set**. A total of 2984 samples are included. Each sample is a time segment (with variable length) from different 5G road test scenes, which contains the information of 23 observable characteristic variables changing with time in this time segment.
  - **Test data set**, the variable characteristics correspond to the training data set one by one, including 600 samples.



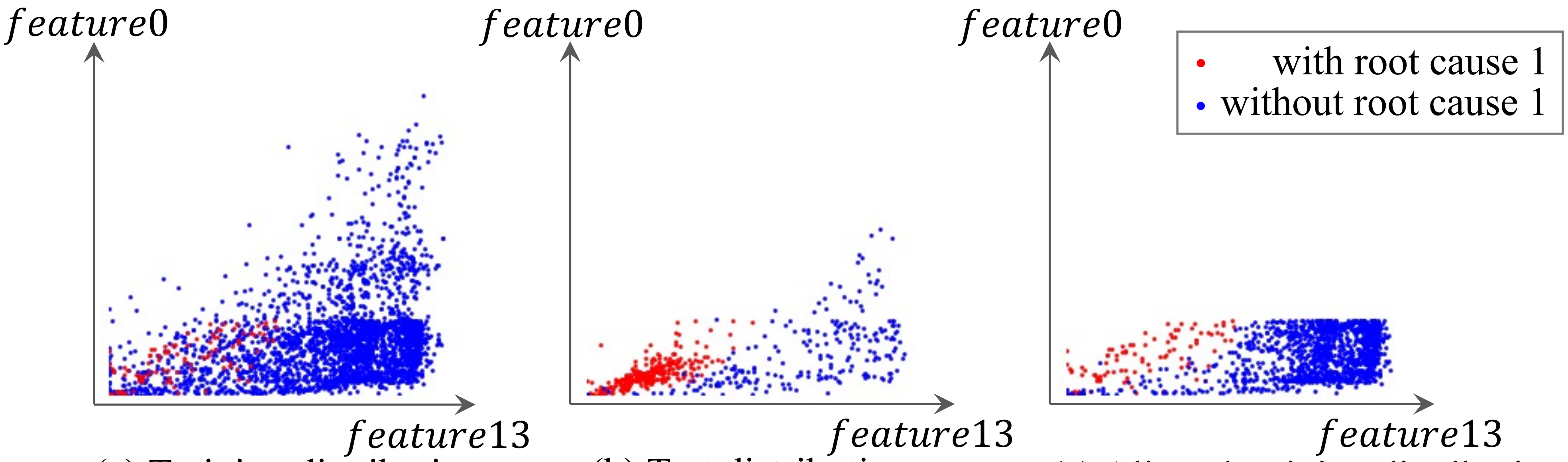
Experiment

- We compare with the following existing baselines:
  - The k-nearest neighbors (k-NN) algorithm.
  - A Bayesian network probabilistic graphical model.
  - CatBoost algorithm.
- We compare with the following ablation studies:
  - CARCL-NCA is a variant of CARCL by removing the causal alignment.
  - CARCL-NM is another variant of CARCL by replacing the multi-stage classifier with a single LightGBM.
  - CARCL-L is the last variant. It removes causal alignment and multi-stage classifier, equivalent to the original LightGBM.

Algorithms	ALL	Ch(R_i)	Ch(R)
k-NN	0.566111	0.472222	0.566111
CatBoost	0.644722	0.642500	0.625277
Naive-Bayes	-0.018333	0.409999	-0.001111
CARCL-NCA	0.662499	0.675555	0.651944
CARCL-NM	0.901388	0.835278	0.910555
CARCL-L	0.627222	0.643055	0.607222
CARCL	**0.922778**	**0.869167**	**0.931945**

Case study

- To show the effectiveness of our approach, we provide a case study with the root cause 1. In detail, we analyzed the cases with feature 13 and feature 0 that is the descendent of the root cause 1, in which feature 13 is the cause of feature 0.
- There is a significant difference between the distribution of the training data set and the test data set without the distribution alignment. After distribution alignment, the distribution of the new training data set and the test data set is similar.



(a) Training distribution (b) Test distribution (c) Aligned training distribution

Figure: Case study on root cause 1, in which the label in test distribution are obtained by the prediction of CARCL.

Part IV

Conclusion

Conclusion

- We propose a root causes location framework based on causal alignment.
- Differ from previous methods, we consider the different distribution of data in training data set and test data set, and conduct the causal alignment processing according to the causal mechanism.
- The excellent performance of the proposed method provides an effective solution for fault root causes localization.
- Furthermore, it guides the future work of root causes localization methods based on the causal alignment.

Final

Thanks for listening

Authors: Yuequn Liu, Wenhui Zhu, Jie Qiao, Zhiyi Huang, Yu Xiang, Xuanzhi Chen, Wei Chen, Ruichu Cai