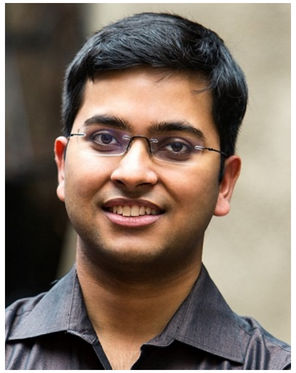
ICLR 25

Robust Root Cause Diagnosis using In-Distribution Interventions

Lokesh Nagalapatti

Ashutosh Srivatsava

Sunita Sarawagi



Amit Sharma

Diagnosing Anomalies in Microservices/Industry

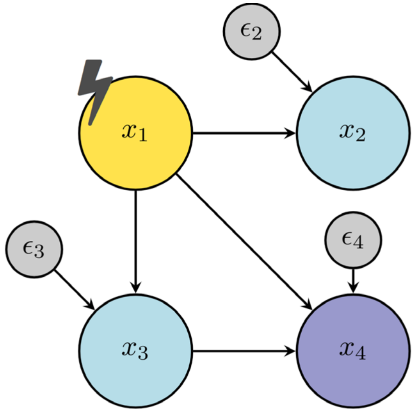
- Given that a fault occurred at a particular component of an industry, goal is to diagnose its root cause.

- The goal of RCD is:
  1. Predict the root cause node that triggered the anomaly at target node.
  2. Propose remediation action at the root cause node to fix the anomaly at the target.

- A Causal Inference problem.

Root Cause diagnosis

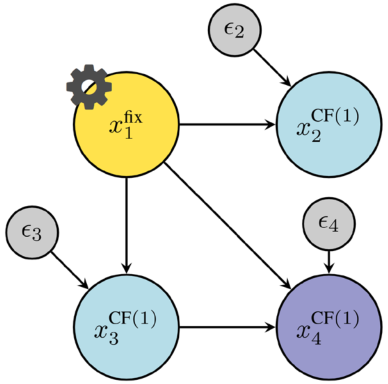
X_1 suffered an abnormal intervention



(a) Root Cause sample

Fault occurred at x_4

Suppose we apply a fix at X_1 and set it to its normal value x_1^{fix}



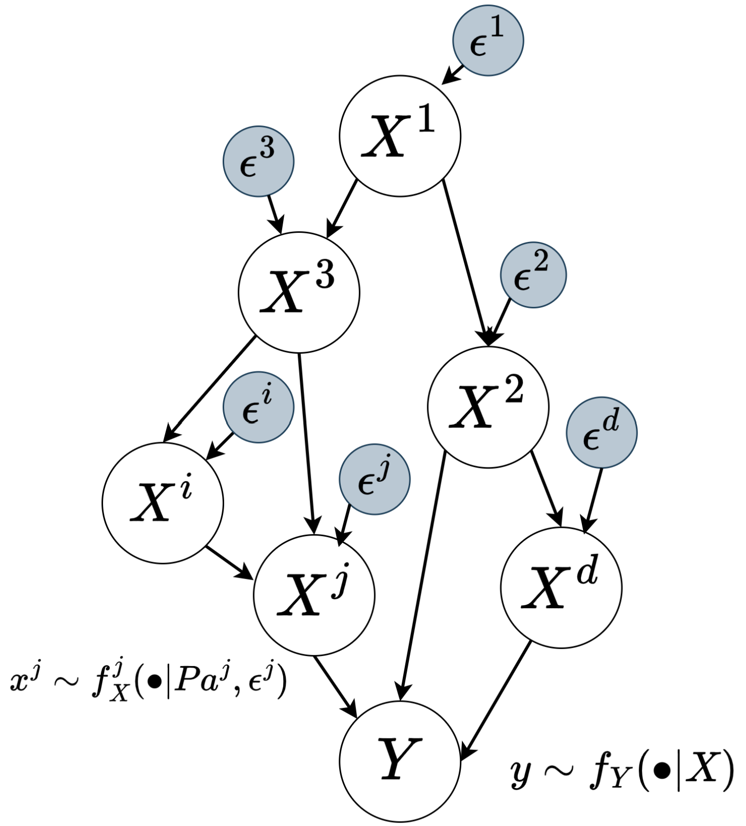
(b) True Counterfactual

As a consequence of the fix, the corresponding CF X_4 should come back to normal value.

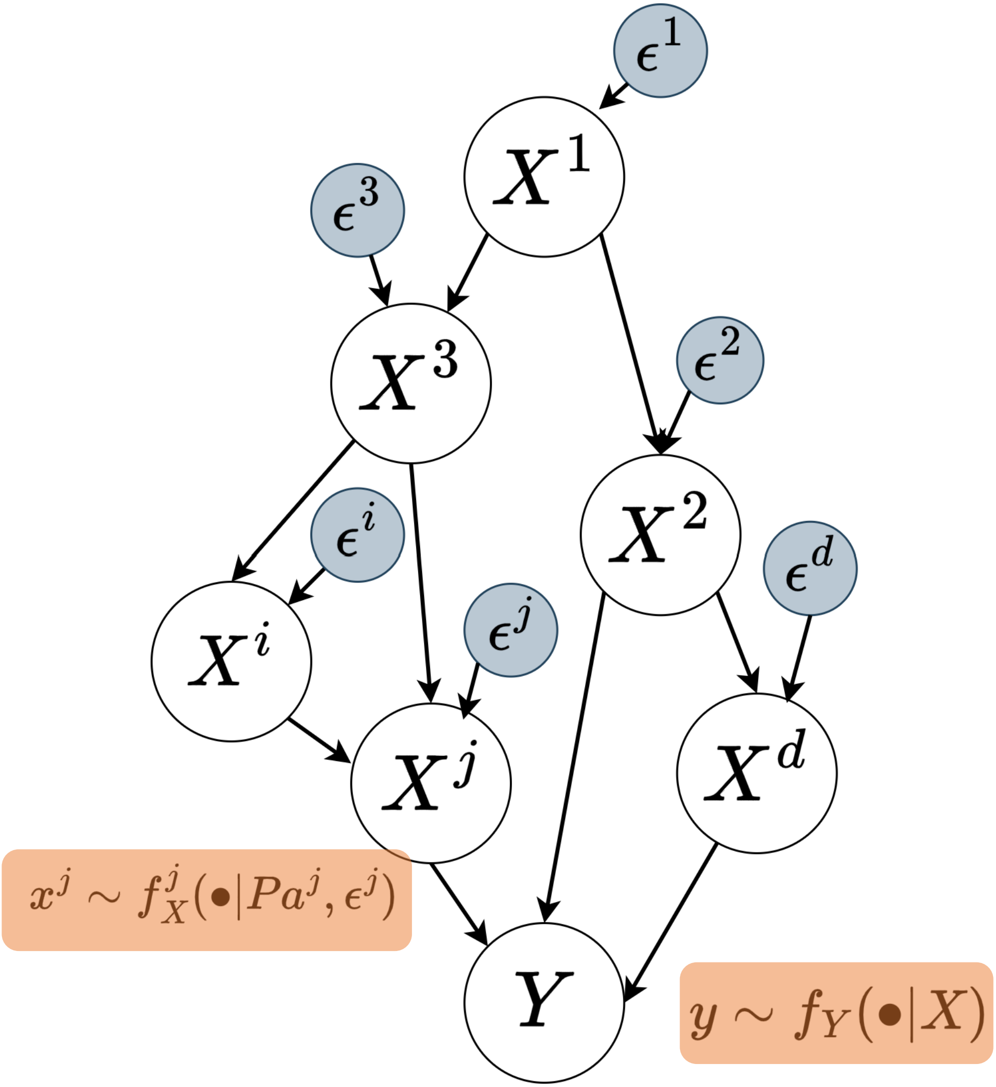
Training Dataset

- We are given a logged dataset of values of nodes observed through time.

- And the Causal Graph connecting the nodes.



Advantage of knowing the causal graph



Further if the node functions are additive in ϵ

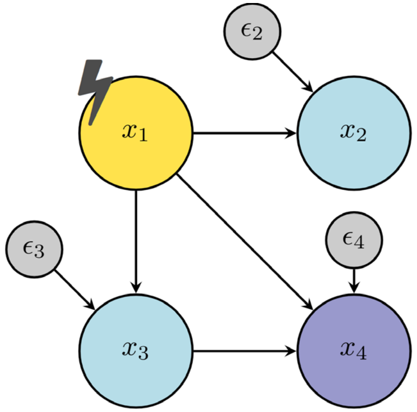
$$x^j = f_X^j(\mathrm{pa}^j) + \epsilon^j$$

Then Abduction is possible:

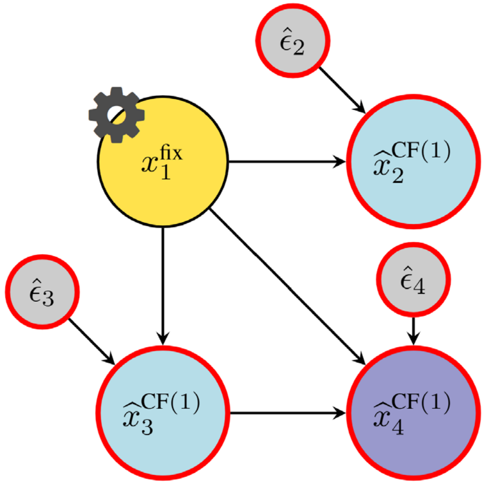
$$\epsilon^j = x^j - f_X^j(\mathrm{pa}^j)$$

Each of these functions can be learned using observational datasets upto good accuracies because each one involves fitting a very low dim regression problem

Prior works Estimate Counterfactuals



(a) Root Cause sample

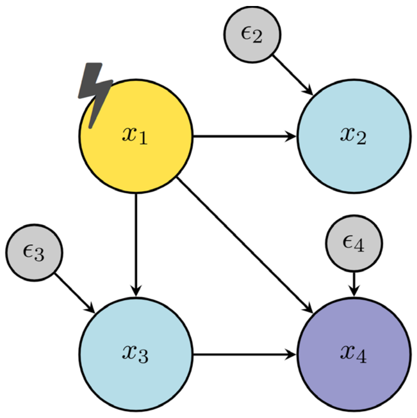


(c) Est. Counterfactual

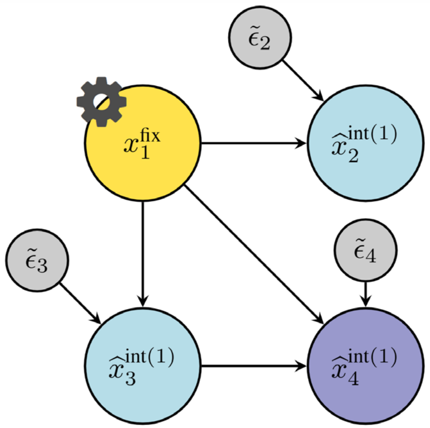
Abduction involves estimation of ϵ_4 at an abnormal x_1

1. Abduction: $\widehat{\epsilon_4} = x^j - f_X^j(x_1)$
2. Action: Set $x_1 = x_1^{\text{fix}}$
3. Prediction: $\widehat{x_4^{\text{CF}}} = f_4\left(\widehat{x_1^{\text{fix}}}\right) + \widehat{\epsilon_4}$

Our Approach: In Distribution Intervention



(a) Root Cause sample



(d) Est. Intervention

1. Action: Set $x_1 = x_1^{\text{fix}}$
2. Prediction: $\widehat{x_4^{\text{int}}} = f_4\left(\widehat{x_1^{\text{fix}}}\right) + \widetilde{\epsilon_4}$
   where $\widetilde{\epsilon_4}$ is a sampled value

Consequence: We always evaluate $\widehat{f_4}$ at in-distribution values

Results on Petshop Dataset

Recall@		Low		High		Temporal	
		k=1	k=3	k=1	k=3	k=1	k=3
Correlation	Random Walk (Yu et al., 2021)	0.00	0.10	0.00	0.20	0.00	0.33
	Ranked Correlation (Hardt et al., 2023)	0.40	0.60	0.70	0.90	0.50	0.67
	ϵ-Diagnosis (Shan et al., 2019)	0.00	0.00	0.00	0.00	0.17	0.17
Causal Anomaly	Circa (Li et al., 2022)	0.60	0.80	0.60	1.00	0.67	1.00
	Traversal (Chen et al., 2014)	0.80	0.80	0.90	0.90	1.00	1.00
	Smooth Traversal (Okati et al., 2024)	0.40	0.60	0.00	0.60	0.50	1.00
Causal Fix	HRCD (Ikram et al., 2022)	0.07	0.21	0.00	0.07	0.25	0.75
	TOCA (Okati et al., 2024)	0.40	0.40	0.20	0.20	0.00	0.00
	CF Attribution (Budhathoki et al., 2022b)	0.40	0.60	0.40	0.70	0.00	0.50
	IDI (Ours)	0.90	0.90	0.90	0.90	1.00	1.00

Thank you!