Dual-Latent Generative Causal Structure Learning with Causal Annealing

Objective : To learn Causal structure in the presence of unobserved confounders with non-linear causal relationships without previously known causal graphs.

NEURAL INFORMATION PROCESSING SYSTEMS

Key Contributions

VAE with dual latent spaces ($Z_{directed}, Z_{bidirected}$) Capturing directed, and bidirected dependencies using ADMG via trainable adjacency matrices A_D, A_B.

Causally aware objective: A causality-aware loss that enforces acyclicity, bow-free structure, and sparsity–entropy balance for meaningful, interpretable edges.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{reconstruction}} + \lambda_{\text{KL}}(\mathcal{L}_{\text{KL\_directed}} + \mathcal{L}_{\text{KL\_bidirected}}) + \lambda_{\text{causal}}\mathcal{L}_{\text{Causal\_ADMG}},$$

Causal annealing: Introduced a novel training strategy that delays causal constraints via a transition epoch, named it as causal transition epoch to improve stable structure learning.

NeurIPS 2025
Workshop CauScien

Dual-Latent Generative Causal Structure Learning with Causal Annealing - Method

NeurIPS 2025
Workshop CauScien

Model Components:

- **Input:** data matrix $X \in \mathbb{R}^{n \times d}$
- **Output:** reconstructed $\hat{X}$, directed A_D, bidirected A_B
- **Initialize:** encoder/decoder parameters, adjacency matrices W_1, W_2

For each epoch $e = 1$ to E:

- Encode X into:

$$(\mu_D, \log \sigma_D^2), \qquad (\mu_B, \log \sigma_B^2)$$

- Compute structure-aware means:

$$\mu_{DA_D} = \mu_D W_1, \qquad \mu_{BA_B} = \mu_B W_2$$

- Structure-aware latents:

$$z_D = \mu_{DA_D} + \epsilon_D \odot \exp(0.5 \log \sigma_D^2), \qquad z_B = \mu_{BA_B} + \epsilon_B \odot \exp(0.5 \log \sigma_B^2)$$

  where $\epsilon_D, \epsilon_B \sim \mathcal{N}(0, I)$

- Form $z = [z_D, z_B]$ and decode:

$$\hat{X} = \mathrm{Decoder}(z)$$

- Estimate adjacency matrices:

$$A_D = f(z_D), \qquad A_B = f(z_B)$$

- Minimize total loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rec}} + \mathrm{KL}_{\text{directed}} + \mathrm{KL}_{\text{bidirected}} + \lambda_{\text{causal}}\, \mathcal{L}_{\text{Causal-ADMG}}$$

Dual-Latent Generative Causal Structure Learning with Causal Annealing: Outco

Causal Structure Learning

Table 1: Performance comparison: F1 scores (F1D for directed, F1B for bidirected edges) on FC, ER(4,6,4), and ER(12,50,10)

Method	FC		ER(4,6,4)		ER(12,50,10)	
	F1D	F1B	F1D	F1B	F1D	F1B
FCI	0.00	0.75	0.50	0.40	0.25	0.33
CAM-UV	0.80	0.67	0.30	0.25	0.38	0.36
RCD	0.00	0.54	0.35	0.35	0.45	0.20
DCD	0.00	0.67	0.25	0.20	0.32	0.18
N-DAG-G	0.50	0.00	0.60	0.00	0.55	0.00
N-ADMG-G	0.49	**0.99**	0.75	0.60	0.60	0.38
N-BF-ADMG-G	0.64	0.93	0.78	0.80	**0.60**	0.40
Proposed (G-ADMG-CL)	**1.0**	0.50	**0.92**	**0.89**	0.51	**0.45**

Causal Inference

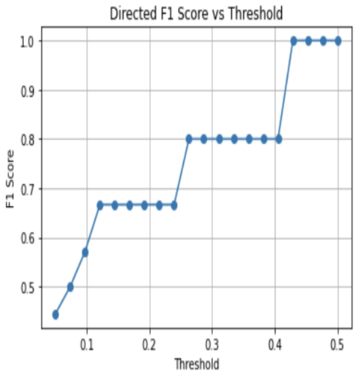
Table 2: Causal inference results using IHDP dataset

Method	RMSE-ATE
FCI	0.13
CAM-UV	0.15
RCD	0.14
DCD	0.16
N-DAG-G	0.12
N-BF-ADMG-G	0.10
Proposed (G-ADMG-CL+P)	0.031

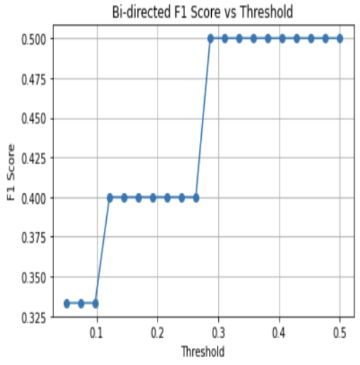
Impact of Causal Annealing

Table 3: Impact of causal annealing on structure recovery (F1).

Method	FC		ER(4,6,4)	
	F1D	F1B	F1D	F1B
G-ADMG-CL (with annealing)	1.00	0.50	0.92	0.89
G-ADMG-CL (no annealing)	0.50	0.50	0.75	0.80

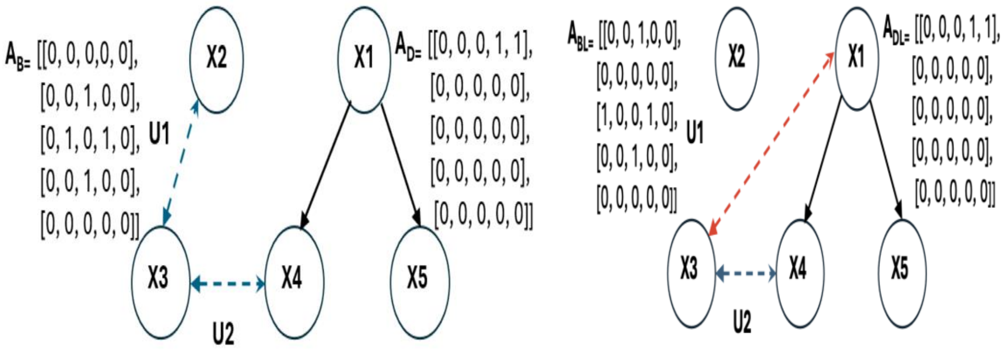


(a) Obtained best F1 score for learned for A_D.

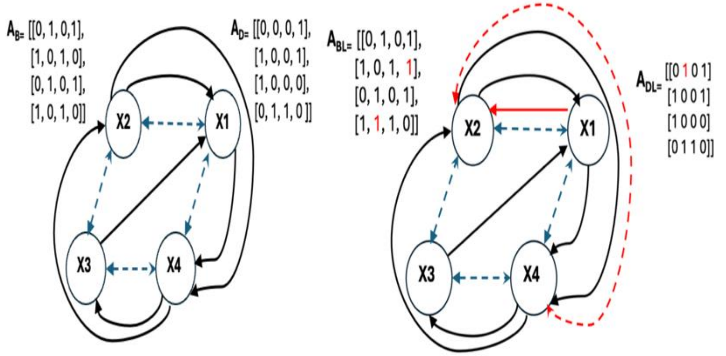


(b) Obtained best F1 score for A_B.

FC data



FC (Fork Collider) data



ER 4, 6, 4

ADMG: (Acyclic Directed Mixed Graph)

Dual-Latent Generative Causal Structure Learning with Causal Annealing: Conclusion

Dual-latent VAE with trainable **directed (W_1)** and **bidirected (W_2)** adjacency matrices
Supports **ADMG-based structure learning**
Learns **directed causal effects** + **latent confounded bidirected influences**
Capability **absent in prior VAE-based causal models**
Causal mixed-graph loss for unified directed–bidirected learning
Causal annealing strategy for stable and progressive structure enforcement
Ablation studies:

- ❑ Mixed-graph loss is essential
- ❑ Causal annealing improves robustness and accuracy

Performance: competitive or improved vs. established baselines
Real-data validity: learned structure improves causal inference (e.g., IHDP)
Future work:

- ❑ Exploring annealing schedules (CTE variants)
- ❑ Hyperparameter sensitivity analysis
- ❑ Compare thresholding strategies
- ❑ Analyze coexistence of directed & bidirected edges in complex graphs

Soma Bandyopadhyay
soma.bandyopadhyay@tcs.com;
somab@iitkgp.ac.in

Sudeshna Sarkar
sudeshna@cse.iitkgp.ac.in