Causal Discovery from Incomplete Data
A Deep Learning Approach

Yuhao Wang⋆ Vlado Menkovski⋆ Hao Wang† Xin Du⋆ Mykola Pechenizkiy⋆

⋆Eindhoven University of Technology †Massachusetts Institute of Technology.

{y.wang9, v.menkovski, x.du, m.pechenizkiy}@tue.nl, hoguewang@gmail.com

AAAI StarAI 2020, New York

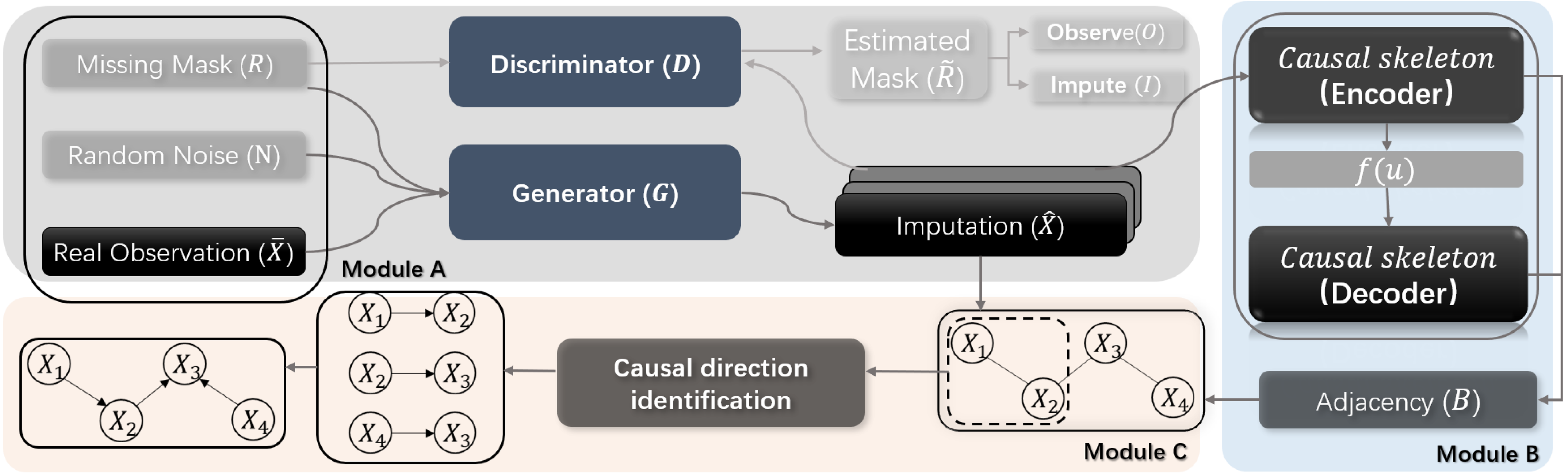
Introduction

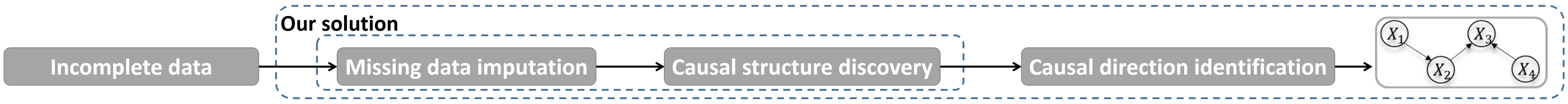
Causal Discovery

- Enable accurate decision making;
- Robust uncertainty inference;
- Reliable fault diagnose;
- Efficient redundancy elimination;
- ...

Current Problem

- Missing data are ubiquitous;
- Existing algorithms on partially observed data may lead to the incorrect inference.





System Framework

In each iteration, G and D take the incomplete data as input and impute the missing values to form $\hat{X}$. The causal structure B is involved as parameters of both SE and SD. We encode $\hat{X}$ into a latent code $f(U)$ through SE, and decode $f(U)$ into $\tilde{X}$ with SD.

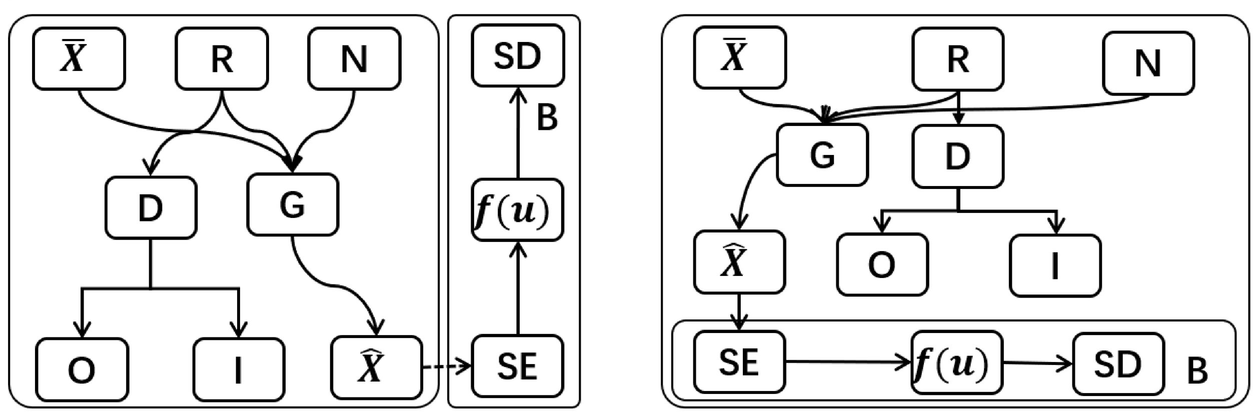


Figure 1:Left: *Imputation first, then structure discovery;*. Right: *Simultaneous imputation and structure learning* ***(Ours)***.

Our Contributions

- A deep learning framework of Imputed Causal Learning (ICL), for iterative missing data imputation and causal structure discovery, producing both imputed data and causal skeletons.
- Leverage the extra asymmetry cause-effect information to enumerate the causal directions and uncover the underlying causal graph.
- State-of-the-art performance on both synthetic and publicly-used real data.

Research Methodology

The General Procedure

- Causal skeleton learning returns a global view of how variables are dependent on each other;
- Direction identification provides a more accurate local view between the matched variable pairs.

Detailed Implementation

Initialize: $R \in \{0,1\}^{n\times d}$, $\bar{X} \in \mathbb{R}^{n\times d}$, $\tilde{\mathcal{G}} \in \mathbb{R}^{d\times d}$, $N = P_n \sim \mathcal{N}(\mu, \sigma^2)$.

Missing Data Imputation: Generate full observations and yield an optimistic estimation from $\bar{X}$ by imputation.

- Missing entries: $\bar{X} = G(R, \bar{X}, (1-R)\odot N)$.
- Imputation: $\hat{X} = R\odot \bar{X} + (1-R)\odot \bar{X}$.

Causal Skeleton Learning: Finding a $\tilde{\mathcal{G}}$ by extracting B during optimization:

$$\tilde{\mathcal{G}} = g(argmin_{\mathcal{G}\in\mathbb{R}^{d\times d}}\mathcal{S}_{\mathcal{D}}(\mathcal{G})), \qquad f(U) = (I - B^T)MLP(\hat{X}, \mathbf{W}_1);$$

$$s.t.\ \ h(\mathcal{G}) = tr[(I+\alpha B\circ B)^d] - d = 0, \qquad \tilde{X} = MLP((I-B^T)^{-1}(f(U), \mathbf{W}_2)),$$

Joint Training: The loss function is formed by the imputation loss and structure learning loss.

$$\min_{\Phi} f(\Phi) = L_i(G, D); \quad \min_{B,\Theta} f(B,\Theta) = -L_e,\ s.t.\ h(B) = 0.$$

Causal Direction Identification: Then orient the edges of $\tilde{\mathcal{G}}$ in a pair-wise way, consequently uncovering the final causal DAG $\mathcal{G}$.

Experimental Results

(1) Synthetic nonlinear data: with MCAR and MAR missingness {10%, 30%, 50%} mechanism;

$$1:\ \ x = 2sin(B^T(x+0.5\cdot 1)) + B^T(x+0.5\cdot 1) + u,$$
$$2:\ \ x = \sqrt{x(B^T(x^2+0.5\cdot 1))} + u.$$

Table 1:Performance comparison (mean and standard deviation) using Structural Hamming Distance (SHD), lower is better.

		30 Var MCAR (Nonlinear 1) (Ideal SHD=7)			50 Var MAR (Nonlinear 2) (Ideal SHD=17)		
		10%	30%	50%	10%	30%	50%
GES	LD-GES	106.0 ± 14.3	109.1 ± 16.9	145.4 ± 13.4	227.2 ± 22.5	224.1 ± 28.6	225.6 ± 28.4
	GAN-GES	107.8 ± 12.2	106.9 ± 14.8	133.1 ± 15.9	228.5 ± 21.3	224.2 ± 25.6	225.6 ± 27.8
	MF-GES	109.3 ± 13.8	108.1 ± 14.8	136.9 ± 16.1	230.6 ± 21.6	224.1 ± 28.5	223.9 ± 26.9
	MC-GES	109.3 ± 13.8	109.1 ± 15.2	132.3 ± 16.2	230.6 ± 21.6	225.4 ± 28.0	225.4 ± 27.2
RFCI	LD-RFCI	22.2 ± 5.2	26.4 ± 8.3	43.3 ± 7.4	44.1 ± 8.3	49.7 ± 8.8	68.2 ± 10.1
	GAN-RFCI	38.6 ± 5.1	39.9 ± 8.3	42.0 ± 7.3	52.3 ± 8.3	66.6 ± 8.7	69.2 ± 10.1
	MF-RFCI	38.9 ± 5.0	39.9 ± 8.3	44.6 ± 7.0	51.0 ± 8.4	66.7 ± 8.8	68.8 ± 9.7
	MC-RFCI	38.8 ± 4.8	39.8 ± 8.3	42.7 ± 7.1	51.7 ± 8.2	66.5 ± 9.1	69.0 ± 10.1
LiNGAM	LD-LiNGAM	22.0 ± 8.4	25.3 ± 10.3	32.6 ± 10.4	41.3 ± 15.2	50.4 ± 17.6	53.9 ± 7.1
	GAN-LiNGAM	20.9 ± 8.4	23.1 ± 10.3	37.0 ± 10.4	43.0 ± 15.2	53.2 ± 17.6	47.6 ± 7.1
	MF-LiNGAM	23.1 ± 7.8	23.5 ± 8.3	37.6 ± 11.2	52.0 ± 16.9	48.2 ± 18.1	52.4 ± 13.6
	MC-LiNGAM	21.5 ± 8.9	29.1 ± 12.3	37.3 ± 12.0	43.6 ± 13.1	51.9 ± 14.0	52.6 ± 11.2
PC	MVPC	40.63 ± 7.9	43.8 ± 9.8	45.7 ± 8.2	72.4 ± 9.7	69.8 ± 9.8	67.3 ± 11.4
	LD-PC	26.2 ± 6.2	27.9 ± 7.6	35.0 ± 6.4	36.0 ± 7.7	38.5 ± 10.4	45.2 ± 8.1
	TD-PC	25.2 ± 5.0	26.6 ± 6.23	28.1 ± 5.1	36.4 ± 7.1	55.5 ± 7.1	52.3 ± 10.4
	GAN-PC	26.0 ± 6.2	26.1 ± 7.6	32.3 ± 6.4	34.2 ± 7.7	38.6 ± 10.4	41.6 ± 7.4
	MF-PC	26.4 ± 5.8	26.2 ± 7.9	33.3 ± 6.8	35.0 ± 8.0	35.3 ± 10.1	41.9 ± 7.0
	MC-PC	27.9 ± 5.9	26.8 ± 8.2	33.3 ± 7.2	34.7 ± 8.0	37.8 ± 10.9	42.2 ± 7.5
	CBR-PC	24.4 ± 4.0	26.6 ± 7.1	27.2 ± 5.2	47.2 ± 6.2	47.5 ± 8.7	46.3 ± 4.9
MMPC	LD-MMPC	22.6 ± 7.3	23.2 ± 7.5	30.7 ± 9.7	45.2 ± 11.4	44.5 ± 11.1	44.0 ± 7.0
	GAN-MMPC	22.0 ± 7.5	23.8 ± 7.2	27.0 ± 9.9	46.0 ± 11.1	48.5 ± 10.5	44.5 ± 6.5
	MF-MMPC	22.8 ± 7.3	25.0 ± 7.2	29.1 ± 9.6	46.3 ± 11.2	48.7 ± 11.2	44.5 ± 6.9
	MC-MMPC	22.4 ± 7.3	25.8 ± 7.2	29.4 ± 9.5	46.3 ± 11.1	48.6 ± 11.2	44.4 ± 7.1
DAG	LD-DAG	12.2 ± 6.2	13.6 ± 9.2	20.0 ± 10.4	30.2 ± 5.9	32.5 ± 4.5	37.9 ± 7.1
	GAN-DAG	11.0 ± 7.7	10.3 ± 6.8	14.4 ± 8.7	23.4 ± 5.5	27.7 ± 3.9	30.5 ± 4.2
	ICL (Ours)	**9.8 ± 3.9**	**7.4 ± 3.8**	**8.4 ± 4.9**	**19.0 ± 4.2**	**25.5 ± 3.8**	**27.3 ± 5.5**

(2) AutoMPG: City-cycle fuel consumption dataset with 398 instances and 7 attributes;
(3) Bioinformatics: A bioinformatics dataset, with 11 cell types and 7466 samples.

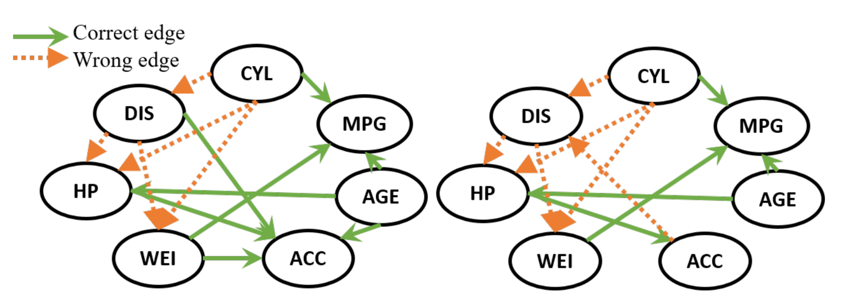


Figure 2:AutoMPG Dataset

Table 2:SHD with respect to the ground gruth

Method	ICL	LD-DAG	GAN-DAG	GES	RFCI	LiNGAM	PC	CBR-PC	MVPC
AutoMPG	**9**	12	11	16	14	14	12	16	14
Bioinformatics	**16**	21	17	18	20	19	20	20	18

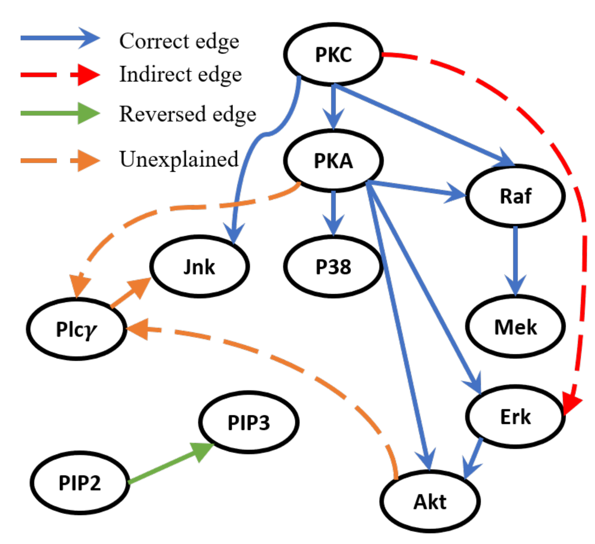


Figure 3:Bioinformatic Dataset

Conclusion and Future Work

In this work, we addressed the problem of incomplete data causal discovery:

- We proposed a deep learning model of ICL to handle this issue.
- Specifically, our ICL model contains a global view of iterative missing data imputation and causal skeleton discovery, and a local view of enumerating causal directions to uncover the underlying causal graph;
- We evaluated the effectiveness of our method on both synthetic and real data.
- For future work, we will generalize our method under more complex conditions such as the existence of confounders.