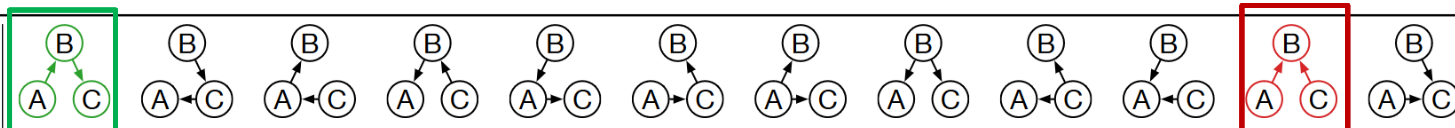
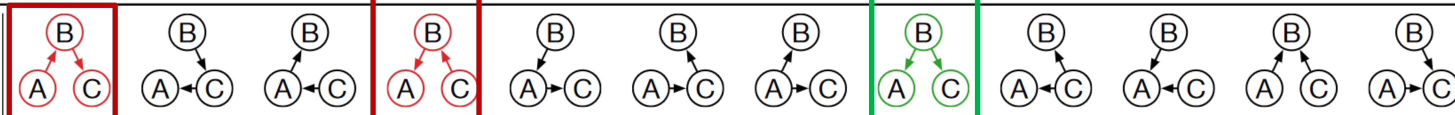
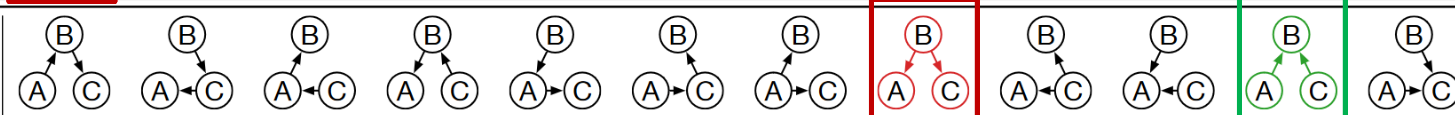
DARING研究背景与意义

东南大学 SOUTHEAST UNIVERSITY

Predicted Graph (Chain)												
Reconstruction Loss	6.00	6.97	6.17	6.17	6.96	6.33	6.16	6.80	6.33	7.00	5.65	7.00
Residuals Mutually Independent?	✓			✓				✓				
Predicted Graph (Fork)												
Reconstruction Loss	1.67	2.17	1.83	1.67	2.17	1.83	1.83	2.00	1.83	2.33	1.78	2.33
Residuals Mutually Independent?								✓				
Predicted Graph (Collider)												
Reconstruction Loss	1.89	2.00	2.22	1.89	2.00	2.22	2.22	1.67	2.22	1.83	2.11	1.83
Residuals Mutually Independent?											✓	

传统可微CD学到的loss最小的结果

ground truth

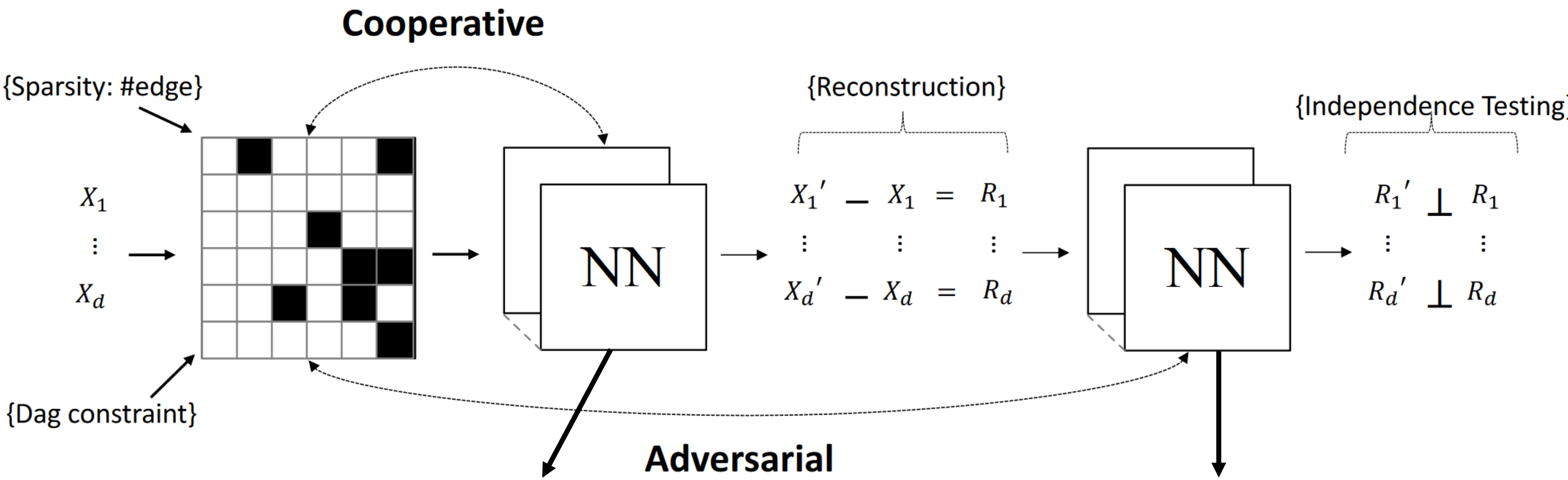
□ 数据生成机制：

- **链式：** $A=\epsilon_A(\sim\mathcal{N}(0,1)), B=A+\epsilon_B(\sim\mathcal{N}(0,4)), C=B/5+\epsilon_C(\sim\mathcal{N}(0,1))$
- **叉式：** $B=\epsilon_B(\sim\mathcal{U}(-2,2)), A=B/2+\epsilon_A(\sim\mathcal{U}(-1,1)), C=B/2+\epsilon_C(\sim\mathcal{U}(-1,1))$
- **对撞：** $A=\epsilon_A(\sim\mathcal{N}(0,1)), C=\epsilon_C(\sim\mathcal{N}(0,1)), B=A/3+C/3+\epsilon_B(\sim\mathcal{N}(0,1/9))$

传统可微因果发现过拟合噪声，导致残差之间存在依赖性，偏离真实因果图

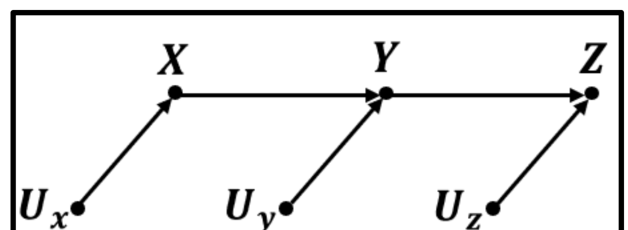
[He Y, Cui P, Shen Z, et al. Daring: Differentiable causal discovery with residual independence[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 596-605.]

DARING Pipeline



生成器学习图结构，最小化重构误差

判别器最大化残差间的相关性



因果图示例

DARING增加残差独立性约束，通过对抗学习优化两套参数

[He Y, Cui P, Shen Z, et al. Daring: Differentiable causal discovery with residual independence[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 596-605.]

DARING模型

东南大学 SOUTHEAST UNIVERSITY

□ 残差独立性度量：

➢ 双变量独立性：

LEMMA 3.2 (DAUDIN [3]). *X and Y are independent if and only if for all functions $h \in L^2_X$, $g \in L^2_Y$,*

$$\mathrm{Cov}[h(X), g(Y)] = 0, \tag{6}$$

where

$$\begin{aligned} L^2_X &= \left\{h(X) \mid \mathbb{E}\left[h(X)^2\right] < \infty\right\}, \\ L^2_Y &= \left\{g(Y) \mid \mathbb{E}\left[g(Y)^2\right] < \infty\right\}, \end{aligned} \tag{7}$$

are square summable functions on X and Y.

➢ 多变量独立性：

THEOREM 3.3. *Let $R = \{R_i\}_{i=1}^d$ be a set of random variables and $R_{-i} = \{R_1, ..., R_{i-1}, R_{i+1}, ..., R_d\}$. All variables of R are mutually independent if and only if $\forall h_i \in L^2_{R_{-i}}$, $\forall g_i \in L^2_{R_i}$, $i \in \{1, \ldots, d\}$,*

$$\mathrm{Cov}[h_i(R_{-i}), g_i(R_i)]| = 0. \tag{8}$$

[He Y, Cui P, Shen Z, et al. Daring: Differentiable causal discovery with residual independence[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 596-605.]

DARING定理证明

东南大学 SOUTHEAST UNIVERSITY

THEOREM. *Let* $R = \{R_i\}_{i=1}^{d}$ *be a set of random variables and* $R_{-i} = \{R_1, ..., R_{i-1}, R_{i+1}, ..., R_d\}$. *All variables of R are mutually independent if and only if* $\forall h_i \in L^2_{R_{-i}}, \forall g_i \in L^2_{R_i}, i \in \{1, \ldots, d\}$,

$$\mathrm{Cov}[h_i(R_{-i}), g_i(R_i)]| = 0. \tag{12}$$

Similar with Equation 7, $L^2_{R_{-i}}$ *and* $L^2_{R_i}$ *are the spaces of square summable functions on* R_{-i} *and* R_i *respectively.*

PROOF. On the basis of Lemma 3.1, $\forall i$, given the condition

$$\mathrm{Cov}[h_i(R_{-i}) \cdot g_i(R_i)] = 0, \quad \forall h_i \in L^2_{R_{-i}}, g_i \in L^2_{R_i},$$

[He Y, Cui P, Shen Z, et al. Daring: Differentiable causal discovery with residual independence[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 596-605.]

DARING定理证明

we have $R_i \perp R_{-i}$, i.e.,

$$P(R) = P(R_i) \cdot P(R_{-i}).$$

Integrate the above function over $R_1, ..., R_{i-1}$, we have

$$P(R_i, ..., R_d) = P(R_i) \cdot P(R_{i+1}, \ldots, R_d).$$

Hence,

$$\begin{aligned} P(R) &= P(R_1)P(R_2, R_3, \ldots, R_d) \\ P(R) &= P(R_1)P(R_2)P(R_3, \ldots, R_d) \\ &= \cdots \\ &= \prod_{i=1}^{d} P(R_i). \end{aligned}$$

As a result, R are mutually independent. 证毕 □

证明：以上过程反向，能推出$P(R) = P(R_i)P(R_{-i})$，即$R_i \perp R_{-i}$，再由引理即可。

[He Y, Cui P, Shen Z, et al. Daring: Differentiable causal discovery with residual independence[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 596-605.]

DARING模型

东南大学 SOUTHEAST UNIVERSITY

□ 提出衡量独立性的统计量：

$$M(R)=\sum_{i=1}^{d}\boxed{\sup_{h_i\in L^2_{R_{-i}},\ g_i\in L^2_{R_i}}}\left\|\frac{\mathrm{Cov}[h_i(R_{-i}),g_i(R_i)]}{\sqrt{\mathrm{Var}[h_i(R_{-i})]}\cdot\sqrt{\mathrm{Var}[g_i(R_i)]}}\right\| \qquad M(R)\in[0,d]$$

求上确界来作为惩罚项

1-范数

h_i和g_i可以通过NN拟合，但在数据有限情况下，参数空间巨大，容易过拟合。

为防止过拟合，令$g_i(R_i)=R_i$，减少参数空间

为适应连续优化框架，用2-范数替换1-范数

$$\mathcal{L}_M(R,\phi)=\sum_{i=1}^{d}\left\|\frac{\mathrm{Cov}[\mathrm{MLP}(R_{-i},\phi_i),R_i]}{\sqrt{\mathrm{Var}[\mathrm{MLP}(R_{-i},\phi_i)]}\cdot\sqrt{\mathrm{Var}[R_i]}}\right\|_2^2$$

残差独立性统计量

DARING提出衡量残差独立性的统计量，优化可微因果发现学习过程

[He Y, Cui P, Shen Z, et al. Daring: Differentiable causal discovery with residual independence[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 596-605.]

DARING模型

东南大学 SOUTHEAST UNIVERSITY

◻ 优化问题：

$$\min_{G,\theta}\max_{\phi}\mathcal{L}(\mathbf{X},G,\theta)=\mathcal{L}_{\text{rec}}(G,\mathbf{X},\theta)+\alpha\mathcal{L}_{\text{DAG}}(G)$$

$$+\beta\mathcal{L}_{\text{sparse}}(G)+\gamma\mathcal{L}_{\text{M}}(X-f(X,\theta),\phi)$$

生成器　判别器

- **重构项：** $\mathcal{L}_{\text{rec}}(G,\mathbf{X},\theta)$
- **残差独立性约束：** $\mathcal{L}_{\text{M}}(X-f(X,\theta),\phi)$

$$\mathcal{L}_{\text{M}}(R,\phi)=\sum_{i=1}^{d}\left\|\frac{\text{Cov}[\text{MLP}(R_{-i},\phi_i),R_i]}{\sqrt{\text{Var}[\text{MLP}(R_{-i},\phi_i)]}\cdot\sqrt{\text{Var}[R_i]}}\right\|_2^2$$

- **DAG约束：** $\mathcal{L}_{\text{DAG}}=\alpha_t h(\mathcal{G})+\frac{\mu_t}{2}|h(\mathcal{G})|^2$
- **稀疏性约束：** **L1，L2正则化**

通过对抗学习优化生成器和判别器参数

[He Y, Cui P, Shen Z, et al. Daring: Differentiable causal discovery with residual independence[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 596-605.]

DARING算法

东南大学 SOUTHEAST UNIVERSITY

Algorithm 1 Causal Discovery with DARING

Input: $\mathbf{X} = \left\{\mathbf{x}^{(k)}\right\}_{k=1}^{n}$ i.i.d. sampled from $P(X)$ and threshold Δ
Output: Causal graph G
Initial $\boldsymbol{G}$, parameters of causality fitting model θ $(\theta_1, ..., \theta_d)$ and parameters of independence test model ϕ $(\phi_1, ..., \phi_d)$
$\boxed{\text{Pretrain } G \text{ and } \theta \text{ to minimize } \mathcal{L}^{(0)} \text{ for } \tau_0 \text{ steps}}$
while not arriving maximal iteration or triggering termination conditions **do**
  for $t = 1$ to τ_1 **do**
    Fix G, θ and calculate $\mathcal{L}_{\mathrm{M}}(R, \phi)$ in Equation 10
    Update ϕ to maximize $\mathcal{L}_{\mathrm{M}}(R, \phi)$
  end for
  for $t = 1$ to τ_2 **do**
    Fix ϕ and calculate total $\mathcal{L}$ in Equation 11
    Update G, θ to minimize $\mathcal{L}$
  end for
end while
Prune the edges less than Δ of G
return: G

为了更好地收敛，先预训练几个epoch

在给定DAG拟合模型下，学习∅

$$\max_{\phi} \mathcal{L}_{\mathrm{M}}(R, \phi) = \sum_{i=1}^{d} \left\| \frac{\mathrm{Cov}[\mathrm{MLP}(R_{-i}, \phi_i), R_i]}{\sqrt{\mathrm{Var}[\mathrm{MLP}(R_{-i}, \phi_i)]} \cdot \sqrt{\mathrm{Var}[R_i]}} \right\|_2^2$$

在给定∅下，学习DAG拟合模型

$$\min_{G,\theta} \max_{\phi} \mathcal{L}(\mathbf{X}, G, \theta) = \mathcal{L}_{\mathrm{rec}}(G, \mathbf{X}, \theta) + \alpha \mathcal{L}_{\mathrm{DAG}}(G) + \beta \mathcal{L}_{\mathrm{sparse}}(G) + \gamma \mathcal{L}_{\mathrm{M}}(X - f(X, \theta), \phi)$$

[He Y, Cui P, Shen Z, et al. Daring: Differentiable causal discovery with residual independence[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 596-605.]