東南大學
SOUTHEAST UNIVERSITY

Discovering Dynamic Causal Space for DAG Structure Learning

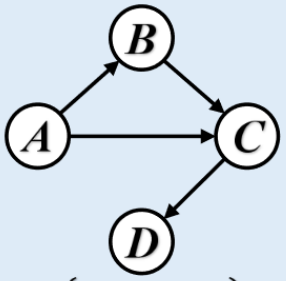
汇报人：易晖洋

日期：2023.08.06

- 研究背景与意义
- CASPER
- 实验结果

1

CASPER研究背景与意义



$A := \epsilon_A(\sim U(-1,1))$

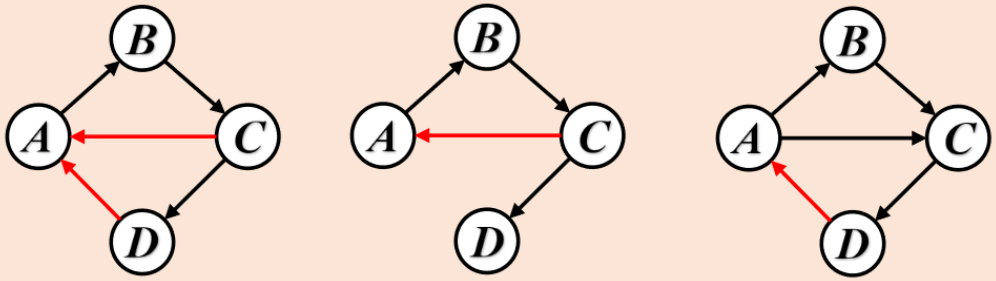
$B := 2A + \epsilon_B(\sim N(0,2))$

$C := A + 0.5B + \epsilon_C(\sim N(-1,1))$

$D := 0.5C + \epsilon_D(\sim U(0,1))$

NOTEARS评分函数对$h(W)$变化不敏感

学习过程：



	Phase-1	Phase-2	Phase-3
$h(W)$	19.16	10.26	4.39
$Score(NOTEARS)$	15.47	14.45	14.39
$Score(CASPER)$	14.43	7.64	1.92

□ 基于NOTEARS框架的组件：评分函数、DAG约束、神经网络

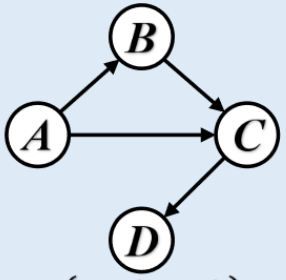
- 评分函数仅考虑数据适应性，未考虑DAG-ness，容易陷入局部最优。
- 评分函数无法量化估计DAG与真实DAG间因果距离。
- DAG约束项系数必须达到无穷大，才满足无环性。

上述实验的非线性版

	Phase-1	Phase-2	Phase-3
h(W)	20.13	10.32	3.91
Score(NOTEARS)	10.20	10.43	9.87
Score(CASPER)	10.53	6.04	1.55

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

CASPER研究背景与意义

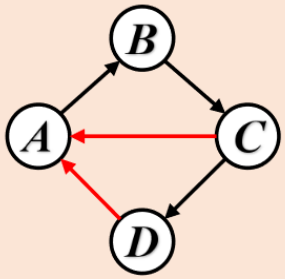


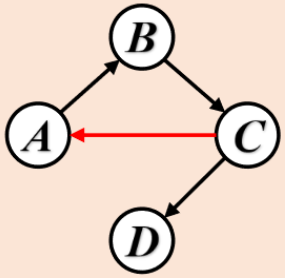
$A := \epsilon_A(\sim U(-1,1))$

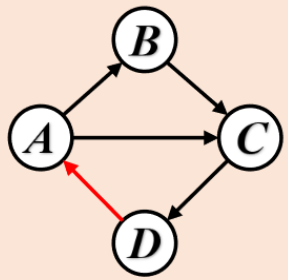
$B := 2A + \epsilon_B(\sim N(0,2))$

$C := A + 0.5B + \epsilon_C(\sim N(-1,1))$

$D := 0.5C + \epsilon_D(\sim U(0,1))$







	Phase-1	Phase-2	Phase-3
$h(W)$	19.16	10.26	4.39
$Score(NOTEARS)$	15.47	14.45	14.39
$Score(\boldsymbol{CASPER})$	14.43	7.64	1.92

□ CASPER框架组件：DAG-ness感知评分函数、DAG约束、神经网络

- 将图结构信息集成到得分函数，考虑了DAG-ness信息，能刻画估计DAG与真实DAG间距离，进一步对噪声鲁棒。

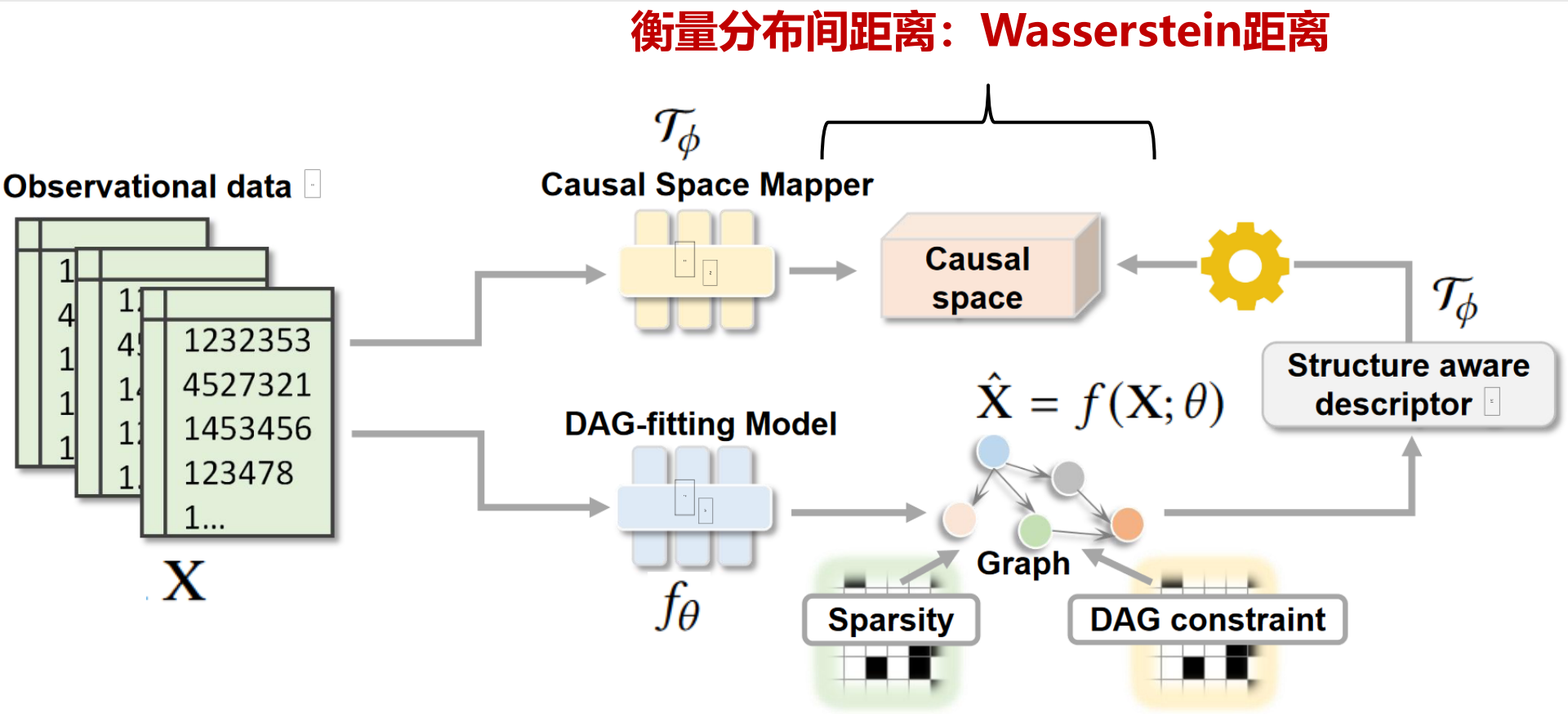
CASPER能感知DAG结构，提升因果发现性能

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

- 研究背景与意义
- CASPER
- 实验结果

2

CASPER Pipeline



$\mathcal{T}_\phi$ 将数据映射到因果空间，能衡量真实DAG与估计DAG间距离

因果空间中能提供DAG-ness信息，进一步得到更准确CD结果

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

CASPER模型

双层优化：

$$\min_{\mathcal{G},\theta} F_{\phi^*}(\mathbf{X};\mathcal{G},\theta) + \mathcal{L}_{\text{DAG}}(\mathcal{G},\alpha_t,\mu_t)$$

$$\text{s.t.} \quad \phi^* \in \arg\max_{\phi \in C(\mathcal{G})} F_\phi(\mathbf{X};\mathcal{G},\theta),$$

$\emptyset \in C(G)$：权重裁剪，保持模型稳定，防止梯度爆炸与消失。将DAG-ness融入评分函数，$\emptyset$的拟合能力随着DAG-ness变化

$$C(\mathcal{G}) := \{\phi : \mathcal{T}_\phi \text{ is continuous}, \|\mathcal{T}_\phi\|_{\text{Lip}} \le g(h(\mathcal{G}))\}$$

$g(\cdot)$ is an increasing function which is $g(x) = \log(1+x)$

$$\|\mathcal{T}\|_{\text{Lip}} := \sup_{a\neq b, a,b\in\mathcal{M}_A} \frac{|\mathcal{T}(a)-\mathcal{T}(b)|}{\|a-b\|}$$

➢ DAG-ness感知评分函数：

Wasserstein距离衡量分布间距离，即衡量真实DAG与估计DAG间距离

$$F_\phi(\mathbf{X};\mathcal{G},\theta) = \left\{\mathbb{E}_{\mathbf{X}\sim P_r}[(\mathcal{T}_\phi(\mathbf{X}))] - \mathbb{E}_{\hat{\mathbf{X}}\sim P_\theta}[(\mathcal{T}_\phi(\hat{\mathbf{X}})]\right\} + \lambda\mathcal{R}_{\text{sparse}}(\mathcal{G})$$

观测数据分布　　重构数据分布

➢ 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

将DAG-ness信息融入评分函数中，提升结构学习准确度

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

CASPER算法

Algorithm 1 CASPER Algorithm for DAG Structure Learning

Input: observational data $\mathbf{X} = \{\mathbf{x}^{(k)}\}_{k=1}^{n}$ sampled from P_r and threshold $\omega > 0$, maximum epoch in the inner loop K_{inner}, maximum epoch in the outer loop K_{outer}
Initialize: initialize the parameters of causal fitting model θ and parameters of causal space model ϕ
for $t = 0$ to τ_0 **do**
  Update θ and $\mathcal{G}$ to minimize F_ϕ and get $\mathcal{G}^{\text{pre}}$
end for
for $k_1 = 0$ to K_{outer} **do**
  Fix causal space model parameters ϕ
  Calculate $F_\phi(\mathbf{X}; \mathcal{G}, \theta) + \mathcal{L}_{\text{DAG}}(\mathcal{G})$ in Equation 12
  Update θ and $\mathcal{G}$ to minimize $F_\phi + \mathcal{L}_{\text{DAG}}$
  for $k_2 = 0$ to K_{inner} **do**
    Fix graph $\mathcal{G}$ and the causal fitting model's parameters θ
    Update ϕ to maximize $F_\phi(\mathbf{X}; \mathcal{G}, \theta)$ in Equation 11
    $c \leftarrow \log(1 + h(\mathcal{G}))$
    $\phi \leftarrow \text{clip}(\phi, -c, c)$
  end for
end for
Prune the edges less than ω of $\mathcal{G}$
return predicted $\mathcal{G}$

为了更好地收敛，先预训练几个epoch

权重裁剪

通过交替训练内外循环，得到更准确的梯度优化+更快地收敛

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

- 研究背景与意义
- CASPER
- 实验结果

3

CASPER实验结果-合成数据

□ 线性、不同节点数和图密度：

Table 1: Linear Setting, for ER graphs of 10, 20, 50 nodes.

	10 nodes				20 nodes				50 nodes			
ER2	TPR↑	FDR↓	SHD↓	SID↓	TPR↑	FDR↓	SHD↓	SID↓	TPR↑	FDR↓	SHD↓	SID↓
Random	0.08±0.07	0.93±0.18	33.2±7.3	95.6±12.2	0.11±0.09	0.89±0.08	56.8±8.7	292.3±45.7	0.04±0.02	0.90±0.03	397.3±12.7	1,082.0 ±182.2
NOTEARS	0.82±0.07	0.09±0.05	5.4±1.6	16.6±5.8	0.82±0.09	0.13±0.04	9.4±4.1	59.4±10.7	0.79±0.06	0.19±0.03	27.6±7.7	427.0±186.1
DAG-GNN	0.83±0.05	0.12±0.05	4.8±1.1	12.9±6.2	0.83±0.02	0.13±0.02	8.7±2.5	48.5±5.3	0.81±0.03	0.13±0.02	24.3±5.5	334.2±120.3
NoCurl	0.84±0.04	0.13±0.03	4.6±1.3	13.2±5.1	0.82±0.05	0.15±0.05	8.9±3.4	50.1±6.7	0.78±0.07	0.15±0.03	25.2±6.0	356.7±165.2
GraN-DAG	0.82±0.03	0.08±0.01	5.2±0.9	14.8±4.9	0.80±0.06	0.14±0.01	8.5±2.9	47.2±8.0	0.82±0.05	0.12±0.01	24.8±7.6	289.1±118.3
DARING	0.85±0.02	0.10±0.01	4.3±1.7	13.4±4.5	0.84±0.05	0.16±0.02	8.9±3.0	46.7±6.5	0.83±0.06	0.13±0.02	23.5±6.2	310.8±159.6
CASPER(Ours)	**0.90**±0.04	**0.07**±0.02	**3.8**±0.8	**11.6**±4.3	**0.89**±0.09	**0.10**±0.03	**7.8**±3.7	**42.4**±7.2	**0.87**±0.05	**0.12**±0.03	**21.8**±5.8	**230.4**±119.8
ER4	TPR↑	FDR↓	SHD↓	SID↓	TPR↑	FDR↓	SHD↓	SID↓	TPR↑	FDR↓	SHD↓	SID↓
Random	0.09±0.17	0.93±0.09	52.3±16.7	80.3±17.7	0.07±0.03	0.90±0.08	86.9±7.0	387.5±52.3	0.09±0.08	0.92±0.08	998.2±45.9	3,399.1±489.2
NOTEARS	0.83±0.06	0.08±0.03	7.4±2.7	28.4±5.8	0.75±0.01	0.28±0.05	32.0±5.4	152.8±27.0	0.51±0.12	**0.27**±0.10	113.4±29.5	943.8±172.2
DAG-GNN	0.82±0.07	0.12±0.01	7.0±1.6	29.4±3.3	0.81±0.02	0.25±0.04	29.5±3.3	138.4±18.9	0.55±0.09	0.28±0.08	115.2±25.4	835.3±154.1
NoCurl	0.86±0.10	0.07±0.02	6.5±2.3	26.0±4.9	0.79±0.03	0.27±0.03	31.3±2.1	142.0±14.9	0.59±0.10	0.29±0.06	105.7±26.2	910.5±129.0
GraN-DAG	0.84±0.04	0.06±0.03	7.8±2.1	25.5±5.0	0.78±0.03	0.26±0.04	29.7±3.4	143.5±17.0	0.52±0.08	0.31±0.05	110.3±23.4	854.3±178.5
DARING	0.83±0.06	0.09±0.01	6.8±1.8	27.8±3.5	0.80±0.02	0.24±0.02	29.3±2.0	139.1±15.4	0.50±0.12	0.33±0.05	118.9±27.0	809.4±165.3
CASPER(Ours)	**0.88**±0.05	**0.06**±0.04	**6.2**±2.1	**25.0**±2.7	**0.85**±0.03	**0.19**±0.02	**27.5**±2.9	**132.0**±16.3	**0.63**±0.10	0.29±0.10	**98.4**±31.1	**735.0**±160.2

CASPER在线性情况下，取得SOTA结果

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

CASPER实验结果-合成数据

□ 非线性、不同节点数和图密度：

Table 2: Nonlinear Setting, for ER graphs of 10, 20, 50 nodes.

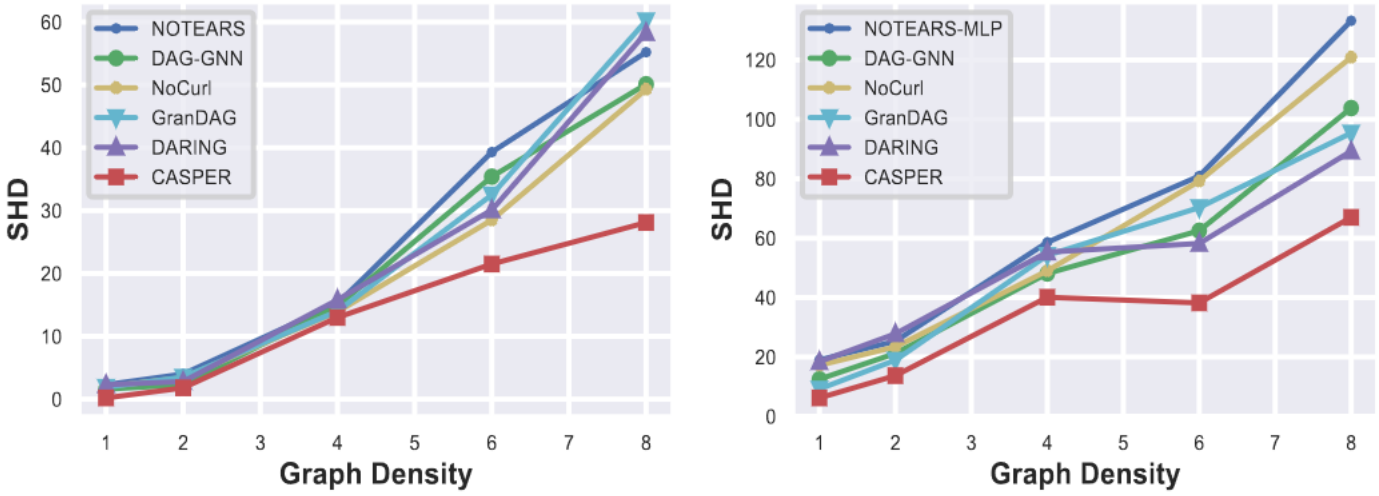
	10 nodes				20 nodes				50 nodes			
ER2	TPR↑	FDR↓	SHD↓	SID↓	TPR↑	FDR↓	SHD↓	SID↓	TPR↑	FDR↓	SHD↓	SID↓
Random	0.06±0.07	0.94±0.18	35.2±7.3	95.6±12.2	0.08±0.09	0.89±0.07	59.8±9.7	392.3±48.7	0.04±0.02	0.92±0.03	486.3±23.7	1,134.2 ±210.3
NOTEARS-MLP	0.75±0.12	0.16±0.09	7.6±2.3	18.3±9.1	0.71±0.12	0.16±0.08	15.3±6.1	99.3±18.4	0.37±0.03	0.19±0.07	70.5±8.7	892.5±146.4
DAG-GNN	0.81±0.09	0.14±0.08	7.0±2.1	14.1±6.3	0.78±0.09	0.12±0.03	10.1±5.8	80.3±12.6	0.41±0.07	0.23±0.05	59.2±6.5	698.4±103.5
NoCurl	0.80±0.07	0.17±0.07	6.7±2.4	15.3±5.0	0.72±0.12	0.19±0.03	12.5±4.3	77.9±12.3	0.49±0.05	0.18±0.06	69.8±7.4	733.5±130.4
GraN-DAG	0.83±0.05	0.12±0.05	5.1±1.9	11.5±3.4	0.81±0.14	0.16±0.09	9.9±4.6	65.4±11.7	0.52±0.04	0.14±0.04	52.8±8.6	635.4±172.8
DARING	0.79±0.09	0.21±0.03	7.7±3.1	18.2±4.8	0.73±0.07	0.20±0.06	13.6±5.2	88.3±24.4	0.50±0.07	0.13±0.05	57.4±9.3	745.2±120.6
CASPER(Ours)	**0.87**± 0.13	**0.11**±0.07	**3.5**±1.8	**7.8**±4.1	**0.85**±0.08	**0.09**±0.04	**7.9**±1.5	**55.1**±9.8	**0.59**±0.06	**0.11**±0.03	**45.2**±6.0	**584.3**±102.7
ER4	TPR↑	FDR↓	SHD↓	SID↓	TPR↑	FDR↓	SHD↓	SID↓	TPR↑	FDR↓	SHD↓	SID↓
Random	0.07±0.16	0.94±0.09	51.4±15.7	82.3±17.7	0.06±0.04	0.93±0.18	96.8±6.9	392.1±42.3	0.07±0.07	0.92±0.05	1,198.2±54.2	4,065.8±584.2
NOTEARS-MLP	0.83±0.14	0.23±0.03	10.5±1.9	28.5±11.2	0.48±0.09	0.27±0.05	55.6±9.3	174.5±32.1	0.28±0.08	0.12±0.06	158.2±10.4	1,603.5±88.9
DAG-GNN	0.87±0.13	0.18±0.03	6.8±1.3	18.7±4.8	0.52±0.03	0.21±0.12	49.2±10.2	150.3±32.7	0.43±0.06	0.15±0.04	150.4±7.2	1,536.9±90.4
NoCurl	0.79±0.08	0.24±0.06	8.5±3.5	15.2±7.7	0.43±0.05	0.23±0.06	53.3±8.4	167.9±34.4	0.33±0.04	0.14±0.06	140.3±7.6	1,468.5±100.2
GraN-DAG	0.90±0.10	**0.14**±0.02	6.4±1.1	5.8±0.9	0.47±0.08	0.25±0.08	47.5±7.0	149.8±28.3	0.42±0.04	**0.06**±0.03	128.6±8.4	1,232.4±96.7
DARING	0.85±0.07	0.18±0.09	7.1±1.6	13.7±5.9	0.48±0.07	0.29±0.10	57.2±4.6	180.0±43.5	0.30±0.05	0.16±0.05	136.9±12.5	1,653.0±78.4
CASPER(Ours)	**0.92**±0.06	0.15±0.04	**4.3**±2.1	**4.1**±1.1	**0.56**±0.04	**0.17**±0.09	**42.3**±5.6	**123.2**±24.5	**0.51**±0.03	0.08±0.04	**118.5**±8.0	**1,150.3**±70.2

CASPER在非线性情况下，取得SOTA结果

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

CASPER实验结果-合成数据

不同图密度、线性和非线性：



(a) Linear Setting (b) Nonlinear Setting

Figure 4: SHD comparisons for different graph density conditions in SF graph with 20 nodes.

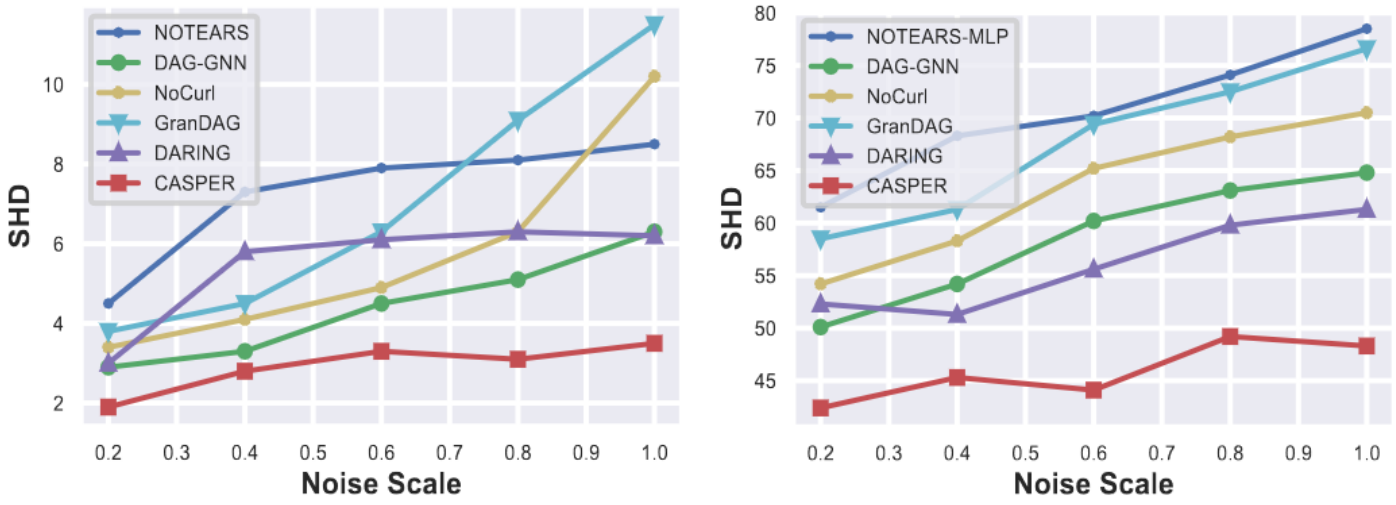
随着节点度的增加，CASPER相对于baselines改进越来越大

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

CASPER实验结果-合成异质数据

加性噪声，均值变化：

$$X_j := f_j(X_{pa(X_j)}) + N_j, \quad j \in \{1, ...d\}, \qquad N(\mu, 1), \mu \in \{0.2, 0.4, 0.6, 0.8, 1\}$$



(a) Linear Setting (b) Nonlinear Setting

Figure 3: SHD comparisons for various noise scales in SF2 graph with 20 nodes.

CASPER对加性噪声均值变化鲁棒，且均取得SOTA结果

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

实验结果-真实异质数据Sachs

Table 3: Empiricle results on Sachs [43] dataset.

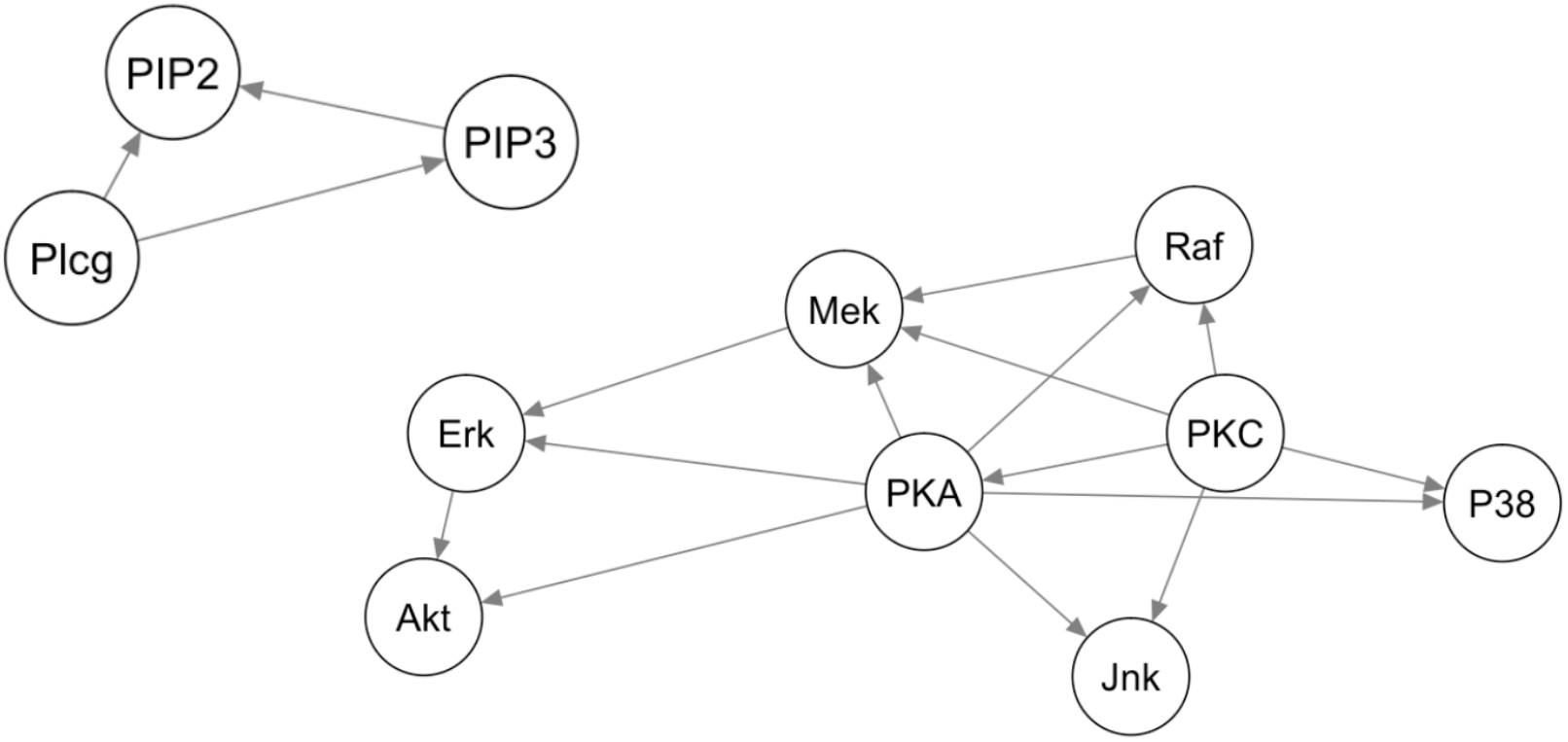
Method	#Predicted Edges	#Correct Edges	SHD	SID
Random	22	1	23	63
GES [6]	34	7	31	54
ICA-LiNGAM [46]	8	4	14	55
FGS [42]	17	5	22	51
GOLEM [31]	6	4	15	53
CD-NOD [19]	18	7	15	-
NOTEARS [58]	20	6	17	48
NOTEARS-MLP [59]	19	7	16	45
DAG-GNN [54]	18	6	19	49
DARING [17]	19	7	16	46
NoCurl [55]	18	5	16	50
GraN-DAG [25]	14	4	16	60
CASPER(Ours)	**15**	**8**	**12**	**42**

CASPER在Sachs数据集上取得SOTA效果

[Liu F, Ma W, Zhang A, et al. Discovering Dynamic Causal Space for DAG Structure Learning[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023.]

Sachs-因果发现benchmark

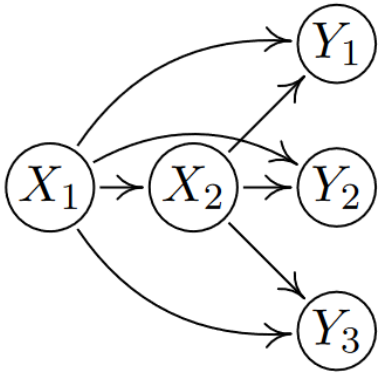
- **Real-world heterogeneous data：11个点，17条边，样本数7466，节点之间有向边表示2个蛋白质表达水平变量间存在因果关系。**



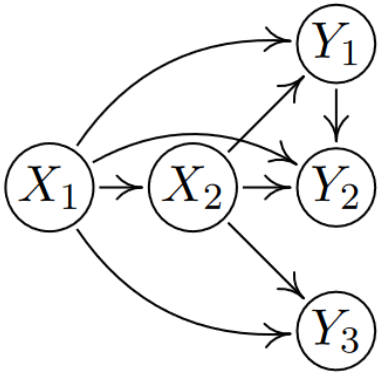
[Sachs K, Perez O, Pe'er D, et al. Causal protein-signaling networks derived from multiparameter single-cell data[J]. Science, 2005, 308(5721): 523-529.]

结构干预距离SID

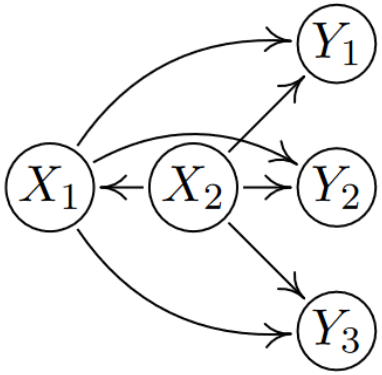
SHD相同，SID不同：



true graph $\mathcal{G}$



graph $\mathcal{H}_1$



graph $\mathcal{H}_2$

$$\mathrm{SHD}(\mathcal{G},\mathcal{H}_1) = 1 = \mathrm{SHD}(\mathcal{G},\mathcal{H}_2) \qquad \mathrm{SID}(\mathcal{G},\mathcal{H}_1) = 0 \neq 8 = \mathrm{SID}(\mathcal{G},\mathcal{H}_2)$$

Definition 4 (Structural Intervention Distance) *Let $\mathbb{G}$ be the space of DAGs over p variables. We then define*

$$\begin{aligned} \mathrm{SID}: \mathbb{G}\times\mathbb{G} &\to \mathbb{N} \\ (\mathcal{G},\mathcal{H}) &\mapsto \#\{(i,j), i\neq j \mid \textit{the intervention distribution from } i \textit{ to } j \\ &\qquad \textit{is falsely estimated by } \mathcal{H} \textit{ with respect to } \mathcal{G}\} \end{aligned} \tag{3}$$

SHD计算图结构差异，SID计算干预分布的差异

[Peters J, Bühlmann P. Structural intervention distance for evaluating causal graphs[J]. Neural computation, 2015, 27(3): 771-799.]

Thanks!
Q&A