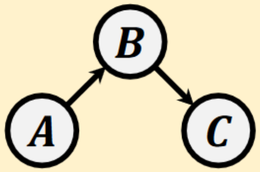
Rescore研究背景与意义



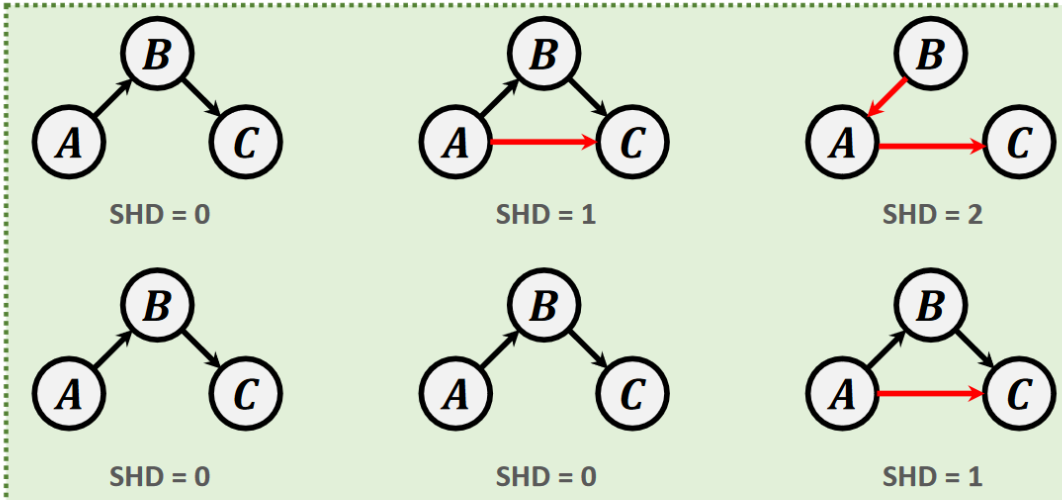
$$A \coloneqq \ \varepsilon_A \sim U(-2{,}2)$$

$$B \coloneqq \frac{1}{2} A + \ \varepsilon_B \sim \begin{cases} N(0{,}1), & P_1 \\ N(0{,}0.1), & P_2 \end{cases}$$

$$C \coloneqq 2B + \ \varepsilon_C \sim N(0{,}1)$$

NOTEARS

NOTEARS + ReScore



$P_1 = 1, P_2 = 0$ 同质性　　$P_1 = 0.2, P_2 = 0.8$ 异质性　　$P_1 = 0, P_2 = 1$ 同质性

◻ 当前基于可微因果发现的2个缺点：

- ➢ 某些情况结构学习准确性低，**线性系统中链结构**容易被 NOTEARS 错误识别 （Reisach et al., NIPS’21）
- ➢ **异质数据**中，NOTEARS 很容易学到冗余因果关系。当前解决异质性的方法依赖于域注释的先验（DICD，CD-NOD (Huang et al., JMLR’20)）

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

Rescore研究背景与意义

东南大学 SOUTHEAST UNIVERSITY

□ 可微CD失败原因（生成模型视角理解）：

- 收集的数据集自然包括**大量的简单**样本和**少量**可能包含**关键因果信息**的样本，**平均评分函数**无法反映样本重要性（Shrivastava et al., CVPR'16）

□ 实验证明ReScore自适应学习样本权重：

- NOTEARS-MLP+ReScore应用于Sachs数据集（9组），学习样本权重
- 每组随机消除500样本，运行NOTEARS-MLP，得到相应SHD和TPR。随着组重要性增加，结果影响越大

Table 5: Performance comparison for removing samples in different groups

Group Index	3	5	7	1	2	6	4	8	0
Avg. ranking	578.4	2856.7	3368.1	3877.0	3949.4	4549.4	4573.2	4590.6	4910.1
SHD w/o group	16	16	17	16	16	17	17	19	19
TPR w/o group	0.529	0.412	0.412	0.412	0.412	0.412	0.412	0.353	0.294

不同重要性样本自然存在于真实数据集中，ReScore能提取这种重要性

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

ReScore模型

东南大学 SOUTHEAST UNIVERSITY

□ 双层优化：

$$\min_{\mathcal{G}} \ S_{\mathbf{w}^*}(\mathcal{G};\mathbf{X}) + \mathcal{P}_{DAG}(\mathcal{G}),$$

$$\boxed{\text{s.t. } \mathbf{w}^* \in \arg\max_{\mathbf{w}\in\mathbb{C}(\tau)} S_{\mathbf{w}}(\mathcal{G};\mathbf{X}),}$$

自适应学习样本权重：在G给定情况下，大量简单样本的loss值小，少量关键样本的loss值大，通过评分函数最大化，简单样本学习得到较小的权重，关键样本学习得到较大的权重。

where $\mathbb{C}(\tau) := \{\mathbf{w} : 0 < \frac{\tau}{n} \le w_1, \ldots, w_n \le \frac{1}{\tau n}, \sum_{i=1}^{n} w_i = 1\}$ for the cutoff threshold $\tau \in (0,1)$

$\tau \to 1$时，权重分布趋于均匀分布。$\tau \to 0$时，权重分布可以更大地偏离均匀分布

➢ 重加权评分函数：

$$S_{\mathbf{w}}(\mathcal{G};\mathbf{X}) = \mathcal{L}_{\mathbf{w}}(\mathcal{G};\mathbf{X}) + \lambda\mathcal{R}_{\text{sparse}}(\mathcal{G}) = \sum_{i=1}^{n} w_i l(\mathbf{x}_i, f(\mathbf{x}_i)) + \lambda\mathcal{R}_{\text{sparse}}(\mathcal{G}),$$

➢ DAG约束： $\mathcal{P}_{DAG}(\mathcal{G}) = \alpha_t h(\mathcal{G}) + \frac{\mu_t}{2}|h(\mathcal{G})|^2$

➢ 稀疏性约束：L1和L2正则化

通过最大化评分函数自适应学习样本权重

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

ReScore算法

东南大学 SOUTHEAST UNIVERSITY

Algorithm 1 ReScore Algorithm for Differentiable Score-based Causal Discovery

Input: observational data $\mathcal{D}$: $\{\mathbf{x}_i : i = 1, 2, ..., n\}$, DAG learner parameters $\theta_{\mathcal{G}}$, reweighting model parameters θ_w, cutoff threshold τ, epoch to start reweighting $K_{reweight}$, maximum epoch in the inner loop K_{inner}, maximum epoch in the outer loop K_{outer}

Initialize: initialize θ_w to uniformly output $\frac{1}{n}$, $k_1 = 0$, $k_2 = 0$

for $k_1 \leq K_{outer}$ **do**
  Fix reweighting model parameters θ_w
  Calculate $\mathbf{w}^*$ by applying threshold $[\frac{\tau}{n}, \frac{1}{n\tau}]$
  Optimize $\theta_{\mathcal{G}}$ by minimizing $S_{\mathbf{w}^*}(\mathcal{G}; \mathbf{X}) + \mathcal{P}_{DAG}(\mathcal{G})$ # Outer optimization in Equation 6
  if $k_1 \geq k_{reweight}$ **then**
    for $k_2 \leq K_{inner}$ **do**
      Fix the DAG learner's parameters $\theta_{\mathcal{G}}$
      Get $\mathbf{w}$ from θ_w by applying threshold $[\frac{\tau}{n}, \frac{1}{n\tau}]$
      Optimize θ_w by maximizing $S_{\mathbf{w}}(\mathcal{G}; \mathbf{X})$ # Inner optimization in Equation 6
      $k_2 \leftarrow k_2 + 1$
    end for
    $k_1 \leftarrow k_1 + 1$
    $k_2 \leftarrow 0$
  end if
end for
return predicted $\mathcal{G}$ from DAG learner

外循环（生成器）：通过内循环确定的样本权重来优化DAG学习器，进一步提升CD性能

内循环（判别器）：通过NN学习样本权重，输入为各样本残差，输出为各样本权重

交替训练内外循环，根据DAG学习器误差学习样本权重

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

实验结果-合成数据实验配置

东南大学 SOUTHEAST UNIVERSITY

- **图类型：** ER或SF
- **节点个数：** 10、20、50
- **图密度：** 节点平均度数为2或4
- **函数形式：** 线性高斯等方差（噪声项服从标准正态）、非线性高斯过程（f_i服从高斯过程，噪声项服从标准正态）

$$X_i = f_i(X_{pa(i)}) + N_i \ , \ i = 1, \cdots, d$$

- **评估指标：**
  - **FDR：** $(R + FP)/P$
  - **TPR：** TP/T
  - **SID：** 衡量干预分布差异
  - **SHD：** 估计DAG转为真实DAG所需添加、删除和反转边的总数。衡量图结构差异

p:预测DAG边数

R:预测DAG中反转边数

FP:预测DAG中不在真实DAG无向骨架中的边数

T:真实DAG边集数

TP:预测DAG中预测正确的边数

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

实验结果-合成数据

□ 线性+非线性、10+50节点数、ER2+ER4：

Table 1: Results for ER graphs of 10 nodes on linear and nonlinear synthetic datasets.

	ER2				ER4			
	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	0.09±0.17	0.93±0.09	52.3±16.7	80.3±17.7
NOTEARS	0.85±0.09	**0.07**±0.07	5.8±2.2	20.8±5.2	0.79±0.11	0.09±0.05	10.0±5.2	25.8±9.9
+ ReScore	**0.89**±0.07 $^{+5\%}$	0.08±0.09 $^{-12\%}$	**4.6**±2.3 $^{+26\%}$	12.8±7.0 $^{+63\%}$	**0.85**±0.04 $^{+8\%}$	**0.05**±0.04 $^{+57\%}$	**7.2**±1.9 $^{+39\%}$	**24.2**±8.4 $^{+7\%}$
GOLEM	0.87±0.06	0.22±0.11	6.5±3.4	13.0±6.7	0.63±0.03	0.16±0.03	17.2±1.3	48.0±13.3
+ ReScore	0.88±0.06 $^{+1\%}$	0.21±0.11 $^{+2\%}$	6.0±3.4 $^{+8\%}$	**12.4**±6.3 $^{+5\%}$	0.66±0.04 $^{+5\%}$	0.17±0.01 $^{-5\%}$	16.2±1.0 $^{+6\%}$	46.7±13.3 $^{+3\%}$
NOTEARS-MLP	0.76±0.17	0.14±0.09	7.0±3.5	17.9±10.0	0.83±0.05	0.21±0.04	10.9±1.9	28.6±12.0
+ ReScore	0.73±0.07 $^{-4\%}$	0.10±0.09 $^{+37\%}$	6.8±2.9 $^{+3\%}$	20.3±9.7 $^{-11\%}$	0.94±0.06 $^{+14\%}$	0.15±0.06 $^{+44\%}$	6.80±2.7 $^{+60\%}$	8.80±12.4 $^{+225\%}$
GraN-DAG	0.88±0.06	0.02±0.03	2.7±1.6	8.70±4.8	0.98±0.02	0.12±0.03	5.4±1.1	3.70±4.71
+ ReScore	**0.90**±0.05 $^{+2\%}$	**0.01**±0.03 $^{+35\%}$	**2.4**±1.1 $^{+13\%}$	**7.20**±3.0 $^{+21\%}$	**0.99**±0.01 $^{+1\%}$	**0.11**±0.01 $^{+12\%}$	**4.80**±0.6 $^{+13\%}$	**0.50**±0.81 $^{+640\%}$

Table 9: Results for ER graphs of 50 nodes on linear and nonlinear synthetic datasets.

	ER2				ER4			
	TPR ↑	**FDR ↓**	**SHD ↓**	**SID ↓**	**TPR ↑**	**FDR ↓**	**SHD ↓**	**SID ↓**
Random	0.04±0.02	0.90±0.03	397.3±12.7	1082.0±182.2	0.09±0.08	0.92±0.08	998.2±45.9	3399.1±489.2
NOTEARS	0.79±0.06	**0.09**±0.03	27.6±7.7	427.0±186.1	0.51±0.12	**0.27**±0.10	133.4±29.5	1643.8±172.2
+ ReScore	**0.88**±0.06 $^{+11\%}$	0.15±0.04 $^{-39\%}$	**26.2**±7.6 $^{+5\%}$	**266.0**±146.4 $^{+61\%}$	**0.52**±0.21 $^{+3\%}$	0.29±0.07 $^{-7\%}$	**130.2**±37.4 $^{+2\%}$	**1453.6**±336.5 $^{+13\%}$
GOLEM	0.80±0.09	0.35±0.09	68.6±19.7	433.5±215.6	0.31±0.11	0.68±0.06	150.6±25.1	1775.4±161.6
+ ReScore	0.82±0.15 $^{+3\%}$	0.33±0.14 $^{+5\%}$	63.4±27.9 $^{+8\%}$	430.2±155.5 $^{+1\%}$	0.39±0.06 $^{+24\%}$	0.66±0.06 $^{+3\%}$	146.3±26.3 $^{+3\%}$	1643.6±114.8 $^{+8\%}$
NOTEARS-MLP	0.32±0.04	0.13±0.08	69.5±4.7	884.4±172.8	0.17±0.02	**0.06**±0.04	167.0±4.1	1607.6±97.0
+ ReScore	0.51±0.08 $^{+59\%}$	**0.10**±0.07 $^{+30\%}$	53.5±8.7 $^{+30\%}$	628.1±120.6 $^{+41\%}$	0.26±0.04 $^{+52\%}$	0.11±0.05 $^{-51\%}$	154.4±6.4 $^{+8\%}$	1437.7±111.1 $^{+12\%}$
GraN-DAG	0.52±0.09	0.15±0.05	51.6±9.3	632.8±140.3	**0.32**±0.04	0.08±0.16	141.6±8.2	1379.0±91.3
+ ReScore	**0.53**±0.06 $^{+3\%}$	0.11±0.02 $^{+36\%}$	**46.0**±6.0 $^{+12\%}$	**581.0**±104.7 $^{+9\%}$	0.31±0.03 $^{-4\%}$	0.06±0.04 $^{+32\%}$	**138.8**±7.5 $^{+2\%}$	**1351.0**±98.2 $^{+2\%}$

ReScore能提升可微CD性能，取得SOTA结果

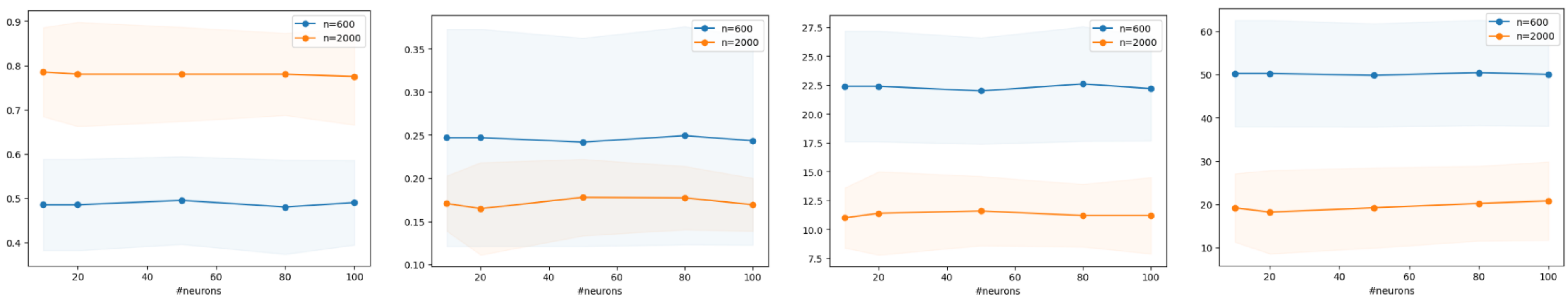
[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

实验结果-超参数影响

东南大學 SOUTHEAST UNIVERSITY

□ 研究自适应权重学习模型中隐藏层神经元数对ReScore影响：

- **实验配置：**
  - **10节点，ER4，非线性高斯过程，样本数600+2000。**
  - **隐藏层神经元数：1, 10, 20, 50, 80, 100**



(a) TPR *w.r.t.* neurons (b) FDR *w.r.t.* neurons (c) SHD *w.r.t.* neurons (d) SID *w.r.t.* neurons

Figure 5: Performance with varying neurons in ReScore model.

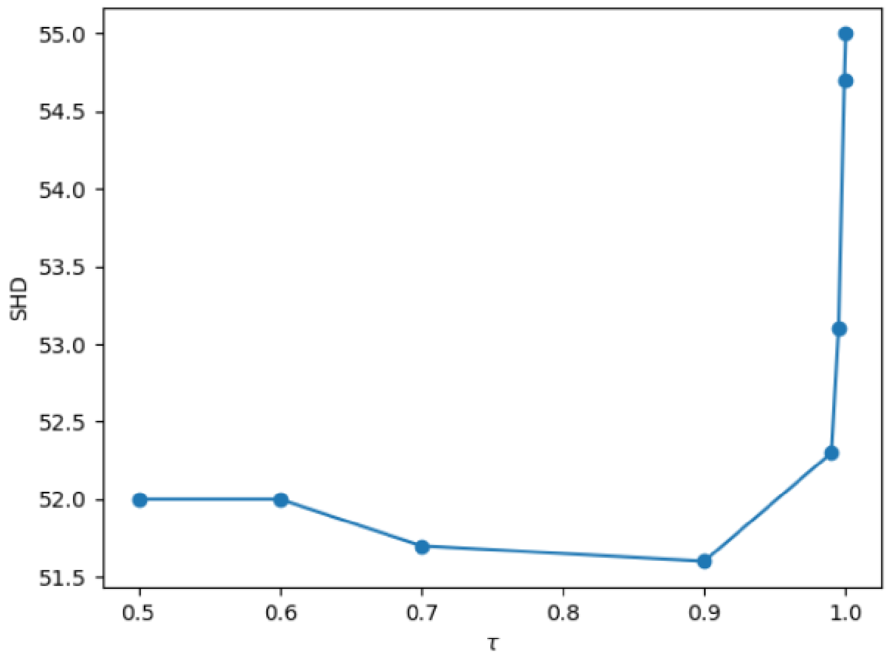
ReScore对隐藏层神经元数量不敏感；样本量越大，性能越好

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

实验结果-超参数影响

东南大学 SOUTHEAST UNIVERSITY

□ 研究阈值τ对ReScore的影响：$n=2000, d=20, ER4, GP$模型



(b) SHD *w.r.t.* threshold τ

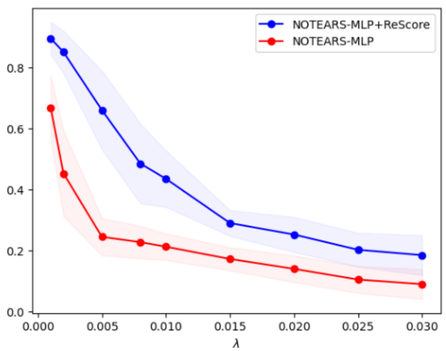
不同实验设置下，最佳阈值通常在[0.7,0.99]内

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

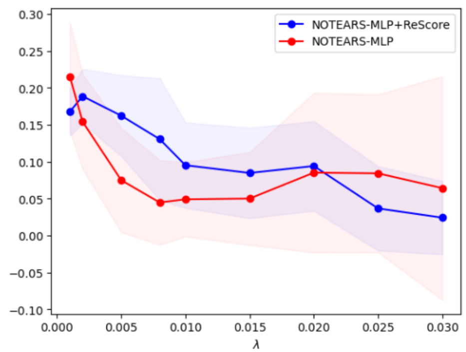
实验结果-超参数影响

☐ 研究L1正则化系数对ReScore影响：

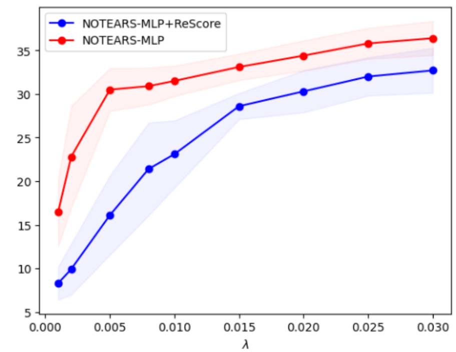
➢ **实验配置：10节点，ER4，非线性**



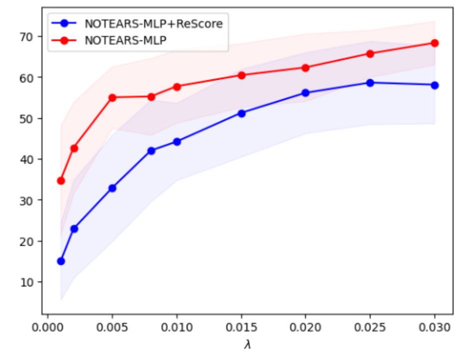
(a) TPR *w.r.t.* λ



(b) FDR *w.r.t.* λ

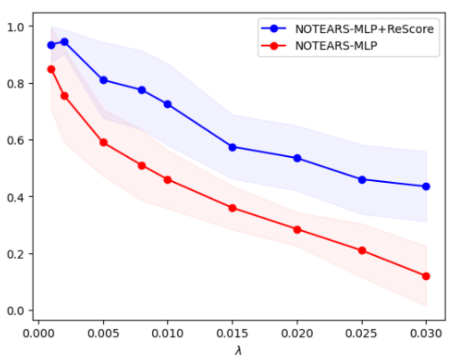


(c) SHD *w.r.t.* λ

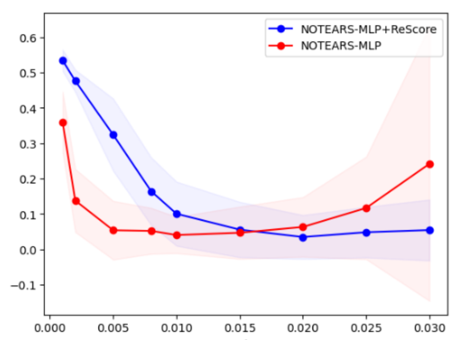


(d) SID *w.r.t.* λ

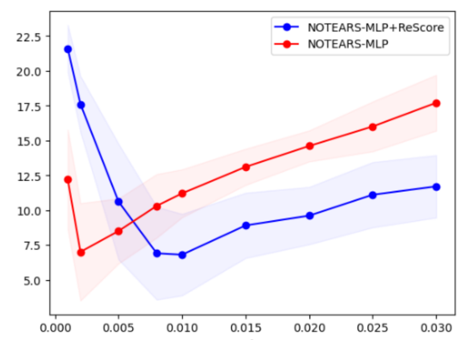
➢ **实验配置：10节点，ER2，非线性**



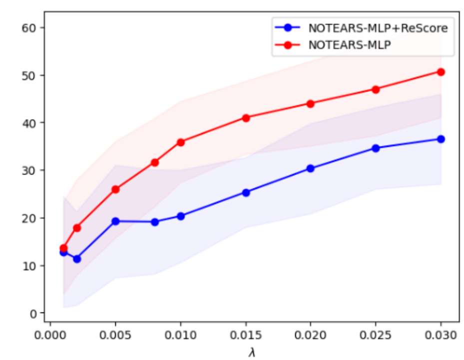
(a) TPR *w.r.t.* λ



(b) FDR *w.r.t.* λ



(c) SHD *w.r.t.* λ



(d) SID *w.r.t.* λ

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]

实验结果-真实异质数据Sachs

東南大學 SOUTHEAST UNIVERSITY

Table 3: The performance comparison on Sachs dataset.

	TPR ↑	FDR ↓	SHD ↓	SID ↓	#Predicted Edges
Random	0.076	0.899	23	63	22
GOLEM	0.176	0.026	15	53	4
+ ReScore	0.294	0.063	**14**	49	6
NOTEARS-MLP	0.412	0.632	16	45	19
+ ReScore	0.412	0.500	**13**	43	14
GraN-DAG	0.294	0.643	16	60	14
+ ReScore	0.353	0.600	**15**	58	15
GES	0.294	0.853	31	54	34
+ ReScore	0.588	0.722	**28**	50	36
CD-NOD	0.588	0.444	15	-	18

ReScore能加强可微CD在真实异质数据Sachs上的性能

[Zhang A, Liu F, Ma W, et al. Boosting Differentiable Causal Discovery via Adaptive Sample Reweighting. In ICLR, 2023.]