Efficient Causal Structure Learning from Multiple Interventional Datasets with Unknown Targets

汇报人：王雲霞

2022-01-08

作者：王雲霞，曹付元，俞奎，梁吉业

报告提纲

一、研究动机
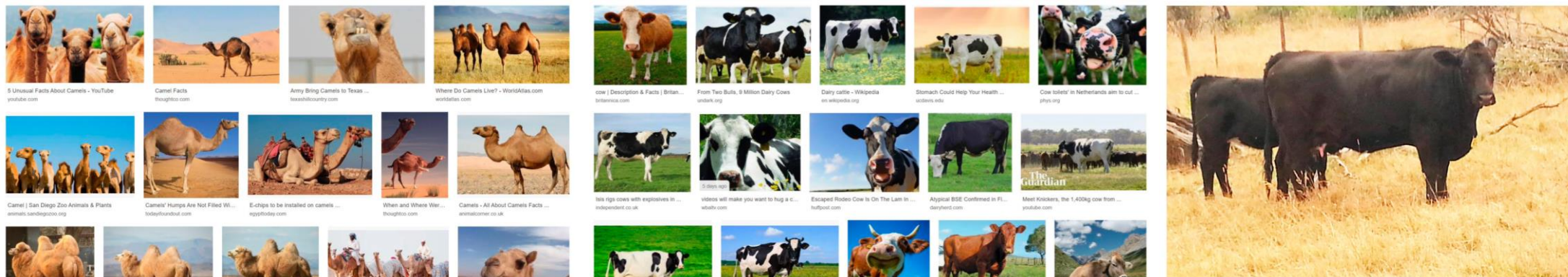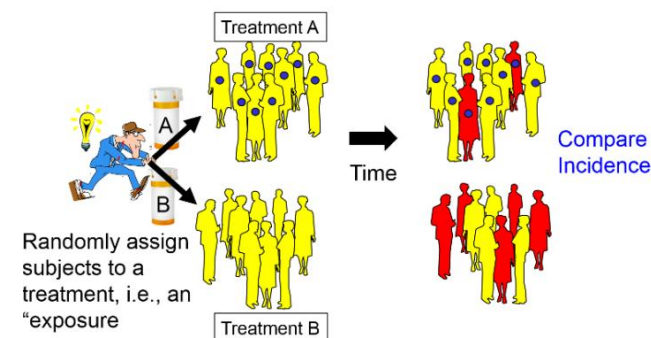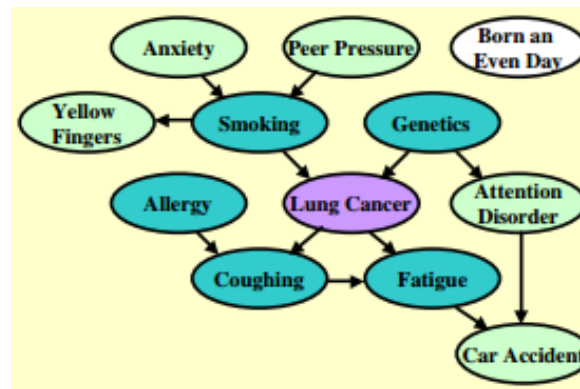
二、研究内容

三、实验分析

四、总结

■ 因果结构学习是什么？

● 发现变量/事物背后的因果关系。

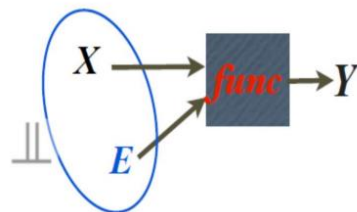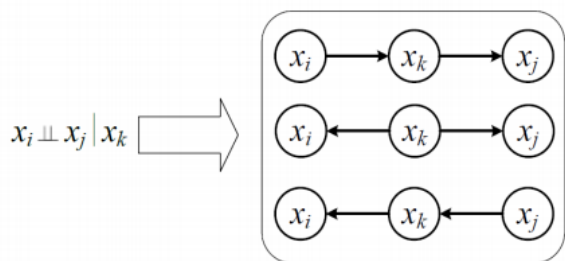■ 为什么研究？

● 揭示事物发生机制；指导干预行为；以及反事实推理。

● 人工智能技术需要进一步提高其鲁棒性和可解释性。

■ 因果结构学习的分类（不同角度）

● 方法角度

✓ **基于约束**： $X_1 \perp\!\!\!\perp X_4 | X_3$
✓ 基于因果结构模型：$X=f_i(PA_i, E_i), i=1,...,n$
✓ 基于打分：G' = arg max $_{G \in DAGs}$ S(D, G)

$$\hat{\mathcal{G}} := \arg\max_{\mathcal{G} \in \mathsf{DAGs}} S(\mathcal{D}, \mathcal{G})$$

$\mathcal{D} = \mathbf{X} \in \mathbb{R}^{n \times d}$ is the dataset.

Many choices of the score function, e.g.,

$$S(\mathcal{D}, \mathcal{G}) = \log p(\mathcal{D}|\hat{\boldsymbol{\theta}}, \mathcal{G}) - \frac{\# \text{ parameters}}{2} \log n.$$

● 数据角度

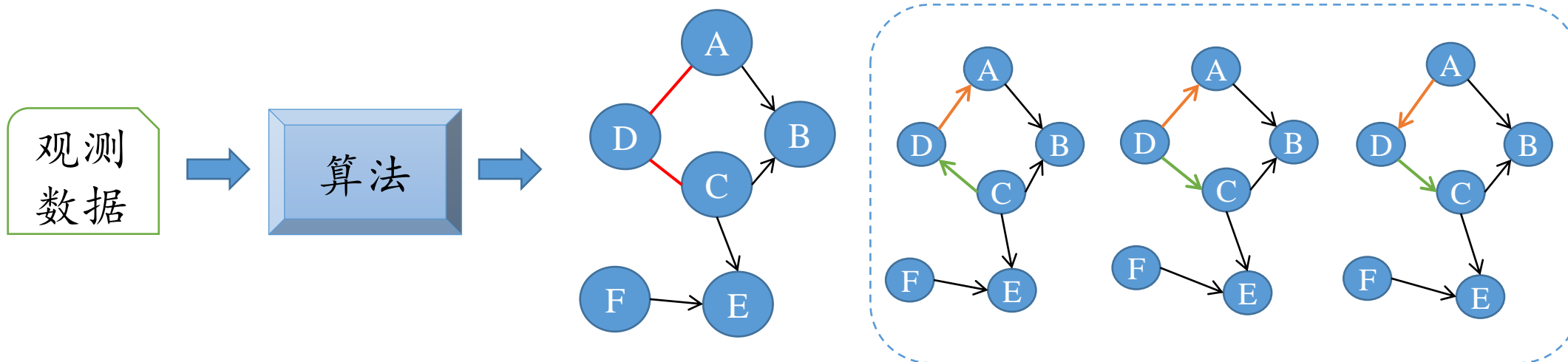✓ 面向单源观测数据的因果结构学习
✓ 面向多源观测数据的因果结构学习
✓ 面向**多源干预**数据的因果结构学习

> 多个不同来源的数据给因果结构带来更丰富的因果信息

> 面向观测数据的方法存在马尔科夫等价类问题，干预可以缓减

■ 马尔科夫等价类

● 马尔科夫等价类：对于两个因果图G和G0，如果他们描述相同的概率分布和条件独立信息，则这两个图是马尔科夫等价的。这样的因果图构成一个马尔科夫等价类。

● V结构：对于三个节点A,B和C，如果B有两个从A和C输入的边可以构成A→B←C，且A和C不相邻，则A,B,C构成一个V结构。

马尔科夫等价类：相同的骨架和V结构

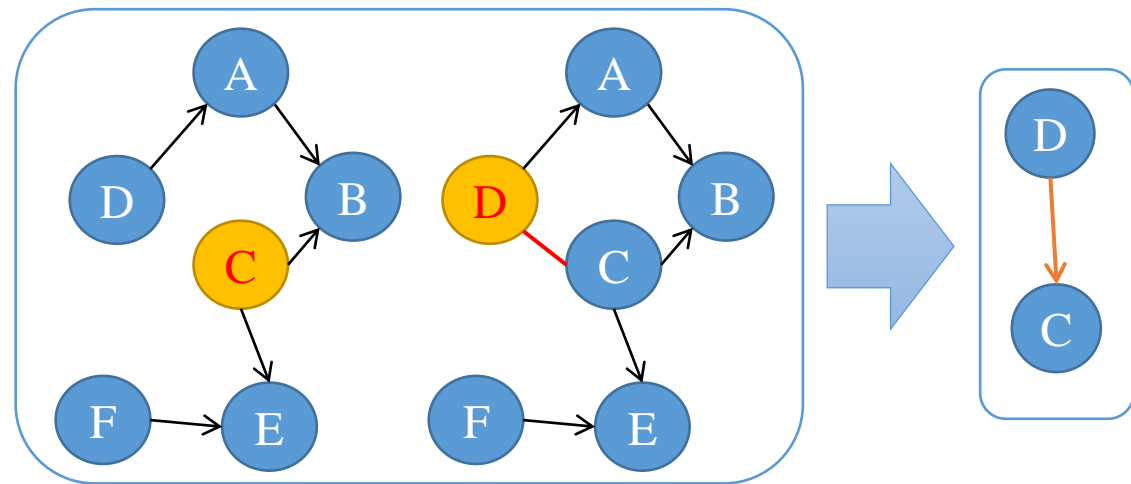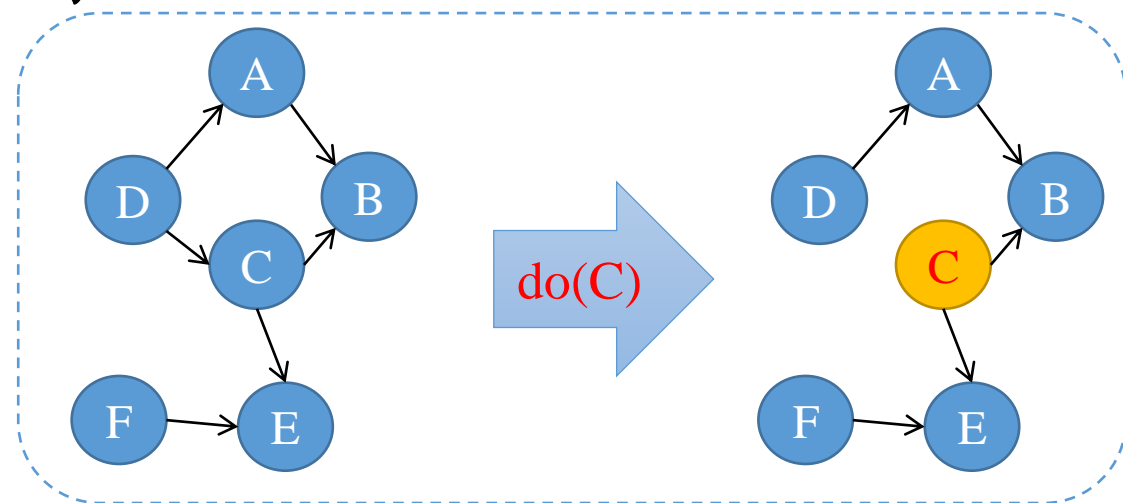■ 面向多源干预数据的因果结构学习

● 干预：
  ✓ 完美干预/硬干预
  ✓ 不完美干预/软干预

● 干预数据集：干预变量得到的数据集（和原始数据集分布不同）

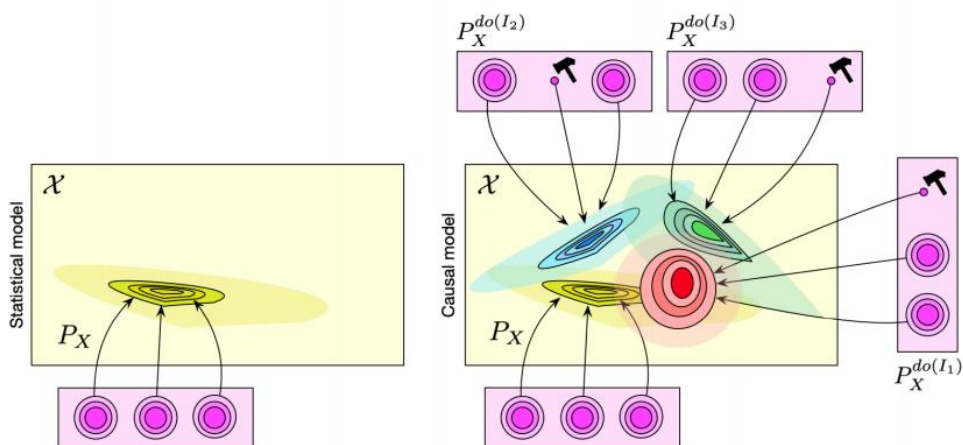● 干预后的图G_i：$G_i=(V,E_i)$，且 $E_i=\{(a,b)|(a,b)\in E, b\notin R_i\}$，其中$R_i$是被干预的变量集合。

■ 挑战

● 未知的干预变量和不相同的数据分布　　● 高维数据 (节点数量较大)

$$P_i(\mathbf{V}|do(R_i)) = \prod_{X_j \in \mathbf{V} \setminus R_i} P(X_j|pa(X_j)) \prod_{X_j \in R_i} P_i(X_j).$$
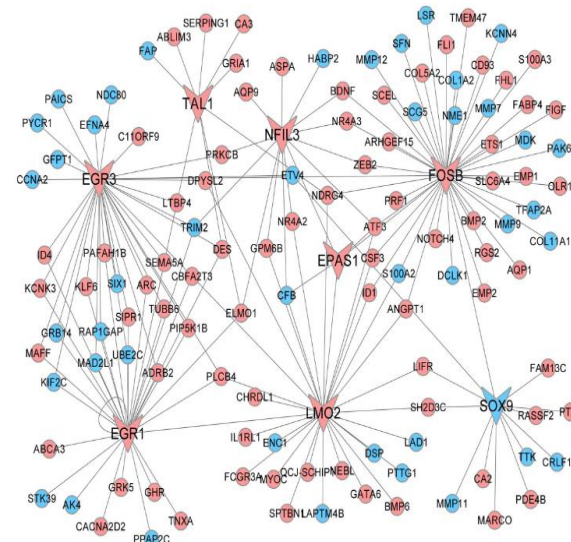
本文关注于降低多个高维且具有未知干预变量的干预数据集中的误发现率。

报告提纲

■ 如何从m个带有未知干预变量的数据集中学习因果结构?
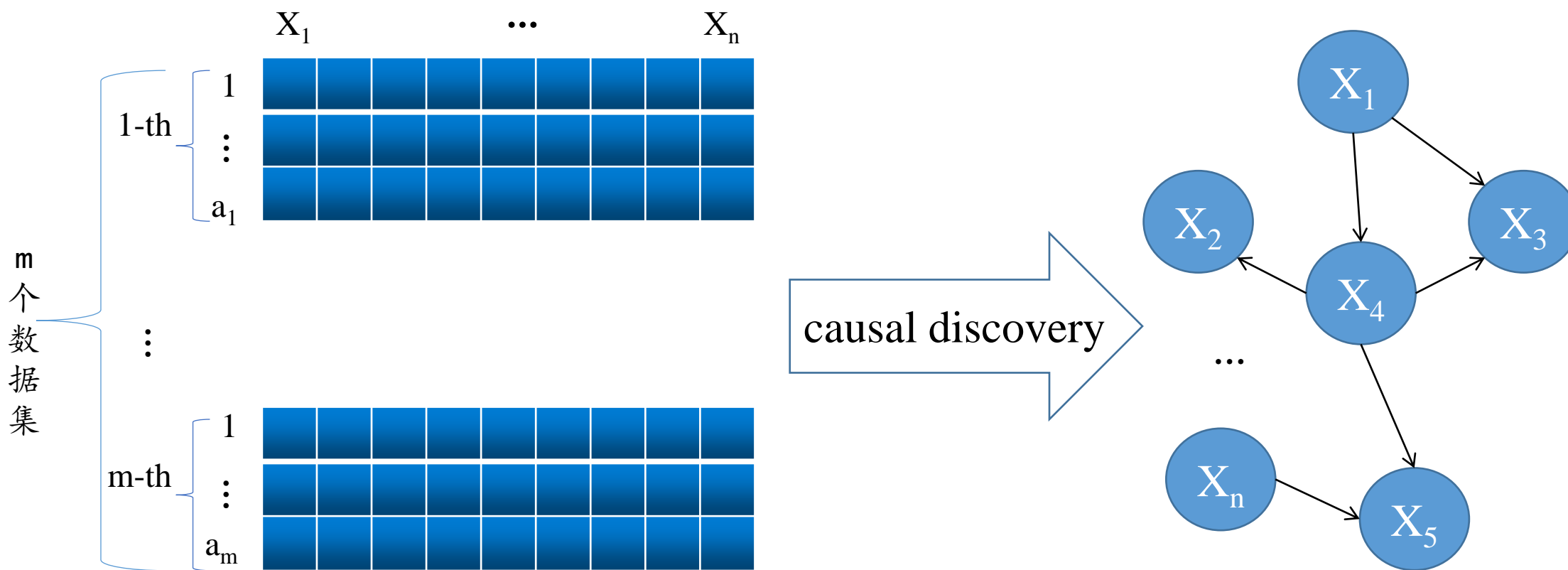
- ## 假设

  - there are no unmeasured confounders

  - P_i is faithful to G_i for any i $\in$ $\{1, \cdots, m\}$

  - 干预变量集R = $\{R_1, \cdots, R_m\}$ is conservative in m interventional datasets

    Definition 4 (Conservative Rule [16]): If $\exists X_j \in \bigcup_{i=1}^{m} R_i$, $\exists R_i \in R$ such that $X_j \notin R_i$, then R is conservative.

■ EMIDGS技术路线

first learned an I-MEC before merging multiple causal DAGs

A skeleton

A smaller I-MEC

graph1

graph2

graphm

The final structure
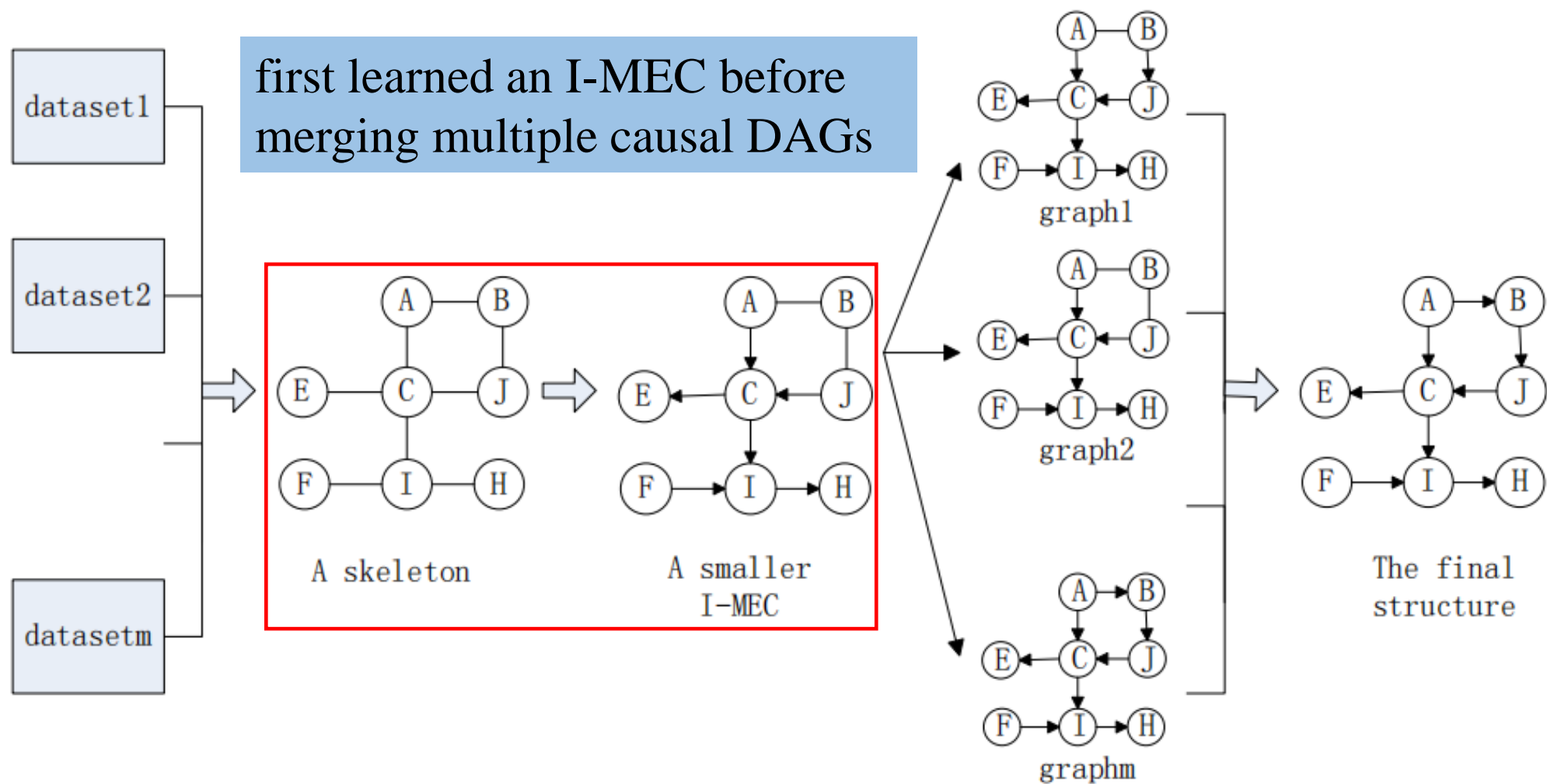
■ EMIDGS技术路线

● step1: **learn a skeleton** by EMIDPC proposed by this paper

● step2: **get an I-MEC** by two Lemmas

● step3: **Search and scoring** based on the learned I-MEC further identifies the remaining unoriented edges.

■ step1： **learn a skeleton** by EMIDPC


```
for X_j ∈ V do
    [pc, sep, kpc, kindep]=EMIDPC(D, X_j, V);
    for Y ∈ pc do
        G(X_j, Y) = 1;
    end for
end for
G_s = G = (V, E_s);
```
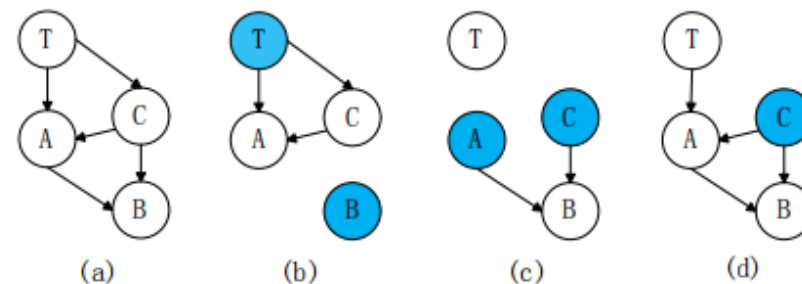

(a)   *(b)*   *(c)*   *(d)*

EMIDPC
(1) Find candidate PC of T
(2) Find PC of T (remove false positives)

Corollary 1 *Referring to Algorithm 2, assuming* $X_j \in cpc(T)$ *and* $\exists S \subseteq \{cpc(T) \setminus X_j\}$, *if* $\exists k \in \{1, \cdots, m\}$ *such that* $X_j \perp\!\!\!\perp T|S$ *and* $\{X_j \cup S\} \subseteq kpc_T(k)$ *in* D_k, *then we know that* $X_j \notin pc(T)$.

Proposition 1 *When R is conservative, the EMIDPC algorithm can learn parents and children of T correctly and efficiently.*

    Proof: *See Appendix B.*

■ step2：get an I-MEC

Lemma 1 *The invariance of V-structures. Suppose R is conservative and there are m intervention datasets without data errors and selection bias and three variables $X, Y, A \in V$. If $\exists S \subseteq V \setminus \{X, Y, A\}$ such that $X \perp\!\!\!\perp Y|S$, $X \not\perp\!\!\!\perp Y|\{S \cup A\}$, and the variables in $\{X, Y\} \cup S$ are dependent on the variable A in the i-th dataset, then X, Y, A forms a V-structure with the collider A, that is $X \to A \leftarrow Y$.*
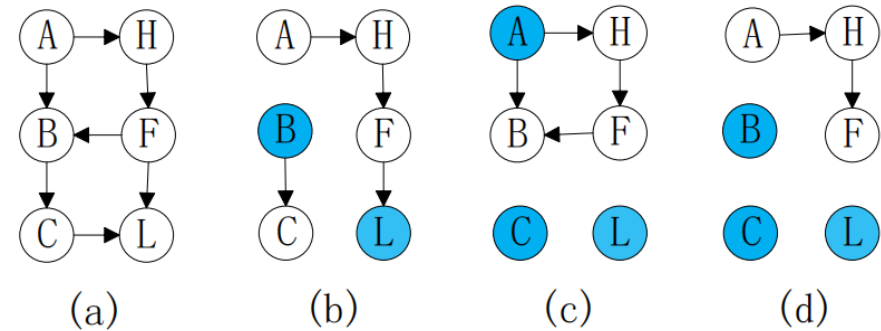
  Proof: *See Appendix A.*

Corollary 2 *Referring to Algorithm 1, if $X \perp\!\!\!\perp Y|S$, and $X \not\perp\!\!\!\perp Y|\{S \cup A\}$, $\exists S \subseteq V \setminus \{X, Y, A\}$, and $\{X, Y, A\} \cup S \subseteq kpc_A(i)$, then X, Y, A forms a V-structure with the collider A, orienting $\langle X, A, Y \rangle$ as $X \to A \leftarrow Y$.*

Corollary 3 *Referring to Algorithm 1, assuming $\langle X, A, Y \rangle$ is identified a V-structure in i-th dataset (D_i) and A is a collider. Then the variables in the set $tem = kindep_A(i) \cap pc(A)$ are children of A.*

Lemma 2 *The unique property of perfect intervention. For m interventional datasets without data errors and selection bias, and $X, Y, A \in V$, assuming a triple $< X, Y, A >$ is determined as a V-structure in the i-th dataset D_i and A is a collider. If $\exists A_1 \in pc(A)$ such that $A \perp\!\!\!\perp A_1$ in D_i, then A_1 is a child of A.*

  Proof: *See Appendix A.*

(a)   (b)   (c)   (d)

Proposition 2 *When R is conservative, the EMIDGS algorithm can discover a theoretically correct I-MEC effectively, which includes more causal relationships.*

  Proof: *See Appendix B.*

■ step3：Search and scoring
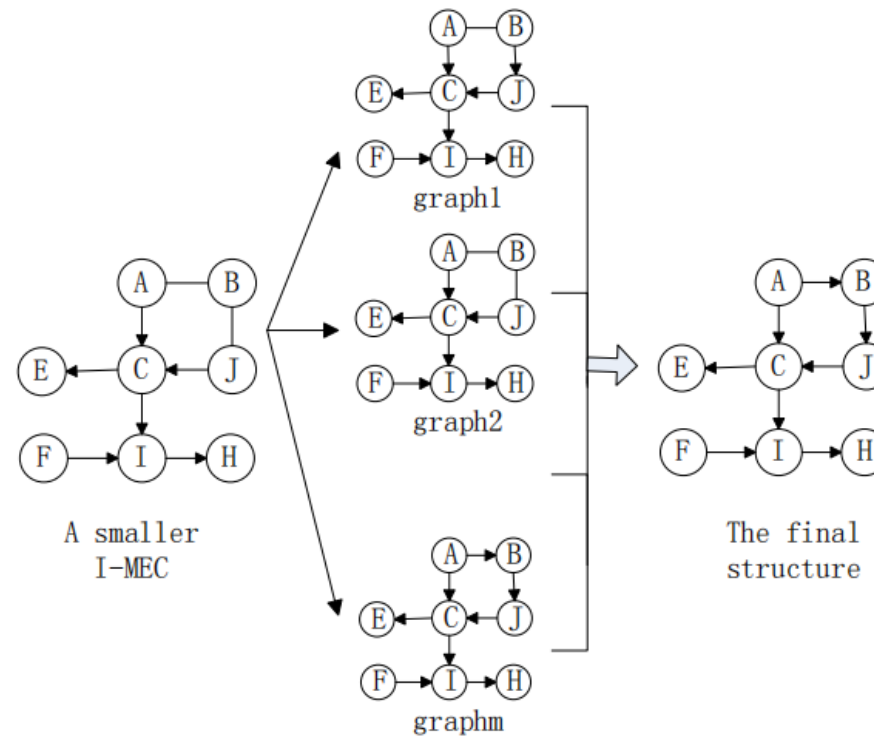
$G_0 = G = (V, E_0)$;

for $i = 1$ to m **do**

    Based G_0, learning a graph $G_f(i) = (V, E_i)$ from the i-th dataset D_i by performing a scoring method on the remaining search space.

end for

Combine $G_f(i)$ to a graph G_f.

如果在某个图$G_f(i)$中确定了e_j的方向，则将有向边e_j添加到最终的图G_f中。如果ej的边在$k(1 \leq k \leq m)$图中存在矛盾，则e_j不添加到G_f中。

Proposition 3 *When R is conservative, the combination of multiple graphs, that is the second step of the orientation phase of EMIDGS, can effectively reduce the search space and improve the efficiency of the EMIDGS algorithm.*

    Proof : *See Appendix B.*

报告提纲

■ 实验设置

Table 1: Description of benchmark BNs.

Net	Nodes	Arcs	Parameters	Average $|MB|$	Type
child	20	25	230	3.00	Medium
insurance	27	52	984	5.19	Medium
alarm	37	46	509	3.51	Medium
win95pts	76	112	574	5.92	Large
munin	186	273	15622	3.81	Very Large
link	724	1125	14211	4.80	Very Large

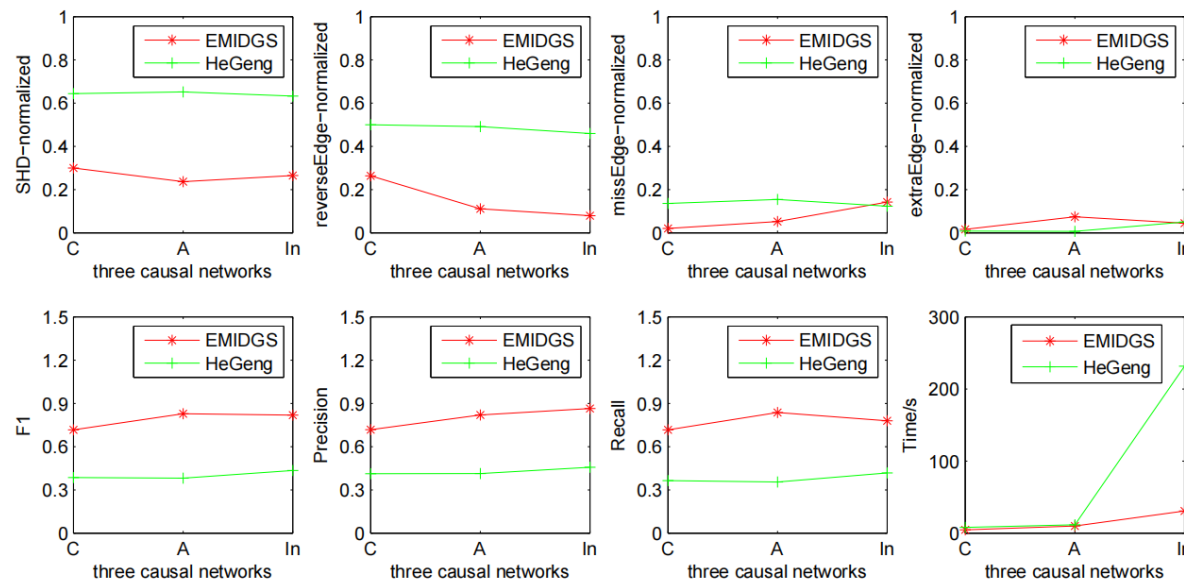three parameters: mD∈{5,10,15,20}, mT∈{2,4,6,8,10}, alpha∈[0.01,0.1]:
- ✓ mT interventional datasets as a group;
- ✓ randomly select q variables to manipulate, q∈[1,mT];
- ✓ different confidence levels alpha.

■ EMIDGS vs. exsiting algorithms
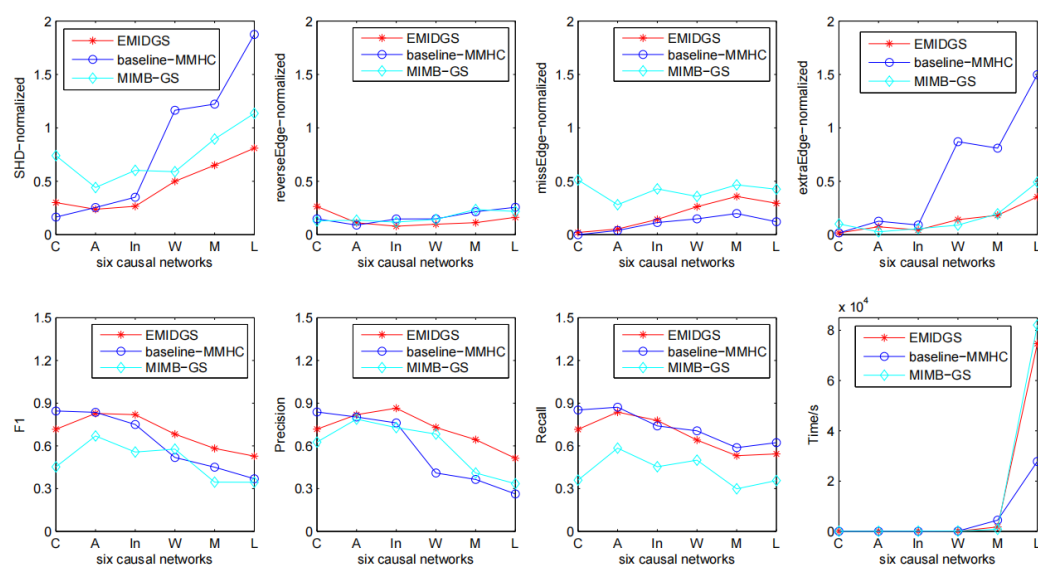
Comparison of performance of EMIDGS, MIMB-GS, and baseline-MMHC algorithms.

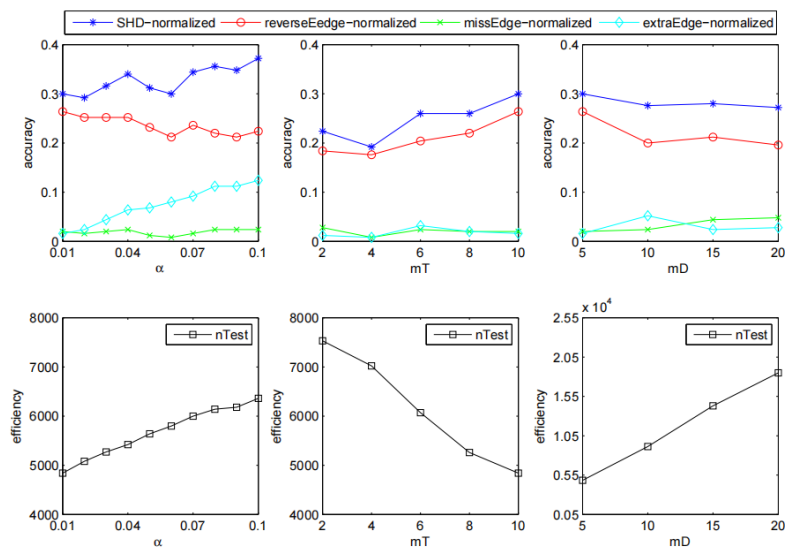Comparison of performance of EMIDGS and the HeGeng algorithms.
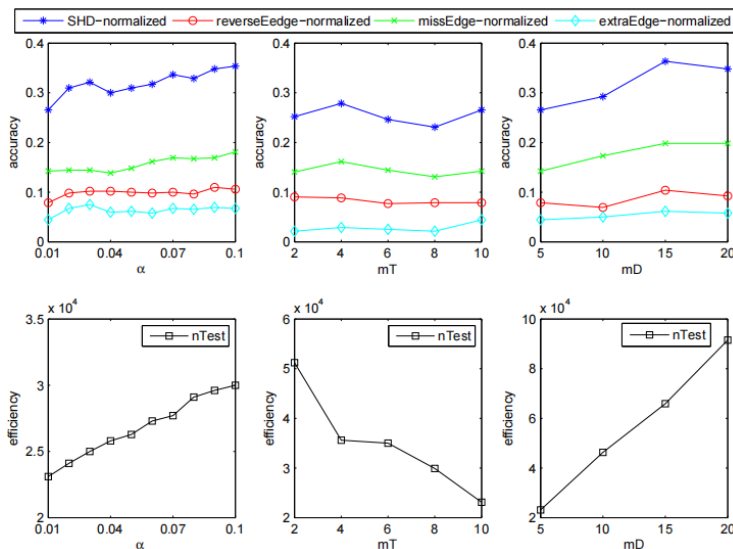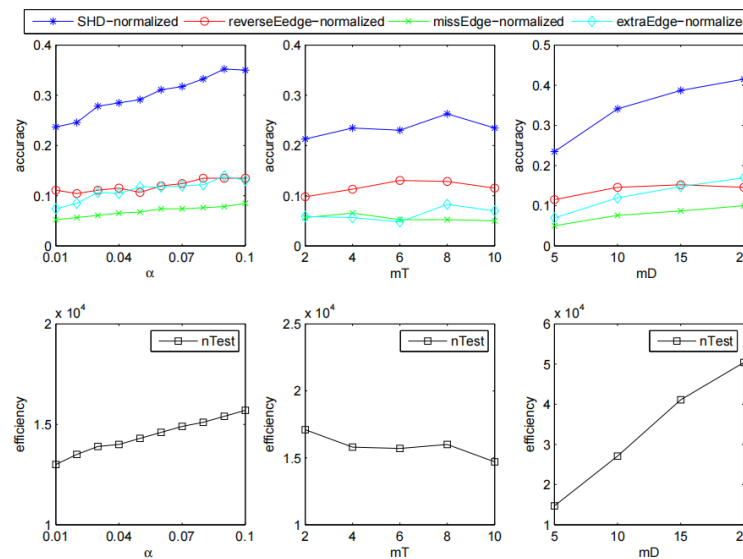
在高维数据中，EMIDGS表现出较明显的优势，其能有效降低误发现率。

■ Different values of three parameters

Comparison of performance of EMIDGS under different values of three parameters on "child".

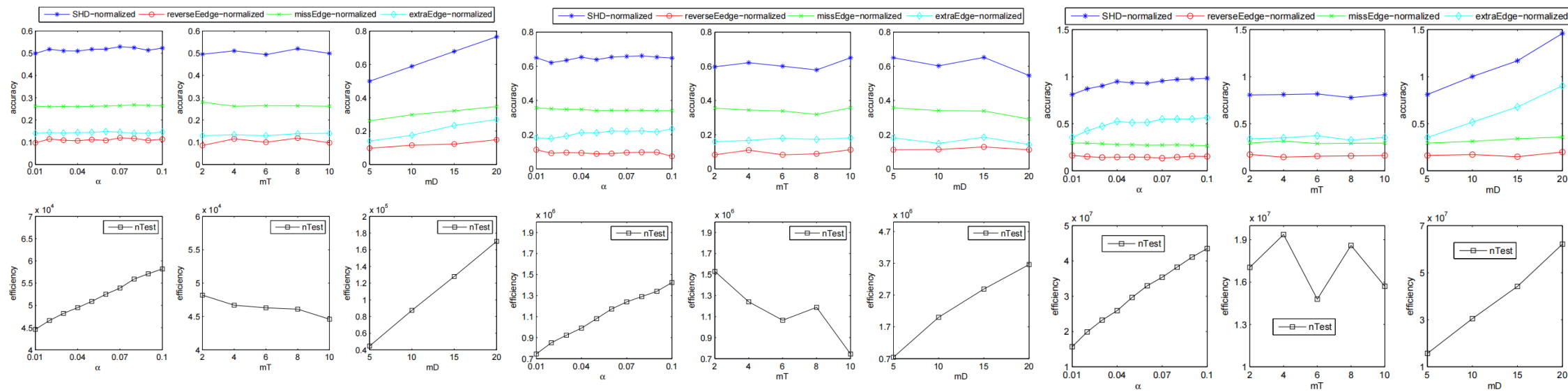Comparison of performance of EMIDGS under different values of three parameters on "insurance".

Comparison of performance of EMIDGS under different values of three parameters on "alarm".

■ Different values of three parameters

Comparison of performance of EMIDGS under different values of three parameters on "win95pts".

Comparison of performance of EMIDGS under different values of three parameters on "munin".

Comparison of performance of EMIDGS under different values of three parameters on "link".

报告提纲

□ 总结

- **EMIGDS算法意义**
  - ✓ 本文提出的局部因果关系发现算法EMIDPC重构因果骨架。
  - ✓ 利用本文提出的两个引理，得到一个包含更多因果信息且理想情况被证明是正确的I-MEC。
  - ✓ 搜索打分有效降低搜索空间。
  - ✓ 实验结果验证该算法有效降低多个高维干预数据集的误发现率。

- **未来工作**
  - ✓ 从具有未知目标且是通用干预的多个干预数据集中学习一个给定目标的局部因果结构。

谢谢指正！