Active Invariant Causal Prediction: Experiment Selection through Stability

1.problem

1. 仅使用观测数据很难恢复causal graph
2. 可识别性只能通过施加intervention或者说进行试验才能得到,但是实验的成本太高了

2.Invariant Causal Prediction

ICP算法从intervention data中恢复label变量Y的直接原因 ,用到的核心思想是 在各个环境(intervention作用在除Y之外的不同的变量之上)中保持不变

ICP最大的优势在于这个方法很少会出现选择的变量是false positive的可能(即明明不是invariant 变量,但缺被判别为invariant 变量)

3.contributions

作者将active learning与ICP进行了结合,框架如下:

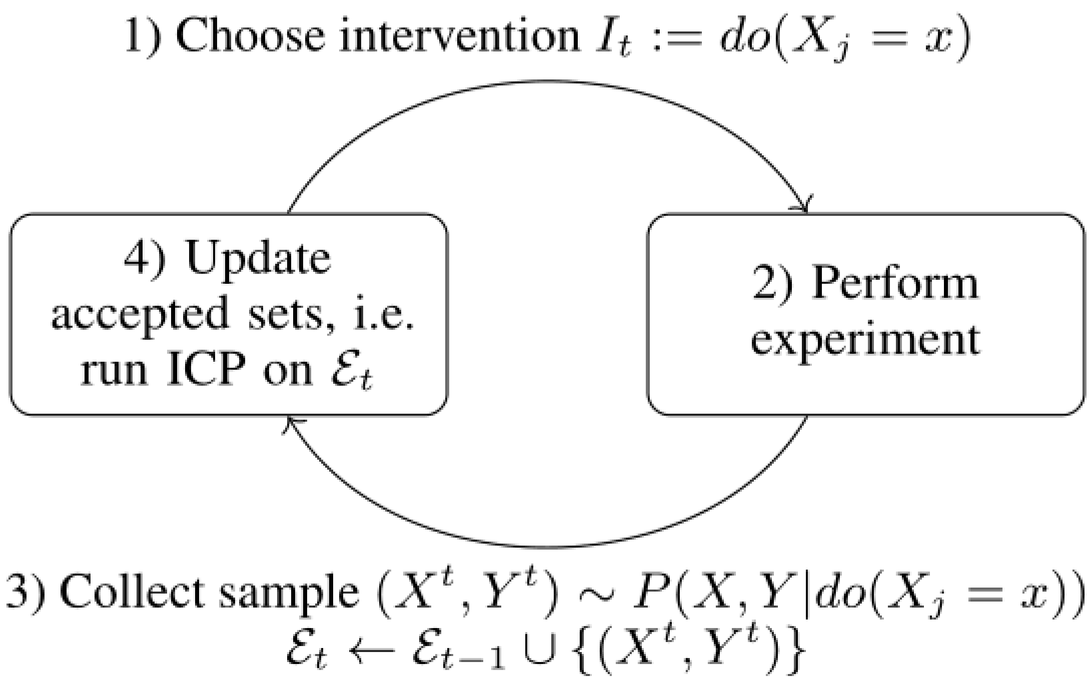


Figure 1: Schematic of A-ICP

主要贡献在于提供了在每一轮循环t中选择intervention的策略.

4.Intervention stable sets

> **设定1**
>
> **Setting 1** (adapted from setting 2 in [30]) Let $X \in \boldsymbol{\mathcal{X}} = \mathcal{X}_1 \times ... \times \mathcal{X}_p$ be predictor variables, $Y \in \mathbb{R}$ a *response* variable and $I = (I_1, ..., I_m) \in \boldsymbol{\mathcal{I}} = \mathcal{I}_1 \times ... \times \mathcal{I}_m$ intervention variables which are unobserved and formalize the interventions present in the collection of intervention environments $\mathcal{E}$. Assume there exists a SCM $\mathcal{S}^{\mathcal{E}}$ over (I, X, Y) that can be represented by a directed acyclic graph $\mathcal{G}(\mathcal{S}^{\mathcal{E}})$, in which the intervention variables are source nodes. Further assume intervention variables do not appear in the structural equation of Y, that is, assume there are no interventions on the response. For each $e \in \mathcal{E}$, there is a SCM $\mathcal{S}_e$ over (I^e, X^e, Y^e) such that $\mathcal{G}(\mathcal{S}_e) = \mathcal{G}(\mathcal{S}^{\mathcal{E}})$, in which only the equations with I^e on the right hand side change with respect to $\mathcal{S}^{\mathcal{E}}$. Furthermore, assume that the distribution of (I^e, X^e, Y^e) is absolutely continuous with respect to a product measure that factorizes.

设定1中将intervention作为变量融入到SCM中,并且不同环境的SCM模型中,图是一致的,即

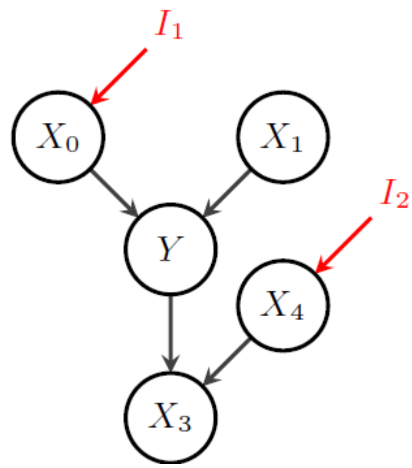
,但是方程可能不同,比如intervention在方程右端时,在各个环境中的作用可能不同

> **definition 2.1**intervention stable set
>
> **Definition 2.1** (intervention stable set [30]). *Let for any set $S \subseteq \{1, ..., p\}$, X_S be the vector containing all variables $X_k, k \in S$. Given setting 1 and a set of environments $\mathcal{E}$, we call a set*
>
> $S \subseteq \{1, ..., p\}$ intervention stable *under $\mathcal{E}$ if the d-separation $I \perp\!\!\!\perp_{\mathcal{G}} Y \mid X_S$ holds in $\mathcal{G}(\mathcal{S}^{\mathcal{E}})$ for any intervention I which is active in an environment $e \in \mathcal{E}$.*

变量集合是stable的当这个集合能让Y与intervention之间d-separate.可以参考下图:

Example A.1. Let $\mathcal{E}$ be a collection of environments with direct interventions on X_0 and X_4, as shown in the graph. Then, the intervention stable sets are

$$\begin{aligned}\mathbb{S}_{\mathcal{E}} =&\{0\},\\ &\{0,4\},\\ &\{0,3,4\}\\ &\{0,1\}\\ &\{0,1,4\}\\ &\{0,1,3,4\}.\end{aligned}$$



使用 表示所有对任意intervention稳定的集合.从定义中可看到在给定intervention stable set中的值之后,在各个intervention的作用下,Y的值都不变(即).

针对这个定义有以下几个推论:

> **推论1**
>
> **Lemma 1** (intervened parents appear on all intervention stable sets). *Let $\mathcal{E}$ be a set of observed environments and let $j \in PA(Y)$ be directly intervened on in $\mathcal{E}$. Then,*
>
> $$S \subseteq \{1, ..., p\} \text{ is intervention stable} \implies j \in S.$$

比如example中的父节点 一致包含在intervention invariant set中.

> **推论2**
>
> **Lemma 2** (sets containing descendants of directly intervened children are unstable). *Let $i \in CH(Y)$ be directly intervened on in $\mathcal{E}$. Then, any set $S \subseteq \{1, ..., p\}$ which contains descendants of i is not intervention stable.*

intervention,Y以及CH(Y)构成了collider结构,那么当CH(Y)被包含在内,则Y与intervention之间就无法d-separate了

> **推论3**
>
> **Lemma 3** (stability of the empty set). *Let $\mathcal{E}$ be any set of environments. Then,*
>
> $$\emptyset \in \mathbb{S}_{\mathcal{E}} \iff \mathcal{E} \text{ contains no interventions on variables in } AN(Y).$$

如果intervention不施加在Y的父节点之上,那么空集包含在intervention stable set之中

> **definition2.2**

> **Definition 2.2** (stability ratio). *Given a set of environments $\mathcal{E}$, the* stability ratio *of a variable $i \in \{1, ..., p\}$ is defined as*
>
> $$r_{\mathcal{E}}(i) := \frac{1}{|\mathbb{S}_{\mathcal{E}}|} \sum_{S \in \mathbb{S}_{\mathcal{E}}} \mathbb{1}\{i \in S\},$$
>
> *i.e. the proportion it appears in the intervention stable sets under $\mathcal{E}$.*

即变量i被包括在intervention stable set中的概率

> **Proposition1**
>
> **Proposition 1** (ancestors appear on at least half of all stable sets). *Let $\mathcal{E}$ be any set of observed environments. Then, for any $j \in \{1, ..., p\}$,*
>
> $$r_{\mathcal{E}}(j) < 1/2 \implies j \notin AN(Y).$$
>
> **Corollary 2.1.** *The parents of the response always have a stability ratio of or above $1/2$.*

结合推论1,当intervention施加在pa(Y)上时,pa(Y)的stability ratio为1,而当intervention全部作用在CH(Y)中时,AN(Y)出现的概率也等于0.5,所以AN(Y)的stability ratio应该大于0.5(但是反过来不一定成立)

5.From stable sets to causal predictors

这一部分作者需要将intervention stable set的内容应用到plausible causal predictor之上.

> **definition3.1**
>
> **Definition 3.1** (plausible causal predictors [27]). *We call a set of variables $S \subseteq \{1, ..., p\}$* plausible causal predictors *under a set of environments $\mathcal{E}$ if for all $e, f \in \mathcal{E}$ and all x*
>
> $$Y^e | X_S^e = x \overset{d}{=} Y^f | X_S^f = x, \tag{1}$$
>
> *i.e. the conditional distribution is the same in all environments. Let $\mathbb{C}_{\mathcal{E}}$ denote the collection of sets which are plausible causal predictors under $\mathcal{E}$.*

即条件分布在各个环境中保持一致.ICP算法的结果就是给出了 的估计值

> **Proposition2**
>
> **Proposition 2** (intervention stable sets are plausible causal predictors). *Let $\mathcal{E}$ be a set of observed environments. Then, for all intervention stable sets $S \subseteq \{1, ..., p\}$, it holds that $S \in \mathbb{C}_{\mathcal{E}}$.*

intervention stable set是plausible causal predictor的一部分,但是作者认为 出现的可能性接近与0,所以就假设

> **推论**
>
> **Corollary 3.1.** *Let $\mathcal{E}_t, \mathcal{E}_{t+1}$ be sets of observed environments such that $\mathcal{E}_t \subseteq \mathcal{E}_{t+1}$. Then, it follows that if S is not a set of plausible causal predictors under $\mathcal{E}_t$, it is not under $\mathcal{E}_{t+1}$ either.*

6.Constructing an active learning policy

ICP方法是非常依赖internvention的位置的,如果没有intervention作用在AN(Y)上面,那么给出的集合完全是一个空集,毕竟空集也可与让Y的预测与intervention之间没关联.

那么作者的方案就是要挑选一些能直接作用在pa(Y)之上的intervention.策略如下:

1. (*Markov strategy*, ("markov")) This strategy selects intervention targets from within the Markov blanket, which contains the parents. Under linearity, in the population setting the Markov blanket can be directly obtained from an ordinary least squares (OLS) regression over all predictors (Appendix E). In the finite regime, we turn to the Lasso [36] to obtain an estimate.
2. (*empty-set strategy*, ("e")) If an observational sample is available, we can test whether the invariance in Eq. (1) holds for the empty set when considering the observational and the interventional sample e_t. If it does, by Lemma 3 we know that the latest intervention target is not upstream of the response, and therefore not a parent. We hence discard the target from future interventions.
3. (*ratio strategy*, ("r")) By Corollary 2.1, a variable is not a parent if it appears on less than half of all intervention stable sets. As an estimate we use the accepted sets (computed based on the environments $\mathcal{E}_{t-1}$) and, if a variable appears on less than half of such sets, we do not add it to the pool of possible intervention targets for the current iteration. Note that unlike in (2.), we do not discard it from future interventions. This is important in the finite regime, where parents may for some iterations appear in less than half of all accepted sets due to testing errors.

其中Markov策略是让intervention作用在Markov blanket之上;empty-set策略测试空集是否是intervention stable set如果是那么先前的一个intervention就没有作用在parent之上,因此要丢弃该intervention;ratio策略是将stable ratio小于0.5的变量暂时移出intervention目标之外.

除此之外,在每一次迭代之中,都将stability为1的变量收集起来,因为stable ratio为1说明这个变量是父节点.所以随着程序的进行intervention target会逐渐减少.

Algorithm 1: A-ICP

Output : $\hat{S}(\mathcal{E}_T)$ estimate of the parents of the response
Input : `policy` an intervention selection policy,
(X^0, Y^0) sample from initial environment,
T number of iterations,
α overall A-ICP significance level

```
𝓔₀ ← {(X⁰, Y⁰)};
accepted sets ← all sets of predictors;
next_intervention ← policy.first_intervention(𝓔₀);
for t = 1 : T do
    perform next_intervention and collect sample (Xᵗ, Yᵗ);
    𝓔ₜ ← 𝓔ₜ₋₁ ∪ {(Xᵗ, Yᵗ)};
    accepted sets, Ŝ(𝓔ₜ) ← ICP(𝓔ₜ, accepted sets, α/T) ;   // see Corollary 3.1
    next_intervention ← policy.next_intervention(accepted sets);
end
return Ŝ(𝓔_T)
```