Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology

SCIENTIFIC REPORTS, JCR(4.997, Q2), JCI(1.05, Q1)

Yiqiao Jin

Supervisor: Bo Fu

Fudan University School of Data Science

2022.11.17

Rough Overview

1. Causal structure discovery algorithms
    - Constraint-based: Fast causal inference(FCI)
    - Score-based: Fast greedy equivalence search(FGES)
2. Database
    - Alzheimer's Disease Neuroimaging Initiative (ADNI)
    - Well-established relationships
    - "Gold-standard" causal graph
3. Study design
    - No knowledge
    - Trivial background knowledge
    - Longitudinal data and trivial background knowledge
4. Guideline
    - Common errors and avoidance
    - Comparisons between the performances of FCI and FGES
5. Conclusions

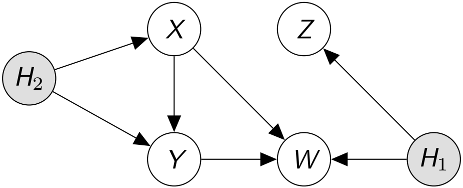
Causal Discovery

Background

The gold standard of causal discovery has typically been to perform planned or randomized experiments (Fisher 1970). There are obvious *practical and ethical considerations* that limit the application of randomized experiments in many instances, particularly on human beings.

What?

To learn *the graphical structure* and recover *the cause-effect pairs* from given *observational* dataset through *computational* methods.



Causal Discovery

Why?

In health science and biomedical applications, discovering *the risk factor or mechanism* that can be altered to *intervene, control, alleviate, and cure diseases* has become the primary research question.

Examples

In a failed Phase 3 trial, amyloid deposition was measured to test the effect of a diabetes drug on reducing AD risk, as an early sign of AD and being associated with diabetes. However, diabetes $\nRightarrow$ amyloidosis.

Mention

- Association $\nRightarrow$ Causation.
- Hypothesis-driven clinical research is based on causal relationship among biomarkers and outcomes.

How?

Combinatoric/search-based approaches, continuous optimization $\cdots$

Definition(Constraint-based algorithm)

Recover the underlying causal structure $\mathcal{G} \in \boldsymbol{\mathcal{G}}$ based on conditional independence constraints in the dataset $\boldsymbol{\mathcal{D}}$ by performing a sequence of statistical hypothesis test.

Definition(Score-based algorithm)

Optimize a properly defined score function $\boldsymbol{\mathcal{S}}(\boldsymbol{\mathcal{G}}, \boldsymbol{\mathcal{D}})$ through a space of possible Causal graphs $\mathcal{G} \in \boldsymbol{\mathcal{G}}$ for the dataset $\boldsymbol{\mathcal{D}}$.

$$\mathcal{G}^* = \arg\max_{\mathcal{G} \in \boldsymbol{\mathcal{G}}} \boldsymbol{\mathcal{S}}(\boldsymbol{\mathcal{G}}, \boldsymbol{\mathcal{D}})$$

Common score functions include the Bayesian Information Criterion(BIC), the Minimum Description Length, and others.

Methodology Selection

In this study, one prominent algorithm was selected from each subtype.

- Fast Causal Inference Algorithm (FCI, constraint-based)
- Fast Greedy Equivalence Search (FGES, score-based)

Similarities and Differences

The table below shows some of the details.

Table: FCI *v.s* FGES

Method	Sufficient	Faithfull	Acyclic	Output
FCI	no	yes	yes	PAG
FGES	yes	yes	yes	CPDAG

Preliminaries

Definition(Sufficiency)

A set of variables $\mathcal{X}$ is causally sufficient if no variable that is a direct cause of two variables in $\mathcal{X}$ is not in $\mathcal{X}$.

Namely, there are no unobserved variables $\mathcal{U}$ in $\mathcal{X}$ that affect the behavior of the causal mechanism generating the dataset $\mathcal{D}$.

Definition(Faithfulness)

If a distribution $\mathcal{P}(\mathcal{X})$ if faithful to the *DAG* $\mathcal{G}(\mathcal{X}, \mathcal{E})$ if

$$\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C} \Longrightarrow \mathcal{A} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{B} \mid \mathcal{C}$$

for all disjoint vertices $\mathcal{A}, \mathcal{B}, \mathcal{C}$.

Preliminaries

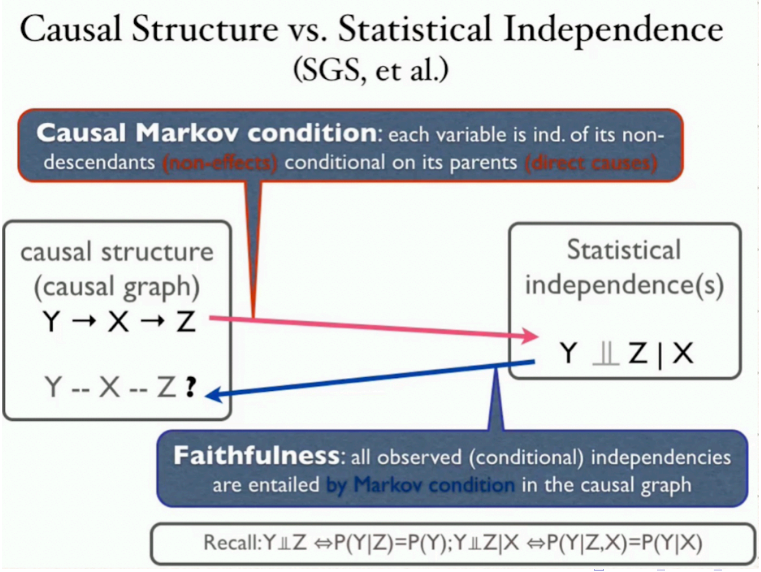
Definition(Markov property)

Given a *DAG* $\mathcal{G}(\mathcal{X}, \mathcal{E})$ and a *joint distribution* $\mathcal{P}(\mathcal{X})$, this distribution is said to satisfy the following conditions

1. the ***Global Markov property*** *with respect to(w.r.t)* the *DAG* $\mathcal{G}$ if
$$\mathcal{A} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{B} \mid \mathcal{C} \Longrightarrow \mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C}$$
for all disjoint vertices $\mathcal{A}, \mathcal{B}, \mathcal{C}$(the symbol $\perp\!\!\!\perp_{\mathcal{G}}$ denotes *d-separation*),
2. the ***Local Markov property*** *w.r.t* the *DAG* $\mathcal{G}$ if each variable is independent of its non-descendants given its parents, and
3. the ***Markov factorization property*** *w.r.t* the *DAG* $\mathcal{G}$ if
$$\mathcal{P}(\mathcal{X}) = \mathcal{P}(X_1, \cdots, X_d) = \prod_{j=1}^{d} \mathcal{P}(X_j \mid \boldsymbol{PA}_j^{\mathcal{G}})$$

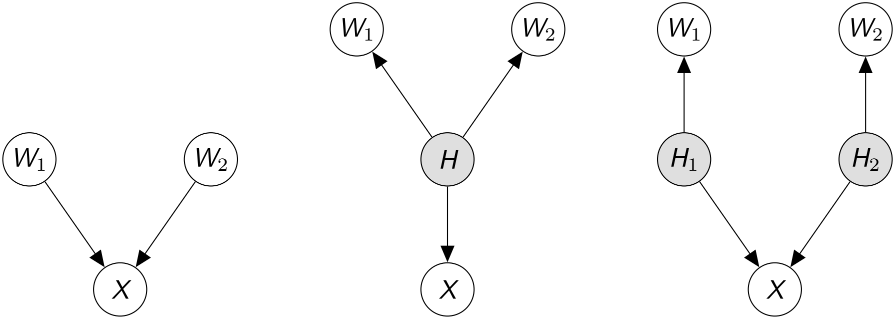
Three definitions are equivalent as long as $\mathcal{P}(\mathcal{X})$ has a density.

Preliminaries



If *DAG* M is insufficient...

DAGs **with the same (in)dependence relationship shows below**



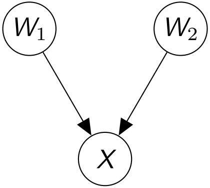
(a) M1 **(b)** M2 **(c)** M3

Where variables $\mathcal{H}, \mathcal{H}_1, \mathcal{H}_2$ denote *hidden/unobserved* confounders.

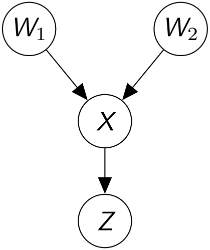
Definition(*Y structure*)

Given a *DAG* $\mathcal{G}$, it is said to satisfy The arc from $\mathcal{X}$ to $\mathcal{Z}$ represents an unconfounded causal relationships

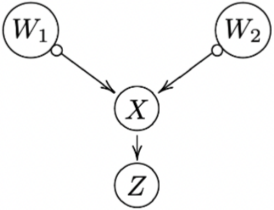
1. $W_1 \rightarrow X \leftarrow W_2$ is a *V structure*,
2. An arc from X to Z,
3. No arcs from W_1 to Z and W_2 to Z.



(d) *X structure*



(e) *Y structure*



(f) Output of *Y structure*

The conditional independency comes from...

1. Input: Conditional independence as prior knowledge.
2. Procedure: Conditional independence test as d-separation set.

Definition(Conditional independence test)

A conditional independence test is defined as

$$H_0 : \mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C} \ \textbf{\textit{v.s}}\ H_1 : \mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C}$$

the null hypothesis H_0 is not rejected iff

$$\hat{C}I(\mathcal{A}, \mathcal{B} \mid \mathcal{C}) > \alpha \xRightarrow{\textit{Ci test}} \mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C} \xRightarrow{\textit{Faithfulness}} \mathcal{A} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{B} \mid \mathcal{C}$$

where α represents the designated significance level.

The main limitation of this approach is related to the exponential growth of the conditioning set $\mathcal{C}$, leading to $2^{\#\{\mathcal{X} \setminus \mathcal{A}, \mathcal{B}\}}$ if unconditionally dependent.

Fast causal inference(FCI)

1. Phase 1
   - Start with a fully connected undirected graph,
   - Remove edges that connect conditional independency,
2. Phase 2
   - Orient edges by identifying the V and Y structures,
   - Orient the remaining edges based on a set of rules.

Fast greedy equivalence search(FGES)

1. Phase 1
   - A. Start with a graph with no edges(independent),
   - B. At each step(until the score can no longer be improved)
     - Adds edges(dependent) with orientation minimizing the BIC,
     - Removes edges decreasing the BIC,
2. Phase 2
   - Remove edges most improving the score until no further edges can thus be removed.

The output of FCI/FGES

		Present Relationships	Absent Relationships
1	A ⟶ B	A is the cause of B	B is not cause of A
2	A —— B	A is the cause of B or B is the cause of A	
3	A ⟷ B	This an unmeasured confounder of A and B.	A is not a cause of B B is not a cause of A
4	Ao⟶ B	Either A is a cause of B or there is an unmeasured confounder of A and B	B is not a cause of A
5	Ao——o B	Exactly one of the following holds: 1. A is a cause of B 2. B is a cause of A 3. There is an unmeasured confounder of A and B 4. Both 1 and 3 5. Both 2 and 3	

The output graph of SEM contains edge type 1; FGES contains edge types 1,2; FCI contains edge types 1,3,4,5.

Alzheimer's Disease Neuroimaging Initiative (ADNI)

	Label
Demographic variables	
AGE	AGE
SEX	SEX
Education Level	EDU
Biomarkers	
Fludeoxyglucose PET	FDU
Amyloid Beta	ABETA
Phophorylated tau	PTAU
Genetics	
APOE epsilon 4 allele	APOE4
Diagnosis	
Diagnosis of AD	DX

Datasets from ADNI

1. **Single cross-sectional**
   - The baseline visit
2. **Longitudinal**
   - Two cross-sections
   - The baseline visit, the visit at the 24 months

Missing data was removed

At least one complete record:

1008 participants

Regular two-year follow-up visit:

266 participants

1. **Level1: No knowledge**
   - No edges are prohibited
2. **Level2: Trivial background knowledge**
   - Edges from demographic v. are prohibited
   - Edges from biomarkers or diagnosis to demographic v. are prohibited
3. **Level3: Longitudinal data and trivial background knowledge**
   - Edges prohibited in Level2
   - Edges pointing from later time point to early time point are prohibited

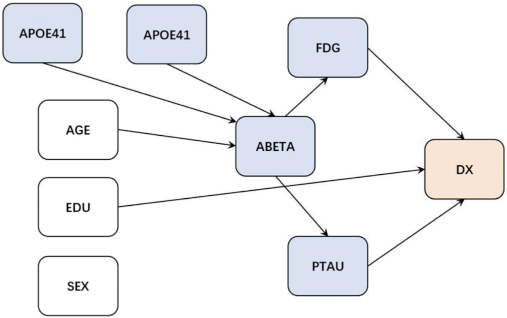


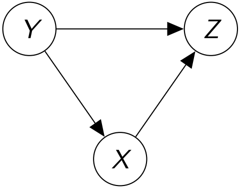
Figure: Gold-standard graph

Evaluation Metrics

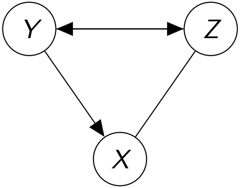
Judgement on the edges(algorithm to gold-standard graph)

- Correct, iff existent edge and the same orientation.
- Semi-correct, iff existent edge and uncontradicted orientation.
- Incorrect, iff nonexistent edge or existent but opposite orientation.

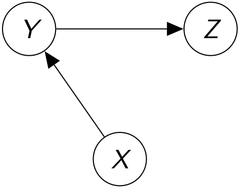
Examples



(a) Gold-standard Graph



(b) Correct for X-Y
Semi-correct for X-Z
Semi-correct for Y-Z



(c) Incorrect for X-Y
Incorrect for X-Z
Correct for Y-Z

Evaluation Metrics

Number

Number of correct, semi-correct, incorrect edges.

Precision

The proportion of (semi-)correct edges over all edges reported.

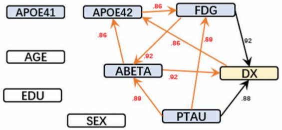
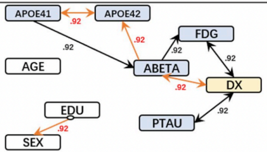
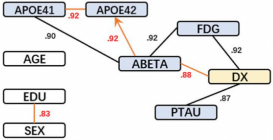
Recall

The proportion of edges in the gold-standard graph (semi-)correctly reported.

Occurrence

The percentage of the adjacency in the result of the 100 bootstrap runs.

No Knowledge

		Precision	Recall	Correct/semi-correct/Incorrect
SEM		0.24	0.30	2.4/0/7.6
FCI		0.24	0.44	1.2/2.4/11.6
FGES		0.46	0.60	1.6/3.2/5.8

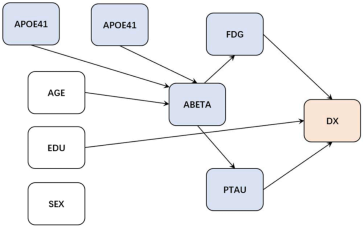
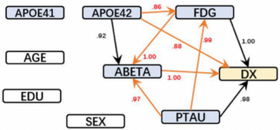
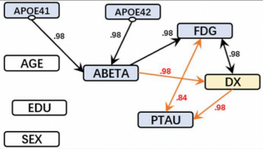
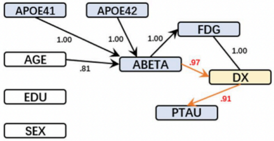


Figure: Gold-standard graph

The edges presented with at least 80 % occurrence rate in the 100 bootstraps samples.

Trivial Background Knowledge

	Precision	Recall	Correct/semi-correct/Incorrect
SEM 	0.31	0.39	3.1/0/6.9
FCI 	0.42	0.55	1.1/3.3/6.5
FGES 	0.71	0.68	4.2/1.2/2.3

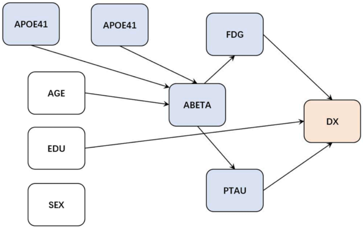
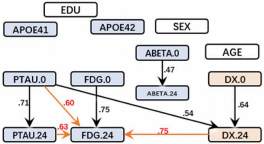
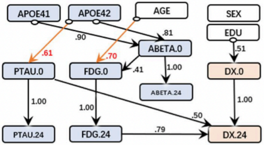
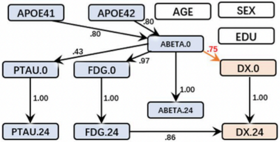


Figure: Gold-standard graph

Significant improvements when trivial background information was added.

Longitudinal and Trivial Background Knowledge

		Precision	Recall	Correct/semi-correct/Incorrect
SEM		0.53	0.36	6.4/0/5.6
FCI		0.56	0.68	5.6/2.7/6.6
FGES		0.76	0.74	9.3/0/3

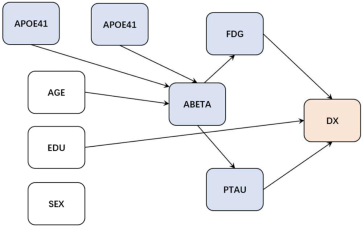


Figure: Gold-standard graph

Significant improvements relative to previous results and FGES recovered a graph with only one incorrect edge.

From algorithm

1. SEM
    - Subject to the assumption of markov property.
    - Suggested modifications are generally not reliable.
2. FGES
    - Subject to the assumption of markov property.
    - Consider a broader array of graphs.
    - Consider the likelihood of the global structure while making local decisions.
    - Computationally expensive to enumerate (and score) every possible graph.
3. FCI
    - Assume that possibly unobserved confounders exist.
    - Subject to sample size.
    - Affected by the incorrect independence tests.
    - Propagate mistakes to large portions of the graph.
4. FGES and FCI outperformed SEM across all three degrees.
5. FGES was better and more stable than FCI.

From experiments

1. Constraint-based algorithm would be a difficult task if conditional independence is unknown(lack of background knowledge).
2. Faithfulness is a strong assumption(probability independence implies graphical separation), and large sample sizes are required to get good conditional independence tests through faithfulness.
3. Add trivial knowledge can eliminate the selection bias to some extent.
4. Increase samples are required to ascertain the validity of conditional independence for constraint-based algorithms.
5. Longitudinal data plays a great role in replenishing background knowledge, and preventing algorithms from making potential mistakes.

Thank you!