Hybrid causal search in latent variable models

Juan Miguel Ogarrio
jogarrio@andrew.cmu.edu

Peter Spirtes
ps7z@andrew.cmu.edu

Joseph Ramsey
jdramsey@andrew.cmu.edu

Department of Philosophy
Carnegie Mellon University
DAG

A -> B -> X -> C

D
Causal search limitations

- Few or no latents
- Strong model assumptions (e.g. linearity, n-factor models)
- Inaccurate with small samples
Outline

- Preliminaries
  - Assumptions
  - Patterns and PAGs
  - GES, FCI
- Greedy Fast Causal Inference
- Simulations
Assumptions

- Partition into observed and latent variables
  - Can only operate on observed variables
- i.i.d. sample, no selection bias (can be relaxed)
Assumptions

- Local Causal Markov assumption
  - d-separation $\rightarrow$ conditional independence

- Causal Faithfulness assumption
  - d-separation $\leftarrow$ conditional independence
Underdetermination

- True model contains no latents
  - Pattern - Markov Equivalence class
- True model may contain latents
  - Partial Ancestral Graph (PAG) - Observational ME class
Patterns - examples
Patterns - examples
PAG example
O-equivalent DAGs
State of the art algorithms: Score-based
GES (Greedy Equivalence Search)

- Score-based
  - Bayesian Information Criterion (BIC) Score
- Outputs Markov and minimal pattern
- Fast Greedy Search (FGS)
  - Optimized version
GES (Greedy Equivalence Search)

- Efficient traversal of search space
  - Two phases
    - Forward phase
      - Single directed edge additions
      - Output Markov
GES (Greedy Equivalence Search)

- Efficient traversal of search space
  - Two phases
    - Forward phase
    - Backward phase
      - Single directed edge removals
      - Markov input, Markov and minimal output
GES (Greedy Equivalence Search)

● **Pros**
  ○ Fast
  ○ Accurate
  ○ No latents → Markov and faithful pattern

● **Cons**
  ○ Latents → Markov and minimal pattern
  ○ BIC score limited to some distributions
State of the art algorithms: Constraint-based
FCI (Fast Causal Inference)

- Constraint-based algorithm
  - Can use any conditional independence test
- Outputs Markov and faithful PAG
  - Can account for latents and selection bias
- Many variants/modifications
  - RFCI, FCI+
Fast Causal Inference (FCI)

● Overview
  ➢ Initial adjacency phase
    → Pre-orientation phase
    → Final adjacency phase
    → Final orientation phase
    → PAG
Fast Causal Inference (FCI)

● Pros
  ○ Can use any independence test
  ○ Accounts for latents

● Cons
  ○ Inaccurate in practice
  ○ Bad worst-case performance
  ○ Non-parametric independence tests are slow
Improvement: pre-process data to improve initial stages of FCI
GFCI (Greedy FCI)

- Hybrid, same assumptions as FCI
- Overview
  - Preprocess data using GES
  - Feed adjacency and unshielded triples to FCI
  - Proceed with FCI
  - Output PAG
GFCI - GES preprocessing

- GES output:
  - Markov and minimal
  - No faithfulness assumption
  - Contains superset of adjacencies
  - Unshielded triples reflect triples in PAG
GFCI - FCI stage

- Remove additional adjacencies
- Copy unshielded triples
- Finalize orientations
GFCI example - DAG
GFCl example - true PAG
GFCI example - GES phase

A  D

B  C
GFCl example - GES phase

A

B

D

C
GFCI example - GES phase
GFCI example - GES phase

A --> B

D --> C
GFCI example - GES phase

Diagram:

- A connected to B
- D connected to C
GFCl example - GES phase
GFCI example - GES phase

A
↓
B
↓
C
→
D
GFCI example - GES phase

\[ B \perp D \mid C \]
GFCI example - GES phase

Diagram:

- A
- B
- C
- D

Connections:
- A to B
- B to C
- C to D
- D to A
GFCI example - FCI phase

A ----> B
  ^    |
  |    v
D ----> C

B ----> D
  ^    |
  |    v
C ----> C
GFCI example - FCI phase

A -> B
B -> C
C -> D
D -> A
GFCI example - FCI phase
GFCI example - FCI phase
GFCI example - FCI phase

A -> B
B -> C
C -> D
D -> A
GFCI example - FCI phase
Simulations
Experiments

- Compare accuracy/speed
- FCI, RFCI, FCI+, GFCI
- Linear Gaussian models
Experiments - parameters

- Graphs
  - DAG sizes: 100, 1000 nodes
  - DAG connectivity: 100, 200 edges / 1000, 2000 edges
  - Latents: 5, 20 latents / 50, 200 latents
- Linear Gaussian models
- Fisher’s Z-test with $\alpha$ : \{0.01, 0.05, 0.1\}
- Around 100 trials per parametrization
Experiments - parameters

- Algorithms
  - FCI
  - RFCI
  - FCI+
  - uRFCI (unbounded RFCI)
  - GFCI (penalty = 4)
Experiments - results

- **Accuracy**
  - GFCI substantially better
  - All algorithms struggle finding ↔ adjacencies

- **Speed**
  - GFCI not as fast, but no practical difference
    - Can be sped up
    - Scaling?
Questions/Comments

jogarrio@andrew.cmu.edu
ps7z@andrew.cmu.edu
jdramsey@andrew.cmu.edu

Code available as part of the Tetrad program

http://www.phil.cmu.edu/tetrad/
https://github.com/cmu-phil/tetrad
Additional slides
Patterns

- Same nodes as DAG
- Directed edges
  - Shared by every DAG
- Undirected edges
  - Distinct for at least two DAGs
Patterns - characterization

- DAG’s pattern/ME class determined by:
  - adjacencies
  - unshielded colliders
PAG

- Nodes same as O nodes in DAG
- Edges
  - Nodes cannot be d-separated
- Edge marks
  - Arrow: non-ancestor in every DAG
  - Tail: ancestor in every DAG
  - Circle: ancestor in some, non-ancestor in some
Edge types

- Directed
- Semi-directed
- Bi-directed
- Undirected
GFCI - Triangle conjecture

- Optimization
  - Additional adjacencies found inside of triangles
Triangle conjecture
GFCI example - DAG
GFCl example - true PAG
GFCI example - intermediate step
GFCI example - intermediate step

\[
Y \perp B \mid \{C, X\} \\
Z \mid B \mid \{C, X, Y\}
\]
GFCI example - additional edges
GFCI example - other parameters
Experiments - parameters

- Graphs
  - DAG sizes: 100, 1000 nodes
  - DAG connectivity: 100, 200 edges / 1000, 2000 edges
  - Latents: 5, 20 latents / 50, 200 latents
Experiments - parameters

- Models
  - Gaussian variables
    - Mean = 0
    - Variance : [1, 3]
  - Edges
    - Coefficients: ±[0.2, 1.5]
Experiments - parameters

- Independence test
  - Fisher’s Z-test
    - $\alpha : \{0.01, 0.05, 0.1\}$
- Random samples
  - Size : 200, 2000
- Around 100 trials
Experiments - accuracy overview
Experiments - result overview
Experiments - time

100 variables, time in msec.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Med.</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFCI</td>
<td>18</td>
<td>90</td>
<td>134</td>
<td>266</td>
<td>1002</td>
</tr>
<tr>
<td>RFCI</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>14</td>
<td>52</td>
</tr>
<tr>
<td>uRFCI</td>
<td>136</td>
<td>332</td>
<td>392</td>
<td>493</td>
<td>1568</td>
</tr>
<tr>
<td>FCI+</td>
<td>14</td>
<td>49</td>
<td>91</td>
<td>145</td>
<td>381</td>
</tr>
</tbody>
</table>
## Experiments - time

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Med.</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFCI</td>
<td>2.330</td>
<td>4.700</td>
<td>6.390</td>
<td>12.760</td>
<td>29.010</td>
</tr>
<tr>
<td>RFCI</td>
<td>0.150</td>
<td>0.798</td>
<td>3.540</td>
<td>6.942</td>
<td>62.900</td>
</tr>
<tr>
<td>uRFCI</td>
<td>1.300</td>
<td>2.610</td>
<td>6.840</td>
<td>16.390</td>
<td>18530.000</td>
</tr>
<tr>
<td>FCI+</td>
<td>1.770</td>
<td>4.500</td>
<td>8.795</td>
<td>13.570</td>
<td>83.140</td>
</tr>
</tbody>
</table>