Causal Discovery from Subsampled Time Series Data by Constraint Optimization

Antti Hyttinen, Sergey Plis, Matti Järvisalo, Frederick Eberhardt, David Danks

University of Helsinki, Helsinki Institute for Information Technology
Mind Research Network and University of New Mexico
California Institute of Technology
Carnegie Mellon University

PGM2016, Lugano
6.9.2016
We consider the discovery of the time series causal structure from data obtained at a coarser \textit{measurement timescale}:

\[ \cdots \ V_1^{t-4} \ V_1^{t-2} \ V_1^t \ \cdots \ \cdots \ \vdots \ \cdots \ \cdots \ \cdots \ V_1^{t-1} \ V_1^t \ \cdots \]

\[ \cdots \ V_2^{t-4} \ V_2^{t-2} \ V_2^t \ \cdots \ \rightarrow \ \cdots \ V_2^{t-1} \ V_2^t \ \cdots \]

\[ \cdots \ V_3^{t-4} \ V_3^{t-2} \ V_3^t \ \cdots \ \cdots \ \vdots \ \cdots \ \cdots \ \cdots \ V_3^{t-1} \ V_3^t \ \cdots \]

- Only every \( u \)-th vector of values is observed (\textit{subsampling rate} \( u \))
- Subsampling induces confounding, and unidentifiability
- Ignoring subsampling can lead to significant errors!

Applications: e.g. fMRI.
We consider the discovery of the time series causal structure from data obtained at a coarser **measurement timescale**:

\[
\ldots \ V_1^{t-4} \ V_1^{t-2} \ V_1^t \ \ldots \ \rightarrow \ \ldots \ V_1^{t-1} \ V_1^t \ \ldots
\]

\[
\ldots \ V_2^{t-4} \ V_2^{t-2} \ V_2^t \ \ldots \ \rightarrow \ \ldots \ V_2^{t-1} \ V_2^t \ \ldots
\]

\[
\ldots \ V_3^{t-4} \ V_3^{t-2} \ V_3^t \ \ldots \ \rightarrow \ \ldots \ V_3^{t-1} \ V_3^t \ \ldots
\]

- Only every \( u \):th vector of values is observed (**subsampling rate** \( u \))
We consider the discovery of the time series causal structure from data obtained at a coarser \textit{measurement timescale}:

\[
\ldots \ V_1^{t-4} \ V_1^{t-2} \ V_1^t \ \ldots \quad \ldots \quad \ V_1^{t-1} \ V_1^t \ \ldots \\
\ldots \ V_2^{t-4} \ V_2^{t-2} \ V_2^t \ \ldots \quad \rightarrow \quad \ldots \quad \ V_2^{t-1} \ V_2^t \ \ldots \\
\ldots \ V_3^{t-4} \ V_3^{t-2} \ V_3^t \ \ldots \quad \ldots \quad \ V_3^{t-1} \ V_3^t \ \ldots 
\]

- Only every \( u \)-th vector of values is observed (\textit{subsampling rate} \( u \))
- Subsampling induces confounding, and unidentifiability
We consider the discovery of the time series causal structure from data obtained at a coarser measurement timescale:

\[ \cdots \; V_{1}^{t-4} \; V_{1}^{t-2} \; V_{1}^{t} \; \cdots \; \rightarrow \cdots \; \]

\[ \cdots \; V_{2}^{t-4} \; V_{2}^{t-2} \; V_{2}^{t} \; \cdots \; \rightarrow \cdots \; \]

\[ \cdots \; V_{3}^{t-4} \; V_{3}^{t-2} \; V_{3}^{t} \; \cdots \; \rightarrow \cdots \; \]

- Only every \( u \):th vector of values is observed (subsampling rate \( u \))
- Subsamping induces confounding, and unidentifiability
- Ignoring subsampling can lead to significant errors!
Introduction to Subsampling

We consider the discovery of the time series causal structure from data obtained at a coarser measurement timescale:

\[
\begin{align*}
\cdots & V_{1}^{t-4} & V_{1}^{t-2} & V_{1}^{t} & \cdots & \cdots & V_{1}^{t-1} & V_{1}^{t} & \cdots \\
\cdots & V_{2}^{t-4} & V_{2}^{t-2} & V_{2}^{t} & \cdots \rightarrow & \cdots & V_{2}^{t-1} & V_{2}^{t} & \cdots \\
\cdots & V_{3}^{t-4} & V_{3}^{t-2} & V_{3}^{t} & \cdots & \cdots & V_{3}^{t-1} & V_{3}^{t} & \cdots 
\end{align*}
\]

- Only every \( u \):th vector of values is observed (subsampling rate \( u \))
- Subsampling induces confounding, and unidentifiability
- Ignoring subsampling can lead to significant errors!
- Applications: e.g. fMRI.
• Adding instantaneous effects in a linear model (see for example Hyvärinen et al 2010).

• Adding instantaneous effects in a linear model (see for example Hyvärinen et al 2010).

• Continuous time approaches, but some processes are inherently discrete time (e.g. salary payment).
• Adding instantaneous effects in a linear model (see for example Hyvärinen et al 2010).

• Continuous time approaches, but some processes are inherently discrete time (e.g. salary payment).

• Recently Plis et al. (UAI2015,NIPS2015) considered modeling subsampling directly, assuming on the system timescale level:
  • discrete time
  • first order Markov: $\mathbf{V}^t \perp \perp \mathbf{V}^{t-k} | \mathbf{V}^{t-1}$
  • no instantaneous effects, or unobserved common causes
  • nonparametric (continuous or discrete values, SVAR processes, or dynamic BNs)
• Adding instantaneous effects in a linear model (see for example Hyvärinen et al 2010).

• Continuous time approaches, but some processes are inherently discrete time (e.g. salary payment).

• Recently Plis et al. (UAI2015,NIPS2015) considered modeling subsampling directly, assuming on the system timescale level:
  • discrete time
  • first order Markov: \( \mathbf{V}^t \perp \perp \mathbf{V}^{t-k} | \mathbf{V}^{t-1} \)
  • no instantaneous effects, or unobserved common causes
  • nonparametric (continuous or discrete values, SVAR processes, or dynamic BNs)

• Corresponding parametric method: Gong et al. (ICML2015) discovered linear models using non-Gaussianity.
Rolled Representation

... $V_{1}^{t-2}$ $V_{1}^{t-1}$ $V_{1}^{t}$ ...

... $V_{2}^{t-2}$ $V_{2}^{t-1}$ $V_{2}^{t}$ ...

... $V_{3}^{t-2}$ $V_{3}^{t-1}$ $V_{3}^{t}$ ...

system t.s.

... measurement t.s.

unrolling
Rolled Representation

system t.s.

... $V_1^{t-2}$ $V_1^{t-1}$ $V_1^t$ ...

... $V_2^{t-2}$ $V_2^{t-1}$ $V_2^t$ ...

... $V_3^{t-2}$ $V_3^{t-1}$ $V_3^t$ ...

unrolling

marginalization

measurement t.s.

... $V_1^{t-2}$ $V_1^t$ ...

... $V_2^{t-2}$ $V_2^t$ ...

... $V_3^{t-2}$ $V_3^t$ ...
Rolled Representation

\[
\begin{align*}
\cdots & \quad V_1^{t-2} \quad V_1^{t-1} \quad V_1^t \quad \cdots \\
\cdots & \quad V_2^{t-2} \quad V_2^{t-1} \quad V_2^t \quad \cdots \\
\cdots & \quad V_3^{t-2} \quad V_3^{t-1} \quad V_3^t \quad \cdots
\end{align*}
\]

unrolling

\[
\begin{align*}
\cdots & \quad V_1 \quad \cdots \\
\cdots & \quad V_2 \quad \cdots \\
\cdots & \quad V_3 \quad \cdots
\end{align*}
\]

\text{marginalization}

\[
\begin{align*}
\cdots & \quad V_1^{t-2} \quad V_1^t \quad \cdots \\
\cdots & \quad V_2^{t-2} \quad V_2^t \quad \cdots \\
\cdots & \quad V_3^{t-2} \quad V_3^t \quad \cdots
\end{align*}
\]

rolling

\[
\begin{align*}
\cdots & \quad V_1 \quad \cdots \\
\cdots & \quad V_2 \quad \cdots \\
\cdots & \quad V_3 \quad \cdots
\end{align*}
\]
Induced confounding

system t.s.

unrolling

marginalization

measurement t.s.

rolling
Induced confounding

system t.s.

\[ \cdots \cdots \vdots \]

unrolling

marginalization

measurement t.s.

\[ \cdots \cdots \vdots \]
Result 1: Deciding whether there is a system t.s. structure compatible with the directed edges of a measurement t.s. structure is **NP-complete** for any fixed $u \geq 2$. 
Result 1: Deciding whether there is a system t.s. structure compatible with the directed edges of a measurement t.s. structure is **NP-complete** for any fixed $u \geq 2$.

Result 2: A constraint satisfaction solution by ASP:

- We encoded the problem (the marginalization operation) using the expressive declarative modeling language
- Solver Clingo (Gebser et al. 2011) uses state-of-the-art SAT-solving techniques to give an exact and complete solution
- ASP is relatively easy and quick to use, the encoding is easily extendable
- Subsampling rate $u$: fixed or free.
Scalability of Enumerating 1000 Solutions

( fixed subsampling rate 2, SAT is our approach, MSL is the previous state of art by Plis et al. (2015) )
Task 2: Finding Structures Compatible with Data

\[ \ldots \ V_1^{t-4} \ V_1^{t-2} \ V_1^t \ \ldots \]

\[ \ldots \ V_2^{t-4} \ V_2^{t-2} \ V_2^t \ \ldots \ \rightarrow \]

\[ \ldots \ V_3^{t-4} \ V_3^{t-2} \ V_3^t \ \ldots \]

\[ \Downarrow \ \Downarrow \ \Downarrow \]

Data

Measurement t.s.

System t.s.
Task 2: Finding Structures Compatible with Data

\[ \ldots \ V_1^{t-4} \ V_1^{t-2} \ V_1^t \ \ldots \]

\[ \ldots \ V_2^{t-4} \ V_2^{t-2} \ V_2^t \ \ldots \rightarrow \]

\[ \ldots \ V_3^{t-4} \ V_3^{t-2} \ V_3^t \ \ldots \]

\[ \overset{\text{data}}{\rightarrow} \overset{\text{measurement t.s.}}{\rightarrow} \overset{\text{system t.s.}}{\rightarrow} \]

- Measurement t.s. structure can be consistently estimated from data: e.g. \( V_1 \rightarrow V_3 \ \Leftrightarrow \ V_1^{t-2} \not\perp V_3^t \mid V_2^{t-2}, V_3^{t-2} \)
Task 2: Finding Structures Compatible with Data

\[ \ldots V_{t-4}^1 V_{t-2}^1 V_1^t \ldots \]

\[ \ldots V_{t-4}^2 V_{t-2}^2 V_2^t \ldots \rightarrow \]

\[ \ldots V_{t-4}^3 V_{t-2}^3 V_3^t \ldots \]

\[ \rightarrow \]

\[ \begin{array}{c}
V_1 \\
V_3 \\
V_2 \\
\end{array} \]

\[ \begin{array}{c}
V_1 \\
V_3 \\
V_2 \\
\end{array} \]

data measurement t.s. system t.s.

- Measurement t.s. structure can be consistently estimated from data: e.g. \( V_1 \rightarrow V_3 \iff V_1^{t-2} \not\perp\!\!\!\perp V_3^t \mid V_2^{t-2}, V_3^{t-2} \)
- Due to finite samplesize, the constraint satisfaction approach will often return UNSATISFIABLE
Task 2: Finding Structures Compatible with Data

\[ \cdots \quad V_1^{t-4} \quad V_1^{t-2} \quad V_1^t \quad \cdots \]

\[ \cdots \quad V_2^{t-4} \quad V_2^{t-2} \quad V_2^t \quad \cdots \quad \rightarrow \]

\[ \cdots \quad V_3^{t-4} \quad V_3^{t-2} \quad V_3^t \quad \cdots \]

- Measurement t.s. structure can be consistently estimated from data: e.g. \( V_1 \rightarrow V_3 \iff V_1^{t-2} \not\perp \!\!\!\perp V_3^t \mid V_2^{t-2}, V_3^{t-2} \)

- Due to finite samplesize, the constraint satisfaction approach will often return UNSATISFIABLE

- Find the system t.s. structure such that the corresponding measurement t.s. structure is optimally close to the estimated (Task 2).
Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
- Objective function is the sum of the penalities
Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalities

- Clingo uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.
Result 3: A Constraint Optimization Solution

Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
- Objective function is the sum of the penalities

- Clingo uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.

- We scale to 11-12 within 10 minutes, depending on the sample size and other specifics
Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalties

- Clingo uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.

- We scale to 11-12 within 10 minutes, depending on the sample size and other specifics

- Previous work by Plis et al. 2015: searching neighbors of the estimated measurement t.s. structure — resembles the uniform weighting scheme.
( fixed subsampling rate 2, average result of the eq. class, 6 nodes, av. degree 3, 200 samples, 100 data sets, linear models )
Causal discovery from subsampled time series data:
Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.

Future work: generalizing the model space, e.g. allowing for unobserved confounding time series.

Thanks!
Causal discovery from subsampled time series data:
- A non-parametric constraint satisfaction approach: Much better scalability than previous state-of-the-art.
- A (first) constraint optimization approach: More accurate than unweighted or unoptimal solutions.

Conclusion
Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.

- A (first) constraint optimization approach:
  More accurate than unweighted or unoptimal solutions.

- Future work: generalizing the model space, e.g. allowing for
  unobserved confounding time series.
Conclusion

Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.
- A (first) constraint optimization approach:
  More accurate than unweighted or unoptimal solutions.
- Future work: generalizing the model space, e.g. allowing for unobserved confounding time series.

Thanks!