Causal Macro Variables

Frederick Eberhardt

(joint work with Krzysztof Chalupka and Pietro Perona)
Causal Discovery

truth
(unknown)

\[ x \rightarrow y \rightarrow z \]

samples

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Causal Discovery

truth
(unknown)

$\xrightarrow{}$

$x \rightarrow y \rightarrow z$

$\xrightarrow{}$

samples
$x \ y \ z$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

$\xrightarrow{}$

inference-algorithm
Causal Discovery

assumptions, e.g.
- acyclicity
- no latent variables
- parametric form
- time order
- etc.

truth (unknown)
\[ \begin{array}{ccc}
  x & y & z \\
  1 & 1 & 1 \\
  0 & 1 & 1 \\
  1 & 0 & 0 \\
  0 & 0 & 0 \\
  \ldots & \ldots & \ldots
\end{array} \]

samples

inference-algorithm
Causal Discovery

truth (unknown)

\[ x \rightarrow y \rightarrow z \]

samples

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

assumptions, e.g.
- acyclicity
- no latent variables
- parametric form
- time order
- etc.

\[ x \rightarrow y \rightarrow z \]

equivalence classes of causal structures

\[ x \rightarrow y \rightarrow z \]

\[ x \leftarrow y \rightarrow z \]

\[ x \leftarrow y \leftarrow z \]
Causal Discovery

truth (unknown)

\[ x \rightarrow y \rightarrow z \]

assumptions, e.g.
- acyclicity
- no latent variables
- parametric form
- time order
- etc.

data

inference-algorithm

\[ x \rightarrow y \rightarrow z \]
\[ x \leftarrow y \rightarrow z \]
\[ x \leftarrow y \leftarrow z \]

equivalence classes of causal structures
Causal Macro Variables

attractiveness

judgement
Causal Macro Variables

symmetry

attractiveness judgement
Causal Macro Variables

symmetry → attractiveness judgement

attractiveness judgement
Causal Macro Variables

- symmetry

→ attractiveness judgement

Examples of symmetry: faces of two people.
Causal Macro Variables

Symmetry

Constitutive relation

Attractiveness judgement

Constitutive relation
Causal Macro Variables

symmetry

constitutive relation

How to construct macro features? Unsupervised?

attractiveness judgement

constitutive relation
The Aim

- account of the construction of causal variables
- applicable to complex macro-level causes
- domain general
- supports an interpretation of causation as invariance under intervention
Ambiguous Manipulation

Total Cholesterol \rightarrow \text{Heart Disease}

Spirtes & Scheines (2002)
Ambiguous Manipulation

Total Cholesterol

HDL

LDL

+ 

- 

Heart Disease

Spirtes & Scheines (2002)
the causal effect of Total Cholesterol on Heart Disease is ambiguous

⇒ Total Cholesterol is over-aggregated, it cannot be described as a cause of Heart Disease
Ambiguous Manipulation

- if HDL and LDL have the same causal effect on Heart Disease then the causal effect of Total Cholesterol on Heart Disease is NOT ambiguous

➡ we can aggregate HDL and LDL into Total Cholesterol, which is a cause of Heart Disease

Spirtes & Scheines (2002)
Ambiguous Manipulation

- if HDL and LDL have the same causal effect on Heart Disease then the causal effect of Total Cholesterol on Heart Disease is NOT ambiguous
  ➡ we can aggregate HDL and LDL into Total Cholesterol, which is a cause of Heart Disease

Spirtes & Scheines (2002)
Ambiguous Manipulation

Total Cholesterol

HDL

LDL

Heart Disease
Ambiguous Manipulation

Total Cholesterol

HDL
LDL

ratio

Heart Disease
Ambiguous Manipulation

- arbitrary choices of variables imply correlated errors
- interventions would be interventions on the variable and the error term
Constructing / Identifying Macro Variable

- Account of the construction of causal variables
- Applicable to complex macro-level causes
- Domain general
- Supports an interpretation of causation as invariance under intervention

- Merge states that have the same causal effect
- Do not merge if an ambiguous manipulation would result
Toy example (discrete spaces)

image

hidden confound

Target behavior (binary)
Toy example (discrete spaces)
Toy example (discrete spaces)

The horizontal, but not the vertical bar, is causal of the target behavior, even though both are predictive of it.
True Macro-Causal Model

P(T=0 | \( \epsilon \)) = 1
P(T=0 | \( I \)) = 0.66
P(T=0 | \( H \)) = 0.33
P(T=0 | \( T \)) = 0

P(T=0 | do \{ \( \epsilon \) \}) = 0.83
P(T=0 | do \{ \( I \) \}) = 0.3
Observational Partition

space of images $\mathcal{I}$

behavior space $\mathcal{T}$

$P(T=0 | \cdot ) = 0$

$P(T=0 | \cdot ) = 1$

$P(T=0 | \cdot ) = 0.33$

$P(T=0 | \cdot ) = 0.66$
Observational Partition

space of images $\mathcal{I}$

behavior space $\mathcal{T}$

$P(T|I)$
**Observational Partition**

- **observational partition**: partitions the space of images according to the equivalence relation induced by the conditional probability of the target behavior $T$ given the image $I$.
• **observational partition**: partitions the space of images 
according to the equivalence relation induced by the conditional 
probability of the target behavior \( T \) given the image \( I \)

\[
i_1 \sim_I i_2 \iff \forall t \in \mathcal{T} P(t \mid i_1) = P(t \mid i_2)
\]
Causal Partition

space of images $\mathcal{I}$

$P(T \mid I) \neq P(T \mid do(I))$

behavior space $\mathcal{T}$
Causal Partition

space of images $\mathcal{I}$

behavior space $\mathcal{T}$

$P(T|I)$

$\neq P(T|do(I))$
Causal Partition

- causal partition: partitions the image space according to the equivalence relation induced by the probability of the target behavior $T$ given an intervention on the image

\[ i_1 \sim_I i_2 \iff \forall t \in T P(t \mid do(i_1)) = P(t \mid do(i_2)) \]
Causal Partition

\[ P(T=0 \mid \text{do}(i_1)) = P(T=0 \mid \text{do}(i_2)) \]

- **causal partition**: partitions the image space according to the equivalence relation induced by the probability of the target behavior \( T \) given an intervention \( I \) on the image

\[ i_1 \sim_I i_2 \iff \forall t \in \mathcal{T} \ P(t \mid \text{do}(i_1)) = P(t \mid \text{do}(i_2)) \]
• **causal partition**: partitions the image space according to the equivalence relation induced by the probability of the target behavior $T$ given an *INTERVENTION* on the image

$$i_1 \sim_I i_2 \iff \forall t \in \mathcal{T} P(t \mid \text{do}(i_1)) = P(t \mid \text{do}(i_2))$$

*macro cause*: the macro cause $C$ of a target behavior $T$ is a random variable whose value stands in a bijective relation to the causal class of the image
Observational vs. Causal Partition

observational partition of $\mathcal{I}$

$P(T|I)$

causal partition of $\mathcal{I}$

$P(T|do(I))$
Causal Coarsening Theorem

For
- multinomial distributions
- no causal feedback
- [technical assumption about the nature of confounding]

the subset of distributions that induce a causal partition that is not a coarsening of the observational partition is Lebesgue measure zero.
Applying the Causal Coarsening Theorem

- learn the observational partition from non-experimental data
- under the assumptions of the theorem, the relevant causal distinctions are a subset of the detected distinctions
- test which distinctions are causal with a few experiments
we found the macro-level climate phenomenon of El Niño supervening on micro-level wind and sea surface temperature data of the equatorial Pacific in an entirely data driven (unsupervised) manner

Chalupka, Bischoff, Perona & Eberhardt (2016)
Multiple Levels of Causal Description

merge states with identical causal effects
Multiple Levels of Causal Description

merge states with identical causal effects

deliberately coarsen
Multiple Levels of Causal Description

$C^*$

*merge*

merge states with identical causal effects

$C$

$E$

deliberately coarsen

$I$

$J$
Multiple Levels of Causal Description

\[ \text{do}(C^* = c^*) \]

merge states with identical causal effects

\[ C \rightarrow C^* \]

deliberately coarsen

\[ E \rightarrow J \]

intervene to some state that maps to \( c^* \)
Causal Macro-Variables

- account of causal macro-variables that
  - turns the question about the existence of causal macro-variables into an empirical question
  - identifies a privileged level of aggregation that retains exactly the causal information of the underlying micro-systems
  - supports a causal interpretation in terms of intervention (and avoids known problems of causal variable definition)
  - is domain-general

- algorithms that discover/construct such causal macro-variables

- applications as proof of concept
can we use the same approach to search for macro-level neural features that are causal of behavior?
Neuroscience-based Psychology

personality

intelligence

fMRI measure

Left hemisphere

Right hemisphere

behavior

Neuroticism

Openness

Agreeableness

Extraversion

Conscientiousness

responses on self reports and questionnaires

performance on cognitive tasks

Neuroscience-based Psychology

explanation

causal inference

causal graph of brain regions

All code available in python from Chalupka’s webpage.
Collaborators

Krzysztof Chalupka

Pietro Perona


All code available in python from Chalupka’s webpage.

Thank you!