Causal inference in graphical models with latent variables. From theory to practice.

Sam Maes, Stijn Meganck, Philippe Leray
sammaes@vub.ac.be

Laboratoire d'Informatique, Traitement de l'Information, Systèmes.

Causality and Probability in the Sciences
Canterbury

15th June 2006

Outline

1. Introduction
2. Causal Inference
3. Learning
4. Parametrisation

Outline

1. Introduction
2. Causal Inference
3. Learning
4. Parametrisation

Introduction

Subtasks of causal modeling with latent variables:

- Structure learning from data:
  - observational
  - experimental
- Learning parameters
- Probabilistic inference
- Causal inference

Problem

No integral approach for all these tasks in the presence of latent variables.

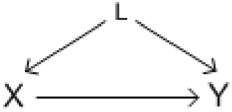
- Causal inference :
  semi-Markovian causal models
- Structure learning from observational data:
  ancestral graph models

Our solution:

- use ancestral graph models for learning
- transform the result into a semi-Markovian causal model

Assumptions

- data generated by an underlying faithful causal DAG
- some variables are latent to the user
- sufficient data and learning algorithms make no errors
- no latent common cause of 2 variables with an immediate causal connection
- no selection bias



Outline

1 Introduction

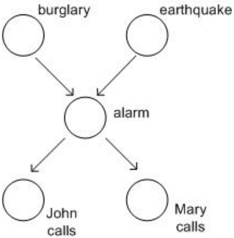
2 Causal Inference

3 Learning

4 Parametrisation

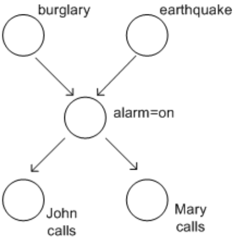
Probabilistic vs Causal Inference

underlying DAG



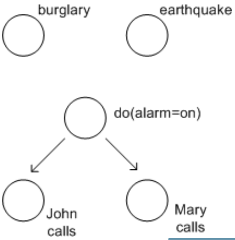
after observation

- instantiate the observed variables
- propagate



after manipulation

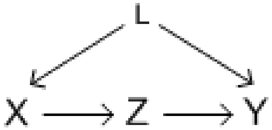
- replacer old causes
- instantiate
- propagate



With latent variables

Causal inference becomes more complicated:

- **replace old causes**
- instantiate
- propagate

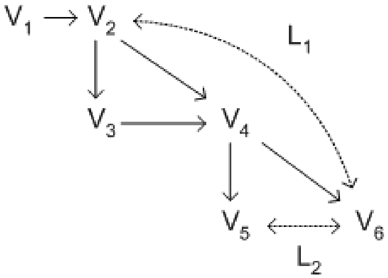


$P(Y = y|do(X = x))$: manipulate variable X and study the effect on Y.

Representation for causal inference

semi-Markoviens causal models (SMCM)

- directed edge represents an autonomous causal relation
- bi-directed edge represents a latent common cause
- importance: every models with arbitrary latent variables can be transformed into a SMCM
- a joint probability distribution: e.g. $P(V_1, \ldots, V_6)$



Inference in SMCMs

causal inference algorithm exists (Tian & Pearl), but:

- no efficient parametrisation
- no probabilistic inference algorithm
- no structural learning algorithm

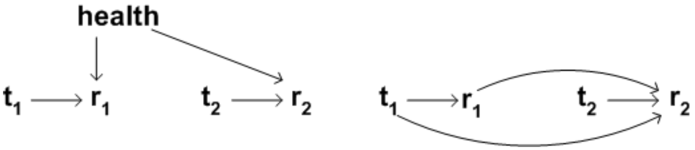
Outline

1. Introduction
2. Causal Inference
3. Learning
4. Parametrisation

Representation for learning

The class of DAGs is not complete under marginalisation.

I.e., a DAG of the observable variables can not exactly represent all the independences between the variables.



Maximal ancestral graphs (MAG)

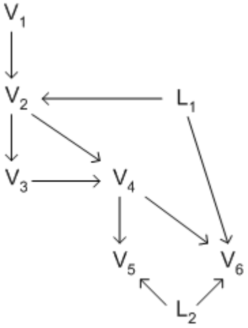
Maximal ancestral graphs without conditionning

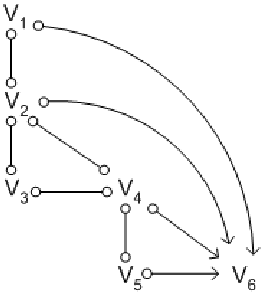
- directed edges have an ancestral meaning $\neq$ causal: there is a causal path in the underlying DAG
- bi-directed edges: latent common causes
- max. 1 edge between 2 variables: ancestral relations absorb latent common causes
- maximal: every absent edge represents an independence

Markov equivalence class

complete partial ancestral graph (CPAG):

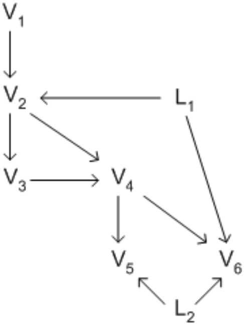
- Fast Causal Inference (FCI)
- Rules to orient edges
- 3 possible edge marks: o, -, >

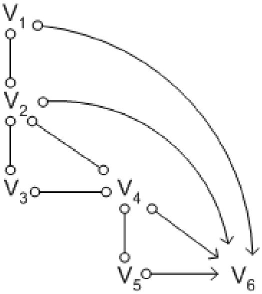




3 possible explanations for each edge:

- causal relation V_1, V_2
- latent variable V_2, V_6
- "inducing path" between V_1 et V_6
  - V_1 can not be separated from V_6 by using observed variables
  - due to the maximality of the models FCI finds an edge





CPAG $\longrightarrow$ SMCM

Perform experiments to differ between cases:

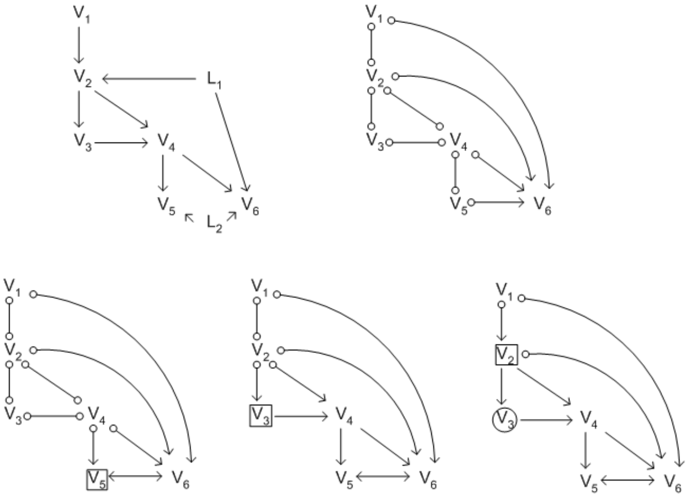
- Type 1: resolve $o\!\!\longrightarrow$
- Type 2: resolve $o\!\!-\!\!o$
- Remove i-false edges

CPAG $\longrightarrow$ SMCM (Type 1)

Type 1: resolve $A \multimap B$

- $exp(A) \not\rightsquigarrow B$: $A \leftrightarrow B$
- $exp(A) \rightsquigarrow B$:
  - $\nexists$ pot.dir. path $A \dashrightarrow B$ of length ≥ 2: $A \longrightarrow B$
  - $\exists$ pot.dir. path $A \dashrightarrow B$ of length ≥ 2:
    block each pot.dir. path by conditioning on a set D
    - $exp(A)|D \rightsquigarrow B$: $A \longrightarrow B$
    - $exp(A)|D \not\rightsquigarrow B$: $A \leftrightarrow B$

Type 1: examples



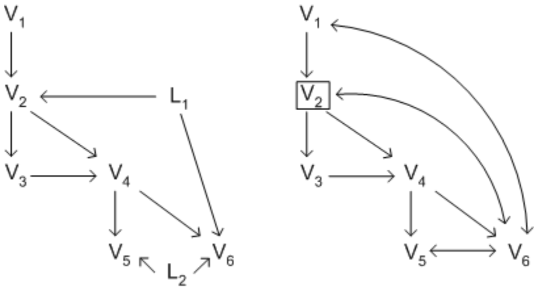
CPAG $\rightarrow$ SMCM (Type 2)

Type 2: resolve o—o

- Easily transformed to Type 1

Remove i-false edges

- recognize the edges $A \leftrightarrow B$ and $A \rightarrow B$ that are possible created due to an inducing path
- block each inducing path between A, B with experiments E
- block each other path between A, B by condit. on a set D
- $exp(E)|D$: $A \not\perp\!\!\!\perp B$: confounding edge is real
  $A \perp\!\!\!\perp B$: remove i-false edge



Outline

1. Introduction
2. Causal Inference
3. Learning
4. **Parametrisation**

Parametrisation for discrete variables

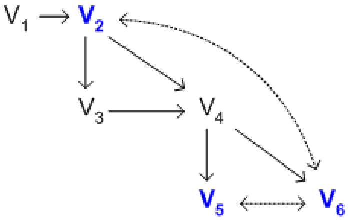
Represent the SMCM with a DAG

- this DAG is an I-map of SMCM
- i.e. all the indep. found in the DAG are also in the SMCM
- use this DAG for:
  - learning parameters
  - probabilistic inference
- for causal inference:
  - use the SMCM for the structure
  - use the DAG for the parameters

Some definitions

C-component

In a SMCM, the set of observable variables that are connected by a bi-directed path belong to the same **c-component** (from "confounded component").



Parametrisation: procedure

Given a SMCM and a topological order, the PR-representation is obtained as follows:

- The nodes are V, the observable variables of the SMCM.
- The directed edges of the SMCM remain.
- The bi-directed edges in the SMCM are replaced:

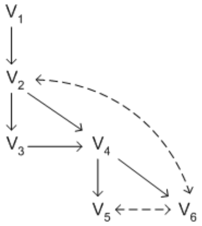
  Add a directed edge from node V_i to node V_j iff:
  - V_i is a topological predecessor in the same c-component
  - a parent of a topological predecessor in the same c-component
  - except if there was already an edge between nodes V_i and V_j.

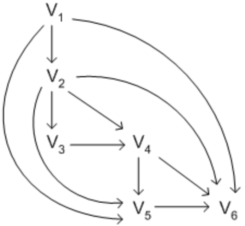
Parametrisation: example

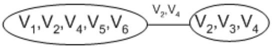
Add an edge $V_i \rightarrow V_j$ iff:

- V_i is a topological predecessor in the same c-component
- a parent of a topological predecessor in the same c-component

with topological order $V_2 < V_5 < V_6$:







Conclusion

From theory to practice.

Somewhat closer, still a lot of assumptions to relax:

- take into account:
  - possible errors in the learning algorithms
  - cost of experiments and impossible experiments
  - rules for marking edges after each experiment
- allow:
  - selection bias
  - confounding between variables with an immediate causal connection
- resolve practical issues such as finding pot. inducing paths and pot. directed paths