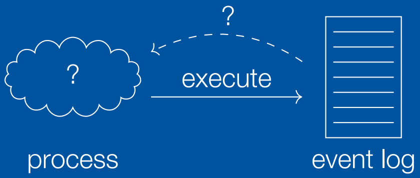
Causal Reasoning over Decisions in Process Models

Sander Leemans, Niek Tax

Process mining



Causal reasoning

Ice cream causes drowning:
it has been shown that the more ice creams are sold,
the more people drown.

correlation $\not\Rightarrow$ causation

causal inference $\neq$ a directly follows b

Causal reasoning in process mining

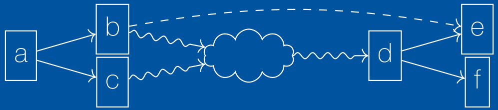
- Interventions → cycle time[1]
- Interventions → outcome likelihood[2]
- Negative outcome → explain[3]

♡

Today:

- Causal soft long-distance dependencies
  → inform interventions
  → inform process redesign

Let's take this a step further:



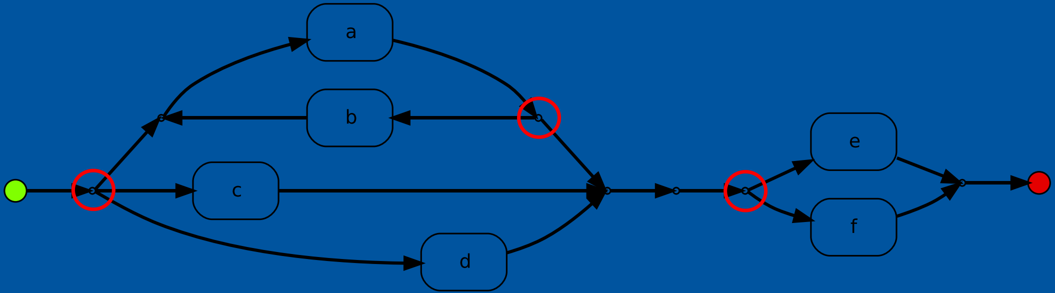
- Increasing likelihood of b increases likelihood of e

[3] Qafari, M.S., van der Aalst, W.M.P.: Case level counterfactual reasoning in process mining. CAiSE Forum 2021

[2] Bozorgi, Z.D., Teinemaa, I., Dumas, M., Rosa, M.L., Polyvyanyy, A.: Process mining meets causal machine learning: Discovering causal rules from event logs. ICPM 2020

[1] Bozorgi, Z.D., Teinemaa, I., Dumas, M., Rosa, M.L., Polyvyanyy, A.: Prescriptive process monitoring for cost-aware cycle time reduction. ICPM 2021

Process trees & choices



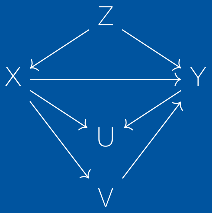
Language: $\{\langle c, e\rangle, \langle c, f\rangle, \langle a, e\rangle, \langle a, f\rangle, \langle a, b, a, e\rangle, \langle a, b, a, f\rangle, \langle a, b, a, b, a, e\rangle, \langle a, b, a, b, a, f\rangle \ldots\}$

[1] Our causal technique also supports Directly Follows Models.

Causal reasoning

Causal graph:

- Nodes: decision variables
- Edges: "there might be a causal relation"
  - missing edges denote information
  - directed acyclic graph



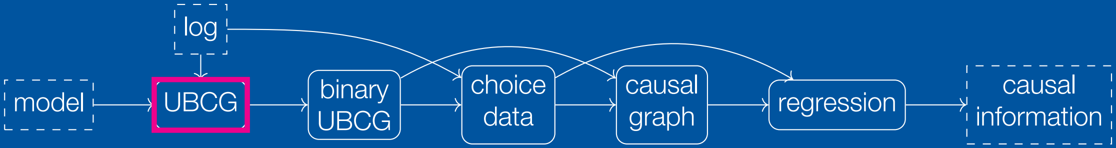
confounding factor

collider

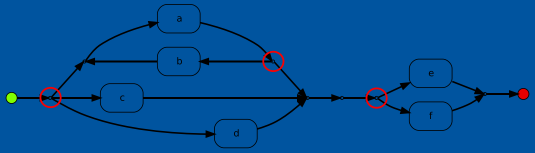
chain

From a causal graph and observational data, in some cases the strength of causal relations can be computed.

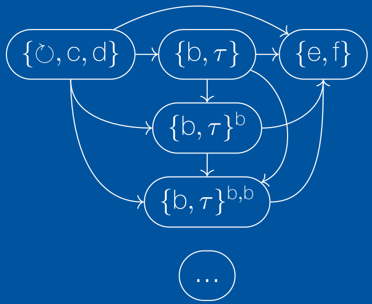
Our approach



Upper bound causal graph

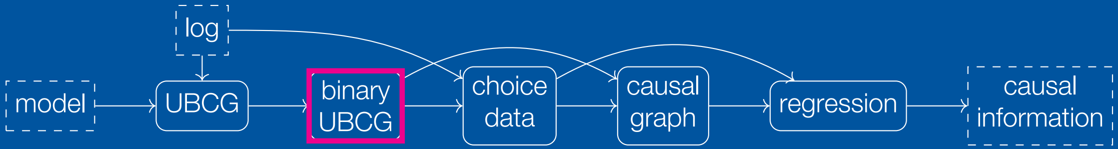


- Nodes: choices
  - Unfold loops
- Edges: add all edges, except if they violate the model



- Process trees
- Directly follows models

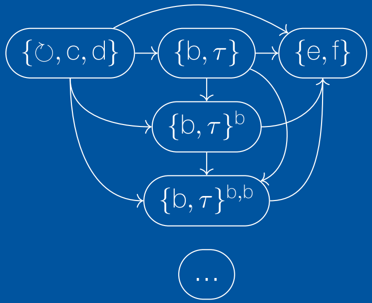
Outline

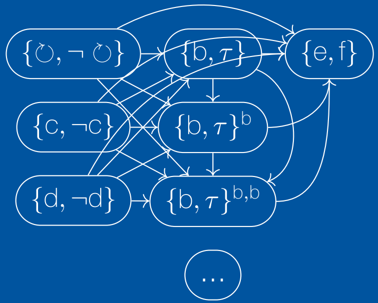


log
model
UBCG
binary UBCG
choice data
causal graph
regression
causal information

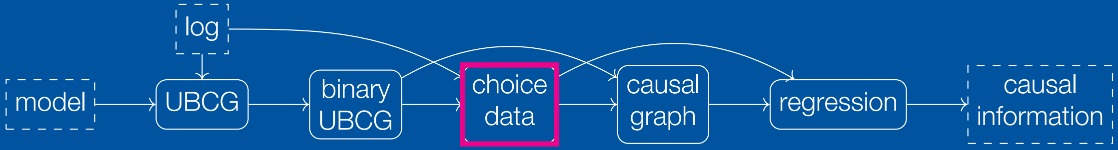
Binary UBCG

- Split nodes into binary choices
- The split nodes have no causal relation





Outline

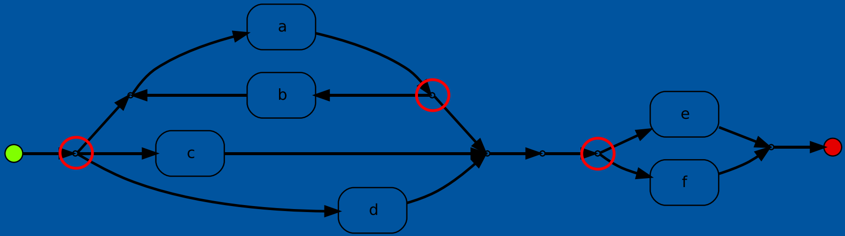


log
model
UBCG
binary UBCG
choice data
causal graph
regression
causal information

Choice data

- Align each trace
- Record choices taken

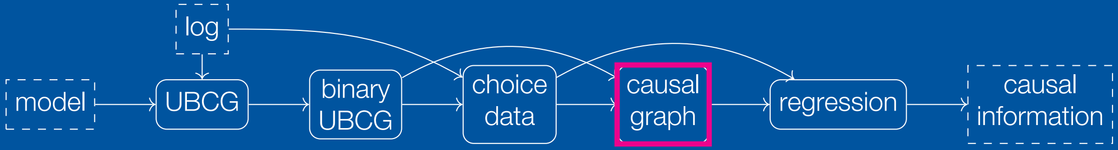
$\{\langle c, f\rangle$
$\langle a, e\rangle\}$



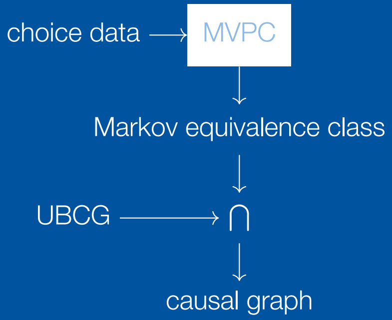
aligned trace	$\{\circlearrowleft, \neg\circlearrowleft\}$	$\{c, \neg c\}$	$\{d, \neg d\}$	$\{b, \tau\}$	$\{b, \tau\}^{b}$	$\{b, \tau\}^{b,b}$	$\{e, f\}$
$\langle c, f\rangle$	$\neg\circlearrowleft$	c	$\neg d$	$\boxed{-}$	-	-	f
$\langle a, \tau, e\rangle$	$\circlearrowleft$	$\neg c$	$\neg d$	τ	-	-	e

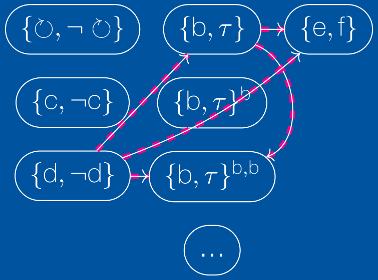
Three-symbol logic

Outline


log
model
UBCG
binary UBCG
choice data
causal graph
regression
causal information

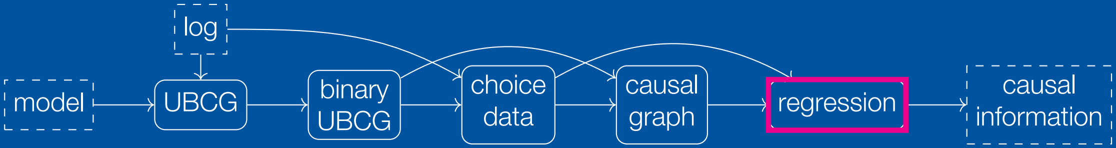
Causal discovery



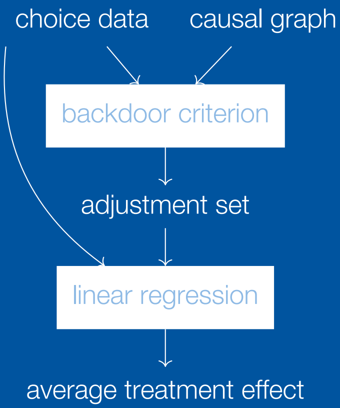


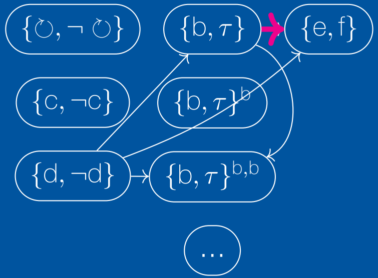
[1] Ruibo Tui. Causal Discovery in the Presence of Missing Data. AISTATS 2019.

Outline


log
model
UBCG
binary UBCG
choice data
causal graph
regression
causal information

Regression





Linear regression:

$\{e, f\} = c_0 + c_1\{b, \tau\} + c_2\{d, \neg d\}$

- Adjustment set $\{\{d, \neg d\}\}$
- b increases the probability of e by c_1

Assumptions

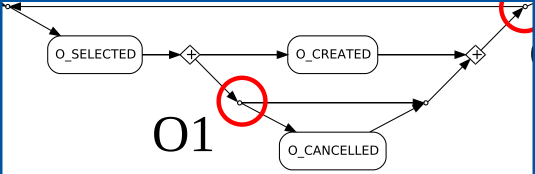
~~Wishful thinking~~ Conjectures

- One causal graph from conjunction with UBCG
- Treatment effect identifiable from causal graph as per backdoor criterion

Assumptions

- Model denotes moments of choice correctly
- No unmodelled confounding factors

Evaluation: applicability



Skipping o_cancelled causes a reduction of o_cancelled by 0.62 in the second loop run.

- 5 real-life logs
- 2 discovery techniques

- Choices: 2 - 288
- UBCG edges: 1 - 23 776
- Causal graph edges: 0 - 7
- Run time: 0.04 - 75s

Conclusion

- Causal relations are there and can be detected
- It’s feasible
- User not flooded

You have been watching...

PM + causal analysis

- UBCG
  - Process trees
  - Directly follows models
- binary UBCG
- choice data
- causal graph
- regression

Gaps

- Visualise & explain
- Linear probability model

Future work

- Confounding factors from event data
- Causal analysis on Petri nets
- Case study

Sander Leemans
s.leemans@bpm.rwth-aachen.de
http://leemans.ch

RWTH AACHEN UNIVERSITY