Causal discovery in manufacturing

Konstantin Göbler\textsuperscript{1,3} \quad Tobias Windisch\textsuperscript{2} \quad Tim Pychynski\textsuperscript{3}
Steffen Sonntag\textsuperscript{3} \quad Martin Roth\textsuperscript{3} \quad Mathias Drton\textsuperscript{1}

\textsuperscript{1}TU München \quad \textsuperscript{2}HS Kempten \quad \textsuperscript{3}Robert Bosch GmbH
Production data

Why?
Correlation sucks

\[ p_{\text{corr}} \approx 0.01 \]

\[ \begin{align*}
X &= N_0,\epsilon_X \\
Y &= X^2 + N_0,\epsilon_Y
\end{align*} \]

\[ X \rightarrow Y \]

\[ p_{\text{corr}} \approx 0.97 \]

\[ \begin{align*}
Z &= N_0,\epsilon_Z \\
X &= 2 \cdot Z + N_0,\epsilon_X \\
Y &= 3 \cdot Z + N_0,\epsilon_Y
\end{align*} \]

\[ Z \leftarrow X \leftarrow Y \]
CausalAI

**Graphical models**

**Acyclic graph**

\[
\begin{align*}
X_1 &= f_1(X_2, X_3, \epsilon_1) \\
X_2 &= f_2(\epsilon_2) \\
X_3 &= f_3(\epsilon_3) \\
X_4 &= f_4(X_1, X_5, \epsilon_4) \\
X_5 &= f_5(\epsilon_5)
\end{align*}
\]

**Statistical equations**

**Causal Discovery**

Learning causal structure from data is hard

- Too many possible structures
- Identifiability: \( Z = 2 \cdot X \) or \( X = \frac{1}{2} \cdot Z \)

⇒ Adapted algorithms required!

What about production data?
How to benchmark causal discovery?

Fundamental benchmarking issue

Benchmarking requires true causal relations.

▶ Expensive to create large benchmark datasets for real problems
▶ Most causal discovery algorithms tested on synthetic data

Specialized algorithms needed

Ground truth cannot be released
How to get a benchmark dataset for manufacturing...

- ...a strong manufacturing flavor?
- ...with identifiable causal structure?
- ...without privacy issues?
Getting a ground truth

\[ f_k(x_{1:k}) = \mathbb{E}(X_k - f_k^*(X_{P_k}^*) | X_{1:t} = x_{1:t}) \]
causalAssembly: Creating realistic semi-synthetic production data

A real assembly line of Robert Bosch GmbH

Sensitive production data

Incorporated domain knowledge

Define as true causal structure

Distributional random forests

Semi-synthetic production data

Unique causal structure

Same flavour
Impressions from the data

Causal relations

Correlations
Benchmarking CausalAI for production data

⇒ Production data yet to tough for vanilla CausalAI

Now its your turn!

boshresearch/causalAssembly

```python
from causalAssembly.models.dag import ProductionLineGraph
df = ProductionLineGraph.get_data()
causal_dag = ProductionLineGraph.get_ground_truth()
```