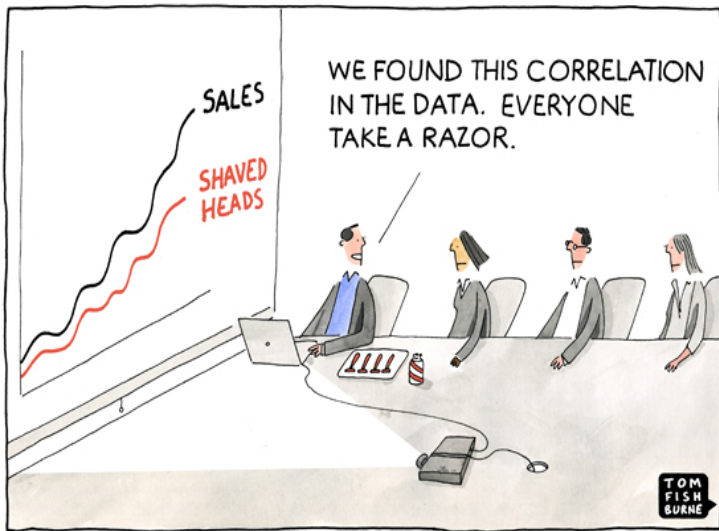


A Gentle Introduction to Structural Causal Models

Daniel Saggau

5/11/2021

Motivation



© marketoonist.com

Introduction (1)

- Machine learning has provided many insights into different problems
- One issue is the consideration of 'What are we actually predicting?'
- Mainstream tools are build on association-based learning
- Associations are not enough for high stake settings
- We want causation and not correlation

Introduction (2)

Causal assumptions differentiate causal models from association learning methods.

Association-based Concepts	Causal Concepts
Correlation	Randomization
Regression	Confounding
Conditional Independence	Disturbance
Likelihood	Error Terms
Odds Ratio	Structural Coefficients
Propensity Score	Spurious Correlation

Table of Contents

- Foundations of SCMs
- Assumptions
 - Comparative Causal Tools
 - Historical Development
- Pearl's Causal Hierarchy
 - Prediction
 - Intervention
 - Counterfactuals
 - Implications
- Graphical Models
 - Features
 - Implications
- Causality and Time

Foundations of SCMs (1)

- System of equations
- Assignment equation $':=$ ' rather than regular equation $'=$ '
- Nonparametric SEM
- Functional form rather than using conditional probabilities
- $\text{SCM} = \text{PO} + \text{Graphical Tools}$
- Exogenous factors are part of the model specification

Error terms

Regression: Omissible outside factor SCM/SEM; Latent influential factor that is pivotal for the model specification but not observable

Consists of graph and assignments: Baseline:

$$C := N_C$$

$$E := f_E(C, N_E)$$

source: Peters, Janzing, and Schölkopf (2017)

Assumptions

- First, there was no sign to express the assignment equation and people used the '=' and one would e.g. write $A = B$.
- Treating an equation as a **algebraic equation** led to confusion because those have no causal information.
- This algebraic equation would imply that $B = A$ because the order has no concrete meaning in algebraic equations.
- The problem is that the equation is symmetric.
- The initial '=' sign was replaced with the ':=' which is asymmetric (Pearl, 2009) and called an **assignment**.
- This misconception has caused a lot of challenges.¹
- As mentioned, we define variables as functions e.g. $A = f_A(B, U_A)$.

¹For more information see Pearl (2009)

Assumptions in Causality

Established Conditions:

Definition: **SUTVA**

'The treatment that one unit receives does not change the effect of treatment for any other unit.' (Hardt and Recht 2021)

Definition: **Consistency**

The outcome Y agrees with the potential outcome corresponding to the treatment indicator.' (Hardt and Recht 2021)

Definition: **Ignorability**

The potential outcomes are conditionally independent of treatment given some set of de-confounding variables. (Hardt and Recht 2021) (perfect RCT)

- First two hold for SCM counterfactuals
- third not testable but can check via backdoor criterion in SCM

SCM Applications:

- Flexible simulations for higher order problems (intervention, counterfactual)
- Graphical visualization via directed acyclic graph

Fundamental Differences (1)

- conflict whether to use graphs or not
- A SEM is a parametric specification used in applied sciences (parameters contested)
- A Bayesian causal network is another popular causal model using conditional probabilities and NO functions
- Differences in performance between BCN and SCM# Performance Evaluation

Implications

Method	CBN	SCM
Prediction	<ul style="list-style-type: none">• Unstable• Volatile to parameter changes• Re-Estimate entire model	<ul style="list-style-type: none">• Stable• More Natural Specification• Only estimate Δ CM
Intervention	<ul style="list-style-type: none">• Costly for Non-Markovian Models• Unstable(Nature CP)• Only generic estimates(Δ CP)	<ul style="list-style-type: none">• Pot. Cyclic Representation• Stable• Context specific
Counterfactuals	<ul style="list-style-type: none">• Impossible• no information on latent factors(ϵ)	<ul style="list-style-type: none">• Possible• Inclusion of latent factors

Table 3: Pearls Hierachy of Causation (2009)

Method	Action	Example	Usage
Association $P(a b)$	Co-occurrence	What happened. . .	(Un-)Supervised ML, BN, Reg.
Intervention $P(a do(b), c)$	Do- manipulation	What happens if . . .	CBN,MDP,RL
Counterfactual $P(a_b a', b')$	Hypotheticals	What would have happened if. . .	SCM ,PO

- ML, BN and regression are at the lowest level in the causal hierarchy
- Prediction methods demand the least information and depend on association alone
- Associational methods ignore external changes outside of our data.
- intervention distribution has information on these external changes.
- intervention distribution is only defined in high level causal methods.

- The second query deals with interventions
- Mathematical Tool: do-calculus
- The do-calculus enables us to study the manipulation of parent nodes
- There are various types of intervention
- One example is **atomic intervention**, where we set a variable to a constant
- In **policy intervention** we specify a different function for an equation
- Off-policy intervention models different intervention that is not in our historical data
- CBN , MDP and reinforcement learning model intervention.

Process is described as follows:

- Ⓐ Abduction: Cast probability $P(u)$ as conditional probability $P(u|\epsilon)$
- Ⓑ Action: Exchange ($X = x$)
- Ⓒ Prediction: Compute ($Y = y$)

Graphical Tools

- Nodes
- Edges
- Parents/Ancestors
-

Graphical Illustration - Probabilistic Model

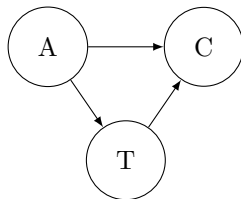


Figure 1: Probabilistic Model

Graphical Illustration - Structural Causal Model

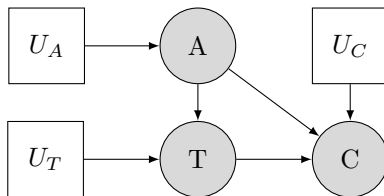


Figure 2: Structural Causal Model

- Time in Physical Sciences: Mechanical and exact
- Time in Social Sciences: Often Vague
- Regular Time Specification is also more vague
- To accommodate that issue, research on differential equation based SCMs started

Graphical Overview

model	predict in IID setting	predict under changing distributions / interventions	answer counter-factual questions	obtain physical insight	automatically learn from data
mechanistic model	Y	Y	Y	Y	?
structural causal model	Y	Y	Y	N	Y??
causal graphical model	Y	Y	N	N	Y?
statistical model	Y	N	N	N	Y

- Hardt, Moritz, and Benjamin Recht. 2021. *Patterns, Predictions, and Actions: A Story about Machine Learning*. <https://mlstory.org>. <http://arxiv.org/abs/2102.05242>.
- Pearl, Judea. 2009. *Causality*. Cambridge university press.
- Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press.