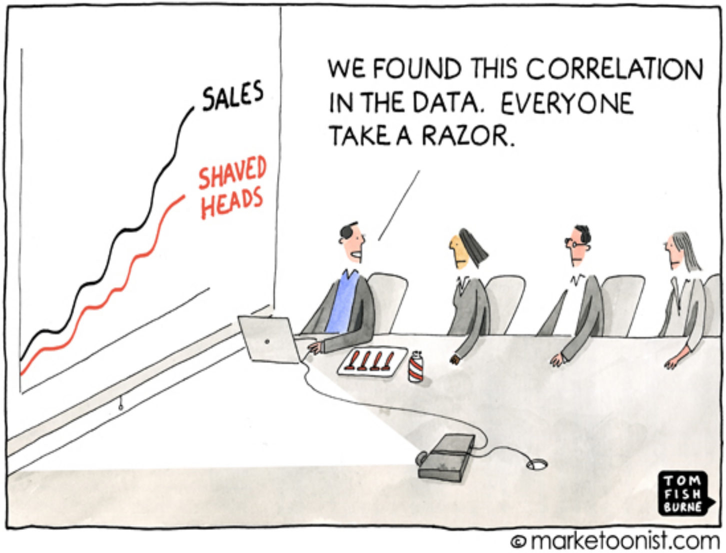
A Gentle Introduction to Structural Causal Models

Daniel Saggau

5/11/2021

Motivation


SALES
SHAVED HEADS
WE FOUND THIS CORRELATION IN THE DATA. EVERYONE TAKE A RAZOR.
TOM FISHBURNE
© 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
- In disciplines like psychology or economics people are less interested in associational learning
- 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 Hierachy
  - Prediction
  - Intervention
  - Counterfactuals
  - Implications
- Graphical Models
  - Features
  - Implications
- Causality and Time

Assumptions (1)

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

Assumptions (2)

Error terms

Regression: Omittable 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 (3)

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. $A = f_A(B, U_A)$.

[1] For more information see Pearl (2009)

Assumptions (4)

Independence

Definition:

SCM Applications:

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

Comparative Causal Tools

Historical Development

- Path Analysis -> SEM -> SCM

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	· Unstable · Volatile to parameter changes · Re-Estimate entire model	· Stable · More Natural Specification · Only estimate Δ CM
Intervention	· Costly for Non-Markovian Models · Unstable(Nature CP) · Only generic estimates(Δ CP)	· Pot. Cyclic Representation · Stable(Nature Eq.) · Context specific(Invariance of Eq.)
Counterfactuals	· **Impossible** · no information on latent factors(ϵ)	· Possible · Inclusion of latent factors

Pearls Causal Hierachy

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

Prediction

Vanilla machine learning (ML), bayesian networks (BN) and regression models (Reg) are at the lowest level in the causal hierarchy (see table 1). These methods demand the least information and depend on association alone. Associational methods ignore external changes outside of our data. The interventional distribution has information on these external changes. The interventional distribution is only defined in high level causal methods.

Intervention

The second query deals with interventions. Here we can use Pearl (2009) 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 (Oberst and Sontag 2019). Causal bayesian networks , Markov Decision Processes (MDP) and reinforcement learning model intervention.

Counterfactuals

Process is described as follows:

a. Abduction: Cast probability $P(u)$ as conditional probability $P(u|\epsilon)$
b. Action: Exchange $(X = x)$
c. Prediction: Compute $(Y = y)$

Graphical Illustration

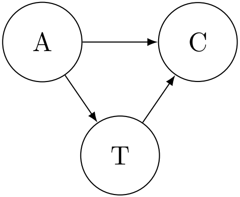


Figure 1: Probabilistic Model

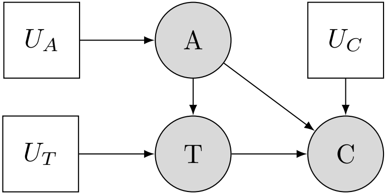


Figure 2: Structural Causal Model

Causality and Time

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

References

Oberst, Michael, and David Sontag. 2019. "Counterfactual Off-Policy Evaluation with Gumbel-Max Structural Causal Models." In *International Conference on Machine Learning*, 4881–90. PMLR.

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, Massachuestts: The MIT Press.