Pearl: Causation, Action, and Counterfactuals

Zeyn Saigol

IR Lab, School of Computer Science

University of Birmingham

24th March 2009

Overview

- Why is causality important
- Probability re-cap
- Causal Bayesian networks
- Inferring causal effects
- Actual causation
- Summary

Why Does Causality Matter?

- Central concept in most sciences – many physical laws describe cause-effect relationships.
- Also key in AI:
  - How should a robot learn causal relationships from interactions?
  - How can a robot make use of causal information provided to it?
- However, often not easy to reason with causality.
- E.g., 1. grass wet => must have rained

  2. break bottle => grass wet

  therefore, if I break bottle, must have rained

Aims of Pearl's Work

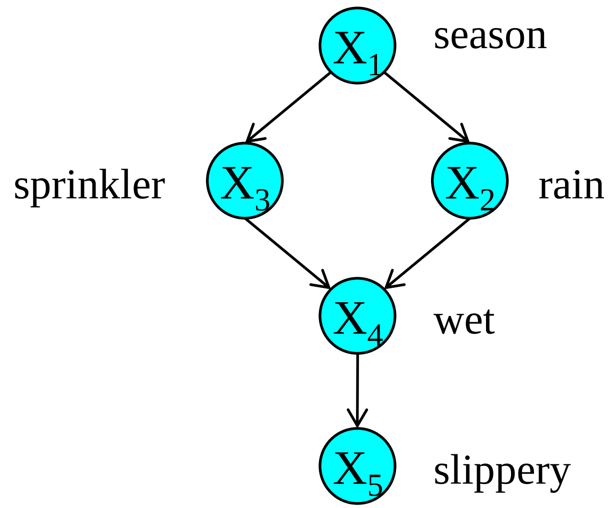
- (at least what I'm aiming to describe)
- Can't directly infer causation from data
  - Famous quote "correlation does not imply causation", e.g. getting dark later doesn't cause it to rain less
- Must start by assuming a causal model of the variables of interest.
- Then aim to:
  - Predict the effect of actions, even if data is not available from performing those actions.
  - Identify the "actual cause" of events.
  - Support counterfactual reasoning.

Probability Re-cap

- Notation: variables in uppercase (X), values lowercase (X=x)
- Conditional probability $P(X \mid Y) = \dfrac{P(X, Y)}{P(Y)}$
- Independence: $I(X,Y) \Leftrightarrow P(X|Y)=P(X)$
- Conditional independence: $I(X,Y|Z) \Leftrightarrow P(X|Y,Z)=P(X|Z)$
- These are *statistical* properties, and can be found from data.

Causal Model

- Consider two classes of variable: exogenous (values determined by factors outside the model) and endogenous (the ones we care about modelling)
- Causal model consists of
  - exogenous (often unobserved) variables $U_1 \ldots U_m$
  - endogenous variables $X_1 \ldots X_n$
  - n functions f_i determining X_i given values for all other variables



$X_1 = U_1$

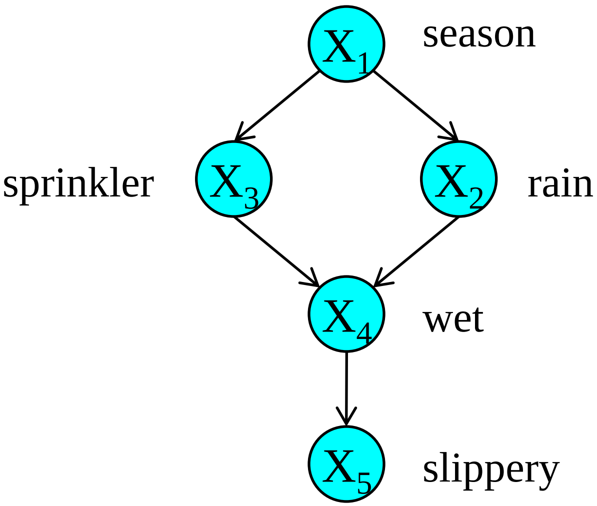
$X_2 = f_2\,(X_1, U_2)$

$X_3 = f_3\,(X_1, U_3)$

$X_4 = f_4\,(X_3, X_2, U_4)$ $\qquad$ $X_5 = f_5\,(X_4, U_5)$

Causal Bayesian Network

- Note lack of arrows is a particularly strong statement – e.g. in example, changing X_2 cannot have any affect on X_3



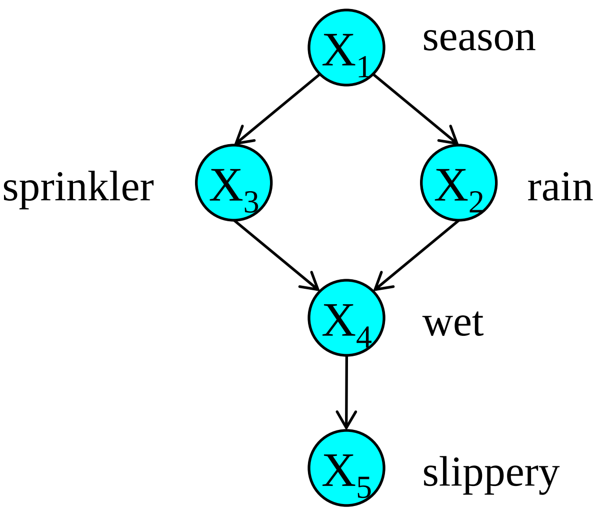
- Graph is a Bayesian network if it is Markovian:
  1. Can be drawn as a DAG (as above)
  2. All exogenous variables are mutually independent
- Allows joint distribution to be written from graph:

$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1)P(x_4 \mid x_2, x_3)P(x_5 \mid x_4)$$

- Each conditional prob can be estimated from data.

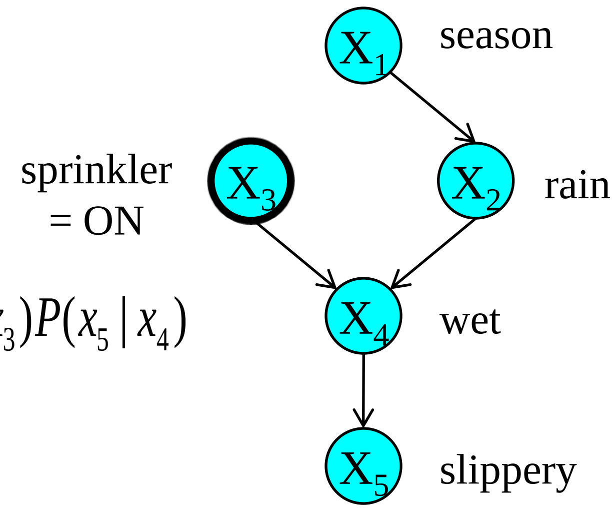
Effects of Actions

- Possible to reason about effects of actions without explicit data on them or knowledge of the forms of the functions f_i
- Action => forcing one of the variables to take a specific value. For example if set X_3, need to:
  - Delete all links to X_3 in the graph
  - Remove the term for X_3 in the distribution



$$P(x_1, x_2, x_3, x_4, x_5) =$$
$$P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1)P(x_4 \mid x_2, x_3)P(x_5 \mid x_4)$$

$$P(x_1, x_2, x_4, x_5 \mid do(x_3 = on)) =$$
$$P(x_1)P(x_2 \mid x_1)P(x_4 \mid x_2, X_3 = on)P(x_5 \mid x_4)$$

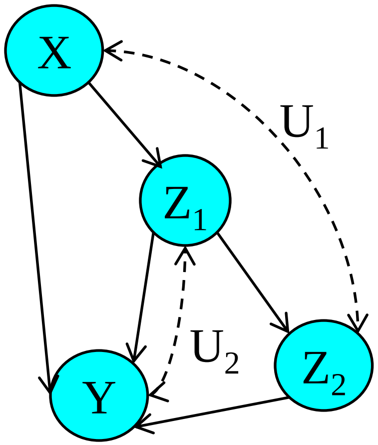


Identifiability of Causal Effects

- If all exogenous variables are known, P(y|do(x)), the effect of X on Y, can always be determined.
- However if some variables in the model are unobserved, the causal effect of X on Y is not always *identifiable*.
- Previous methods for determining this depended on a complex process of working out if paths between X and Y were blocked.
- Tian and Pearl (2002) introduces a simpler criterion.

Identifiability - Tian and Pearl

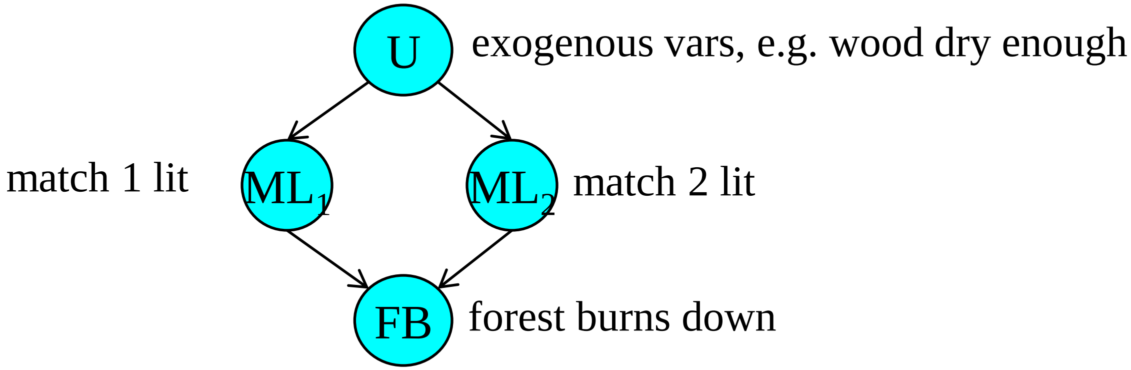
- Introduce graphical notation for unobserved variable that may affect two endogenous variables A and B: a dashed, bidirected line connecting A and B.



- Any nodes that are not ancestors of Y can be removed from the graph.
- Then, P(y|do(x)) is identifiable if there are no bidirected paths connecting X to any of its children.

Actual Causes

- How can we determine what actually causes an event?
- Example: two arsonists set fire to different parts of a forest. Either fire would be sufficient to burn down the whole forest – so which is the "actual cause" of the forest burning down?

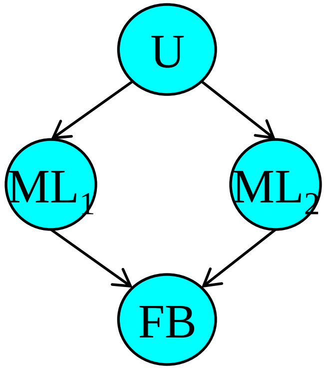


Actual Causes (cont'd)

- Halpern and Pearl (2005) show that an action X=x is the actual cause of event p if (loosely):
  1. X=x and p are true in the model
  2. Given a split of V into Z and W, where $X \subseteq Z$ and Z can be thought of as the "active causal process" X->p

     (a) Setting X=x' and W=w' entails ¬p

     (b) p remains true if X=x when W=w' and Z=z, where z is the original values of Z in the model
  3. X is minimal
- Example: disjunctive (one match suffices to ensure FB) and conjunctive (both matches needed)



- $X=\{ML_1\}$, $Z=\{ML_1, FB\}$, $W=\{ML_2\}$

Summary

- The effects of interventions can often be identified from non-experimental data.
- The conditions under which we can identify such effects can be found by simple graphical criteria.
- We can use graphical models to reason about the actual cause of an event.
- Similar causal effect analyses can support conterfactual reasoning, where the outcome of events that did not happen can be predicted.

References

- Pearl, J. (2003). Statistics and Causal Inference: A Review. *Sociedad de Estadıstica e Investigacion Operativa Test*, **12**(2):281–345.
- Tian, J., and Pearl, J. (2002). A General Identification Condition for Causal Effects. *Eighteenth National Conference on AI*.
- Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.
- Pearl, J. (1995). Causation, action, and counterfactuals. In A. Gammerman (Ed.), *Computational Learning and Probabilistic Reasoning*, John Wiley and Sons, New York, Chapter 15, 235–255
- Halpern, J.Y., and Pearl, J. (2005). Causes and explanations: A structural-model approach – Part I: Causes. *British Journal of Philosophy of Science*, **56**:843–887.