Reasoning about causal belief

Kaibo Xie

Institute for Logic, Language and Computation

July 27, 2018

Causal Logic in Halpern (2016)

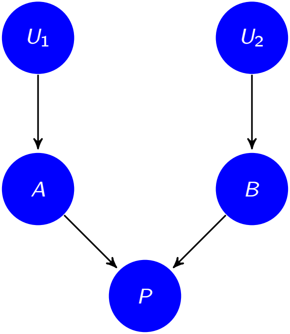
Signature

a signature is defined as a triple $(\mathcal{U}, \mathcal{V}, \mathcal{R})$ where $\mathcal{U}$ is the set of exogenous variables and $\mathcal{V}$ is the set of endogenous variables, and $\mathcal{R}$ is a function that indicates the range of possible values of each causal variables.

Causal Model

Given such a signature $\mathcal{S}$, a causal model is a pair $(\mathcal{S}, \mathcal{F})$ where $\mathcal{F}$ associates with every endogenous variable X a function denoted F_X which characterize the value of X given the value of all the other variables in $\mathcal{U} \cup \mathcal{V}$.

An Example



A and B stand for two assassins. P stands for whether the president is killed. U_1 and U_2 represent external factors that determine whether assassin A or B will shoot the president.

Exogenous variables: $\mathcal{U} = \{U_1, U_2\}$;

Endogenous variables:$\mathcal{V} = \{A, B, P\}$.

Structural equations: $A = U_1$, $B = U_2$, $P = A \vee B$

What if we prevent A from shooting

Suppose in the real world, A receives order to shoot the president and B does not, in this case the president is killed ($U_1 = A = 1$, $U_2 = B = 0$, $P = 1$). What will happen if we prevent A from shooting? Is the president alive in this case?

What if we prevent A from shooting

Suppose in the real world, A receives order to shoot the president and B does not, in this case the president is killed ($U_1 = A = 1$, $U_2 = B = 0$, $P = 1$). What will happen if we prevent A from shooting? Is the president alive in this case?

Intervention

1. Set the value of A to 0: replace $\mathcal{F}$ with $\mathcal{F}_{A=0}$ where $\mathcal{F}_{A=0}$ is the result of replacing the equation for A in $\mathcal{F}$ by $A = 0$ (by turning $\mathcal{F}_A$ into constant functions whose output is 0) and leaving the remaining equations untouched.
2. Check: whether in all possible solutions to the structural equations obtained after setting A to 0 (namely $\mathcal{F}_{A=0}$), $P = 0$ holds whenever $U_1 = 1$, $U_2 = 0$

Halpern's Causal Language

(Basic) causal formula

- A basic causal formula is in the form of $[Y_1 = y_1, ..., Y_k = y_k]\phi$ where $Y_1, ..., Y_k$ are distinct causal variables, and ϕ is a boolean combination of formulas in the form of $X = x$.
- A causal formula is a boolean combination of basic causal formulas.

Halpern's Causal Language

(Basic) causal formula

- A basic causal formula is in the form of $[Y_1 = y_1, ..., Y_k = y_k]\phi$ where $Y_1, ..., Y_k$ are distinct causal variables, and ϕ is a boolean combination of formulas in the form of $X = x$.
- A causal formula is a boolean combination of basic causal formulas.

For instance,
$[Y_1 = y_1, ..., Y_k = y_k]X = x$ means: in all possible solutions to the structural equations obtained after setting Y_i to y_i, $i = 1, ..., k$, The random variable X has value x

Reasoning about causal belief

1. If $Y = y$ had been the case, $X = x$ would been the case
2. It is believed that setting the value of Y to y results in X having the value x
3. After revising my belief state with $Z = z$, it is believed that setting the value of Y to y results in X having the value x

Epistemic operators

- Knowledge: $K\phi$ (the agent knows ϕ)
- Belief: $Bel\phi$ (The agent believes ϕ)
- Conditional Belief: $Bel^{\psi}\phi$ (The agent believes ϕ after revising its belief with ψ)

Epistemic operators

- Knowledge: $K\phi$ (the agent knows ϕ)
- Belief: $Bel\phi$ (The agent believes ϕ)
- Conditional Belief: $Bel^{\psi}\phi$ (The agent believes ϕ after revising its belief with ψ)

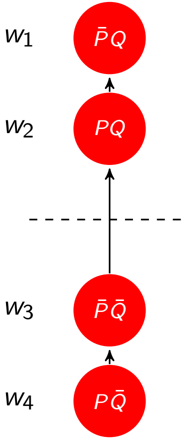
For instance $\wedge_{x\in\mathcal{R}(X)} \vee_{y\in\mathcal{R}(Y)} K[X = x](Y = y)$

Epistemic model (basic)

An epistemic model is a tuple $(W, \leqslant, \Pi)$

1. W is a set of possible worlds
2. $\leqslant$ is a plausibility ordering over W
3. Π is an information partition over W: for each $w \in W$, $\Pi(w)$ tells us which possible worlds are indistinguishable for the agent

Epistemic Model



$w_1 \leqslant w_2 \leqslant w_3 \leqslant w_4$
$\Pi(w_1) = \Pi(w_2) = \{w_1, w_2\}$; $\Pi(w_3) = \Pi(w_4) = \{w_3, w_4\}$

Causal Epistemic Model

A causal epistemic model is a tuple $\langle \mathcal{S}, \mathcal{F}, \Pi, \leq \rangle$

1. $\mathcal{S}$ is a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{R})$
2. $\mathcal{F}$ is a set of structural equations, for each $X \in \mathcal{V}$, $\mathcal{F}_X$ is a function from $(\times_{Z\in\mathcal{U}}\mathcal{R}(Z) \times (\times_{Y\in\mathcal{V}-\{X\}}\mathcal{R}(Y))$ to $\mathcal{R}(X)$. $\mathcal{F}$ has no causal loops.
3. Π is an information partition over W where $W = \times_{X\in\mathcal{U}\cup\mathcal{V}}\mathcal{R}(X)$.
4. $\leqslant \subset W \times W$ is a total pre-order on W satisfying certain constraints. $\leqslant$ is known as the plausibility ordering.

Combine the languages

Let $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, the language for $\mathcal{S}$, write $\mathcal{L}(\mathcal{S})$, is defined as follows: $X = x$ (if $X \in \mathcal{U} \cup \mathcal{V}$ and $x \in \mathcal{R}(X)$ $|\phi \wedge \psi| \neg\phi | [X_1 = x_1, ..., X_i = x_i]\phi$ (if $X_1 = x_1, ..., X_i = x_i$ is a sequence of distinct atomic sentences with $X_1, ..., X_i \in \mathcal{V}$ and ϕ is a formula without intervention operators$|Bel\phi \in \mathcal{L}(\mathcal{S})|Bel^{\psi}\phi|K\phi$

Semantics

Let $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$ be a signature and $M = \langle \mathcal{S}, \mathcal{F}, \Pi, \leqslant \rangle$ is a causal epistemic model.

Boolean Cases

Let w be a possible world in W in the form of $(y_1, ..., y_n)$.
$M, (y_1, ..., y_n) \models X_i = x_i$ $(1 \leqslant i \leqslant n)$ if and only if $x_i = y_i$.
The boolean combinations are defined in the usual way.

Epistemic Operators (Baltag and Smets (2006))

Belief

$M, w \models Bel\phi$ if and only if ϕ holds on the most plausible worlds in $\Pi(w)$

Believing ϕ means ϕ is true at the most plausible worlds

Conditional Belief

$M, w \models Bel^{\psi}\phi$ if and only if for any $s \in Min_{\leqslant}(\{t \in W | M, t \models \psi\} \cap \Pi(w))$, $M, s \models \phi$. It means that the agent has the conditional belief "given ψ, then ϕ" if and only if ϕ holds on the most plausible ψ worlds.

Believing ϕ conditional on ψ means ϕ is true at the most plausible ψ-worlds.

Find the counterpart of Halpern's intervention operator

We need to define the output of setting the value of X to x in $(\vec{u}, \vec{v})$, write $f_{X=x}((\vec{u}, \vec{v})) = (\vec{a}, \vec{b})$

The output of setting the value of X to x

$\mathcal{F}_{X=x}$ is defined as the result of replacing the equation for $X_1, ..., X_n$ in $\mathcal{F}$ by $X_1 = x_1, ...X_n = x_n$ (namely $\mathcal{F}_{X_1}, ..., \mathcal{F}_{X_n}$ become constant functions whose output are $x_1, ..., x_n$ and leaving the remaining equations untouched). Let $\vec{V} = V_1, ..., V_n$, $\vec{v} = v_1, ..., v_n$, $\vec{v'} = v'_1, ..., v'_n$, $\vec{b} = b_1, ..., b_n$. Define $(\vec{a}, \vec{b})$ as: $\vec{a} = \vec{u}$; for any $1 \leq i \leq n$, $b_i = v'_i$ if $X \rightsquigarrow V_i$, otherwise $b_i = v_i$.

$Y \rightsquigarrow Z$ means "Y affects Z" as an abbreviation for the formula $\bigvee_{\vec{X} \subset \mathcal{V}, \vec{x} \in \times_{X \in \mathcal{V}} \mathcal{R}(X), y \in \mathcal{R}(Y), \vec{u} \in \times_{U \in \mathcal{U}} \mathcal{R}(U), z \neq z' \in \mathcal{R}(Z)} ([\vec{X} = \vec{x}, Y = y] Z = z' \wedge [\vec{X} = \vec{x}, Y = y] Z = z)$

The intervention operator

The truth condition of the a sentence with intervention operator

Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$ and a causal epistemic model $M = \langle \mathcal{S}, \mathcal{F}, \Pi, \leq \rangle$. Let w be a possible world in W in the form of $(y_1, ..., y_n)$.
$M, w \models [\vec{X} = \vec{x}]\phi$ if and only if $M, f_{\vec{X}=\vec{x}}(w) \models \phi$ where $f_{\vec{X}=\vec{x}}$ is defined as before.

Plausibility ordering is not arbitrary

Belief should consistent with agents' causal information.

Constraint on the plausibility ordering

For any $w_1, w_2 \in W$ if $w_1 \prec w_2$ and $w_2 \not\prec w_1$ then $w_1 < w_2$ (where $w_1 \prec w_2$ is defined as there is $X \in \mathcal{V}$ such that w_1 complies $\mathcal{F}_X$ and w_2 does not.)

Applications of the combination in the philosophy of language

The distinction between indicative conditionals and subjunctive conditionals
$Bel[\vec{X} = \vec{x}]\phi$ does not imply $Bel^{\vec{X}=\vec{x}}\phi$

Example from Adams (1970)

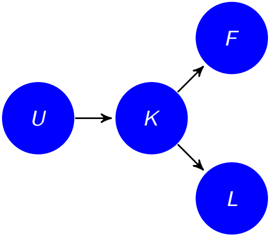
"If Oswald *had* not killed Kennedy, then someone else *would have*" does not imply "If Oswald *did* not kill Kennedy, then someone else *did*"

Subjunctive conditional

Kratzer's King Ludwig Example

King Ludwig of Bavaria likes to spend his weekends in Leoni Castle. Whenever the king is in the castle, the lights will be on and the royal flag will be up. A traveler watches the castle from a distance and sees that the lights are on. The flag, however, is not up. He says

If the flag had been up, the king would have been in the castle.



Nobody will agree that the king would be brought into the castle by hoisting the flag.

$[L = 1]K = 0$ is rejected in this model under the setting.

Condition of Acceptance in Two Readings

Ontic Reading

A counterfactual "if $\vec{X} = \vec{x}$ had been the case then $\vec{Y} = \vec{y}$ would have been the case" is accepted in the ontic reading (on the actual world) iff

$$Bel[\vec{X} = \vec{x}]\vec{Y} = \vec{y}$$

is true at the actual world.

Epistemic Reading

A counterfactual "if $\vec{X} = \vec{x}$ had been the case then $\vec{Y} = \vec{y}$ would have been the case" is accepted in the epistemic reading (at the actual world) iff

$$[\vec{X} = \vec{x}]Bel(\vec{Y} = \vec{y})$$

is true at the actual world.

Thank you!