Approches logiques de la causalité

Yves Moinard

Approches logiques de la causalité

Yves Moinard

INRIA Bretagne-Atlantique, IRISA, Rennes

LISE, journée de la causalité, ESCP Paris, 7 décembre 2011

Causalité en logique

- La causalité a très tôt été considérée comme une notion fondamentale, en particulier en philosophie
(voir SEP Standford Encyclopedia of Plilosophy).
Aristote et la logique médiévale (et donc la théologie) lui accordaient déjà une place très importante.
- La logique moderne et l'informatique ont donc très tôt étudié cette notion, en la distinguant de l'implication. La causalité n'est pas l'implication, même si elle est liée à l'implication.
- La causalité n'est pas non plus corrélation: 'La marée basse ne cause pas la marée haute 6 heures plus tard".
- Les seules approches considérées ici sont "purement logique".

Approche logique de la causalité

- Comment représenter des énoncés causaux, et qu'en déduire?
- On s'intéresse en particulier à la prédiction (quel est l'état final d'un système? quel est l'histoire du système?).
  Quels sont les modèles (description des états en général ou réponse à des questions plus précises) possibles.
- Au diagnostic (qu'est-ce qui a causé l'état actuel du système?)

Approche logique de la causalité

- La causalité: une relation entre états.
- La cause doit être ce qui produit l'effet.
  Une explosion à l'air libre cause du bruit.
- La "cause" produit toujours l'effet.
  (Exclut de nombreuses formes de causalité, par exemple exclut: fumer cause le cancer.)
- On n'aborde pas ici la notion temporelle (même si elle est indissociable de la notion de causalité). On évoquera juste une notion de succession, grâce à un indice entier qui croît de 1 (jamais plus) entre ce qui cause et l'effet.

Quelques notions générales classiques: le problème du cadre

- Problème très profond, dont les philosophes se sont emparés.
- La loi d'inertie (problème du cadre, frame problem). Définie dès les premières approches "logiques" de la causalité. Évocation (?) chez Bertrand Russel. Parmi ses cinq postulats: 'The Postulate of Quasi-permanence´. (1948, p. 487)
- Russel pas souvent cité pour la causalité car il a aussi écrit:

  > The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm. (Russell, 1913, p. 1).

- Dans toute description logique, il faut des formules qui expriment non seulement ce qui ***est*** causé, mais aussi ce qui ***n'est pas*** causé.
- Exige règles avec exceptions non nécessairement connues à l'avance.

Le problème du cadre (cf Murray Shanahan)

- Un exemple fameux (non donné ici): the Yale shooting problem (turquey...).

Un petit exemple:

- Peindre un objet X avec la couleur C cause que X a la couleur C.
- Déplacer X vers le lieu L cause que X se trouve en L.
- Situation initiale: A est “Rouge”, et se trouve dans la “Maison”.
- On peint A en "Bleu" puis on le déplace vers le “Jardin”.
- On s’attend bien sûr à avoir: A est Bleu et dans le Jardin. Avec les approches logiques classiques, ce n’est pas nécessairement le cas.
- On obtient “A est dans le Jardin”, mais déplacer A pourrait modifier sa couleur!

Le problème du cadre et les logiques non monotones

- Il faut expliciter, grâce à de formules logiques, le fait que déplacer un objet ne modifie pas sa couleur. Ce sont toutes ces formules qu'on appelle les ***axiomes du cadre***. Cela traduit la ***loi (de sens commun) de l'inertie***. Chacun sait cela, mais les approches logiques exigent de l'expliciter en formules
- Il faut donc une formule qui traduise en particulier:
  Si on déplace un objet X vers un lieu L, et si l'objet était de couleur C, alors il est de couleur C après le déplacement,
  et de même pour les déplacements.
- Mais ces règles sur les "non effet" (ce qui n'est ***pas causé***) ont de multiples exceptions, et on ne peut [ni ne veut] prévoir dès le départ toutes les exceptions possibles: déplacer un objet dans un pot de peinture cause une modification de couleur,...

Le problème du cadre et les logiques non monotones

- Il faut donc que certains faits vrais puissent devenir faux quand on ajoute des informations (des formules), ce qui est impossible en logique classique (dite donc ***monotone***).
- Il est donc souvent recommandé d'utiliser des méthodes logiques ***non-monotones***:
  - Logique des défauts de Reiter ($\sim$ 1980): règles d'inférence du genre si on a A et si on n'a pas non B (B est possible) alors on a C.
  - Circonscription de McCarthy ($\sim$ 1980): ajout d'axiomes à la logique classique qui minimisent (pour l'inclusion) l'extension de certaines formules (ici les exceptions).

Un exemple de propriétés d'opérateur causal

A causal theory of abduction, Alexander Bochman, J. of Logic and Computation 17(5), 2007

Un opérateur binaire $\Rightarrow$ est ***causal*** s'il satisfait:

- Renforcement à gauche: Si $A \models B$ et $B \Rightarrow C$ alors $A \Rightarrow C$.
- Affaiblissement à droite: Si $A \Rightarrow B$ et $B \models C$ alors $A \Rightarrow C$.
- Et: Si $A \Rightarrow B$ et $A \Rightarrow C$ alors $A \Rightarrow B \wedge C$.
- Ou: Si $A \Rightarrow C$ et $B \Rightarrow C$ alors $A \vee B \Rightarrow C$.
- Réflexivité du vrai et du faux: $\top \Rightarrow \top$, $\bot \Rightarrow \bot$.

Par rapport à l'implication classique, on n'a pas la réflexivité (presque tout le monde semble d'accord là-dessus) (sauf pour le vrai et pour le faux).
On n'a pas non plus la contraposition de l'implication classique:
Si $A \rightarrow B$ alors $\neg B \rightarrow \neg A$.
On n'a pas non plus la ***déduction faible***: Si $A \Rightarrow B$ alors $\top \Rightarrow A \rightarrow B$.
On n'a donc pas non plus ici: Si $A \Rightarrow B$ alors $A \rightarrow B$.

Logique de causalité universelle de McCain et Turner

Causal theory of action and change, Norman McCain and Hudson Turner (AAAI 1997,...)

- Il s'agit d'un formalisme fondateur (seulement évoqué ici). Ce formalisme assez puissant a donné un système qui tourne, CCALC (causal calculator) (cf site du Texas Action Group at Austin).
- ***Tout ce qui est causé est avéré (satisfait).***
- Le principe de causalité universelle affirme que la réciproque est vraie: ***Tout ce qui est avéré est causé***.
- Ce principe (très fort philosophiquement parlant) a l'avantage de la simplification de la théorie obtenue. Et il peut aisément être contourné pour certains faits (dit ***exogènes*** dans la théorie).
- Il n'est pas nécessaire de connaître les causes des faits (pour décrire une "histoire de faits"). Il suffit de connaître les conditions qui font que les faits sont "causés".

Logique de causalité universelle: un minuscule exemple

- $Q \Rightarrow P$. $Q \Rightarrow Q$. $\neg Q \Rightarrow \neg Q$.
- On n'a pas $\neg P$ (car rien ne peut le causer), on a donc P.
- Donc, P est "causé", ce qui exige d'avoir aussi Q.
- Donc, $\{P, Q\}$ est le seul modèle causal possible de cette petite théorie causale.

Notons qu'ici Q est ***exogène***, ce qui est traduit par les deux lois causales le concernant.
Remarquons que les formules logiques utilisées en interne pour traduire les lois causales utilisent un indice i.
Ainsi, $A \Rightarrow B$ sera traduit par $(A_i \wedge \cdots) \rightarrow B_{i+1}$.

- Une "loi causale" $A \Rightarrow B$ signifie: Si A est établi, alors B est causé. [On peut ainsi la traduire aussi avec un opérateur causal unaire $\mathtt{C}$ (au lieu du binaire $\Rightarrow$). La loi s'écrivant alors: $A \rightarrow \mathtt{C}B$.]

Logique de causalité universelle et programmation logique

- De nombreux papiers décrivant divers exemples, dont “le singe et la banane” avec ce formalisme, et en particulier en CCALC (traduction non donnée ici):
  Un singe veut atteindre une banane accrochée au plafond, et pour cela il doit déplacer une boîte sous la banane puis monter sur la boîte.
  On peut raffiner et introduire une boîte trop petite,...
- Ce formalisme se rapproche de la programmation logique, dans sa variante "ensembles-réponses" (answer set programming “ASP”). Les considérations qui on amené à ASP contiennent en effet des réflexions sur le raisonnement causal. Il est donc facile de traduire ces règles en règles d'un programme logique, au lieu d'un outil dédié comme CCALC.
  La minimisation des exceptions (axiome du cadre) se fait grâce à l'opérateur `not` (négation par défaut: `not A` est vrai si on ne peut pas déduire `A` dans l'ensemble réponse considéré).

Divers: Logique contrefactuelle et causalité

- Logique contrefactuelle et causalité:
  la causalité peut s'exprimer grâce à un opérateur modal ***contrefactuel***.
  On considère ici que si A cause B, alors, si on n'a pas B, on n'a pas A. (cf Lewis 1973 et 1986).
- Voici ce qu'en pense Peter Menzies (auteur de l'article à ce sujet dans SEP):

> D'intenses discussions ces trente dernières années ont remis en cause l'adéquation de la causalité en termes contrefactuels. De nombreux raffinements sont apparus afin de mieux s'approcher de la notion de causalité...

Un petit formalisme causal

- Les formalismes causaux ont l'inconvénient d'être plutôt complexes. Par exemple, Jerome Lang, Fangzhen Lin et Pierre Marquis on étudié la complexité d'un "noyau" de différents formalismes, et sont arrivés à la conclusion que cette complexité est grande, même pour des théories (données) causales très simples (Causal Theories of Action: A Computational Core, IJCAI 2003).
- Comme toujours, il y a un compromis à trouver entre expressivité et complexité. De nombreuses théories causale moins ambitieuses que celles présentées jusqu'ici existent.
- En voici une (collaboration avec Philippe Besnard et Marie-Odile Cordier). Elle suit les motivations présentées en début d'article, mais elle est d'expressivité limitée.

Le formalisme causal

Données

- **C**: ensemble de formules causales construites sur des atomes classiques et causaux comme $On(alarm)$ $cause$ $Heard(bell)$.
- **O**: ensemble d'atomes ontologiques: $soft\_bell \rightarrow_{est\_un} bell$, $Heard \rightarrow_{est\_un} Perceived$.
- **W**: ensemble de formules classiques (exprimant des faits incompatibles, co-occurrents,...): $Heard(soft\_bell) \rightarrow \neg Heard(loud\_bell)$.

Résultats

- **Atomes d'explications**: α *explique* β *car_poss* Φ: α *est une explication de* β **car** Φ *est* ***poss****ible.*

$[\{\alpha, \beta\} \cup \Phi$ ensemble d'atomes classiques concrets: $On(alarm)]$.

Le formalisme causal: prédicats et ontologie

Partant des liens IS-A links entre constantes ou prédicats, une relation IS-A$_{aux}$ entre atomes classiques est calculée. ▸ hurry up

- *Heard* essentiellement existentiel pour son paramètre: (***Heard some***) $soft\_bell \longrightarrow_{est\_un} \; bell \quad \rightsquigarrow \quad Heard(soft\_bell) \longrightarrow_{est\_un\_aux} Heard(bell)$.
- *I_Like* essentiellement universel pour son paramètre: (***I_Like all***):

$$\cdots \rightsquigarrow I\_Like(bell) \longrightarrow_{est\_un\_aux} I\_Like(soft\_bell).$$

- $Heard \longrightarrow_{est\_un} \; Perceive \rightsquigarrow Heard(bell) \longrightarrow_{est\_un\_aux} Perceive(bell)$.

Le formalisme causal: inférer les explications

Le cas general: "ontologie en cassis" $\nwarrow IS\!-\!Aaux \nearrow$

Si $\quad \alpha$ **cause** β, $\quad \gamma \rightarrow_{\text{est\_un}} \beta$, $\quad \gamma \rightarrow_{\text{est\_un}} \delta$,

alors $\quad \alpha$ **explique** δ **car_poss** $\{\alpha, \gamma\}$.

Exemple: $C = \{On(alarm)\ cause\ Heard(bell)\}$,
$O = \{loud\_bell \rightarrow_{est\_un} bell,\ loud\_bell \rightarrow_{est\_un} loud\_noise\}$.

Comme $W \not\models \neg(On(alarm) \wedge Heard(loud\_bell))$,
on déduit l'atome d'explication $On(alarm)\ explique\ Heard(loud\_noise)\ car\_poss\ \{On(alarm), Heard(loud\_bell)\}$

et a fortiori

$(On(alarm)\ explique\ Heard(loud\_bell)\ car\_poss\ \{On(alarm), Heard(loud\_bell)\})$, et
$(On(alarm)\ explique\ Heard(bell)\ car\_poss\ \{On(alarm)\})$.

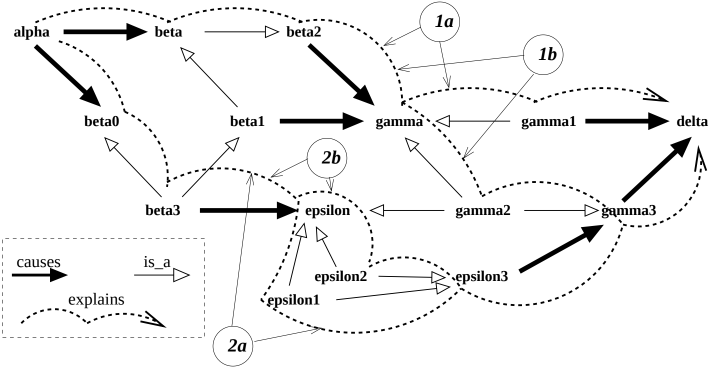
Transitivité de l'explication

L'explication est transitive (ou presque)
(Peu de changements si l'opérateur causal était transitif).

Schéma de transitivité de l'explication

Si $\left(\begin{array}{l} W \not\models \neg(\bigwedge_{\varphi \in \Phi \cup \Psi} \varphi), \\ \alpha \textit{ explique } \beta \textit{ car\_poss } \Phi \\ \beta \textit{ explique } \gamma \textit{ car\_poss } \Psi \end{array} \right)$ alors α *explique* γ *car_poss* $(\Phi \cup \Psi)$

Un exemple avec chemins multiples (quatre chemins d'explication optimaux de α vers δ)



1 α *explique* δ *car_poss* $\{\alpha, \gamma_1\}$, α *explique* δ *car_poss* $\{\alpha, \gamma_2\}$,

2 α *explique* δ *car_poss* $\{\alpha, \beta_3, \epsilon_1\}$, α *explique* δ *car_poss* $\{\alpha, \beta_3, \epsilon_2\}$.

Merci