Argumentation-based Causal and Counterfactual Reasoning

Lars Bengel[1], Lydia Blümel[1], Tjitze Rienstra[2], Matthias Thimm[1]

[1] Artificial Intelligence Group,
University of Hagen, Germany
[2] Department of Advanced Computing Sciences,
Maastricht University, The Netherlands

September 12, 2022

Definition

An argumentation framework is a pair $F = (\mathbf{A}, \Rightarrow)$ where $\mathbf{A}$ is a set whose elements are called *arguments* and where $\Rightarrow \subseteq \mathbf{A} \times \mathbf{A}$ is called the *attack relation*.

Definition

A set $E \subseteq \mathbf{A}$ is:

- *conflict-free* if for all $a, b \in E$ we have $a \not\Rightarrow b$.
- *stable* if E is conflict-free and for every $a \in \mathbf{A} \setminus E$ there is a $b \in E$ such that $b \Rightarrow a$.

Phan Minh Dung. “On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games”. In: *Artificial Intelligence* 77.2 (1995), pp. 321–358

Definition

A *causal model* is a triple (U, V, K) where U and V partition the set of atoms into, respectively, a set of background and explainable atoms. K consists of a set of *Boolean structural equations*, one for each atom $v \in V$. A Boolean structural equation for v is a formula of the form $v \leftrightarrow \phi$.

Definition

A *causal model* is a triple (U, V, K) where U and V partition the set of atoms into, respectively, a set of background and explainable atoms. K consists of a set of *Boolean structural equations*, one for each atom $v \in V$. A Boolean structural equation for v is a formula of the form $v \leftrightarrow \phi$.

Definition

A *causal knowledge base* is a knowledge base $\Delta = (K, A)$ where K is a causal model and where A is a set of *background assumptions*, at least one for each background atom. A *background assumption* for an atom u is a literal $l \in \{u, \neg u\}$.

Definition

Let $\Delta = (K, A)$ be a knowledge base. We define the *AF induced by* Δ $F(\Delta)$ as the AF $(\mathbf{A}, \Rightarrow)$ where

- $\mathbf{A}$ is the set of all pairs (Φ, ψ) such that
  - $\Phi \subseteq A$,
  - $\Phi \cup K \nvdash \bot$,
  - $\Phi \cup K \vdash \psi$, and if $\Psi \subset \Phi$ then $\Psi \cup K \nvdash \psi$.
- $\Rightarrow \subseteq \mathbf{A} \times \mathbf{A}$ such that $((\Phi, \psi), (\Phi', \psi')) \in \Rightarrow$ iff (Φ, ψ) *undercuts* (Φ', ψ'), i. e., for some $\phi' \in \Phi'$ we have $\phi' \equiv \neg\psi$.

Definition

Let $\Delta = (K, A)$ be a knowledge base. We define the *AF induced by* Δ $F(\Delta)$ as the AF $(\mathbf{A}, \Rightarrow)$ where

- $\mathbf{A}$ is the set of all pairs (Φ, ψ) such that
  - $\Phi \subseteq A$,
  - $\Phi \cup K \nvdash \bot$,
  - $\Phi \cup K \vdash \psi$, and if $\Psi \subset \Phi$ then $\Psi \cup K \nvdash \psi$.
- $\Rightarrow \subseteq \mathbf{A} \times \mathbf{A}$ such that $((\Phi, \psi), (\Phi', \psi')) \in \Rightarrow$ iff (Φ, ψ) *undercuts* (Φ', ψ'), i. e., for some $\phi' \in \Phi'$ we have $\phi' \equiv \neg\psi$.

Proposition

Let $\Delta = (K, A)$ be a knowledge base. Then $\phi \mid\!\sim_{\Delta} \psi$ if and only if every stable extension E of $F(K \cup \{\phi\}, A)$ contains an argument with conclusion ψ.

Claudette Cayrol. "On the Relation between Argumentation and Non-monotonic Coherence-Based Entailment". In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal Québec, Canada, August 20-25 1995, 2 Volumes*. Morgan Kaufmann, 1995, pp. 1443–1448

Example

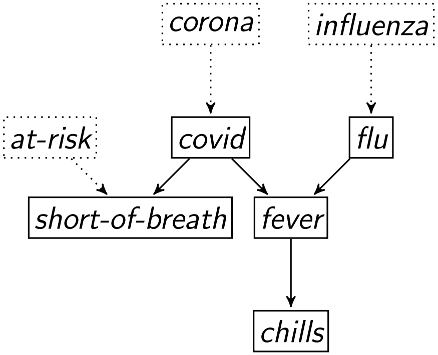
Example

Consider the *causal knowledge base* $\Delta = (K, A)$, where

$$K = \begin{array}{lcl} \mathit{covid} & \leftrightarrow & \mathit{corona} \\ \mathit{flu} & \leftrightarrow & \mathit{influenza} \\ \mathit{fever} & \leftrightarrow & \mathit{covid} \vee \mathit{flu} \\ \mathit{chills} & \leftrightarrow & \mathit{fever} \\ \mathit{short\text{-}of\text{-}breath} & \leftrightarrow & \mathit{covid} \wedge \mathit{at\text{-}risk} \end{array}$$

$$A = \{\mathit{at\text{-}risk}, \neg \mathit{at\text{-}risk}, \neg \mathit{corona}, \neg \mathit{influenza}\}$$

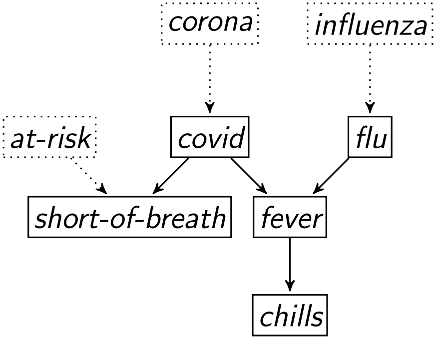
Example

$K =$ 

$$A = \{\mathit{at\text{-}risk}, \neg \mathit{at\text{-}risk}, \neg \mathit{corona}, \neg \mathit{influenza}\}$$

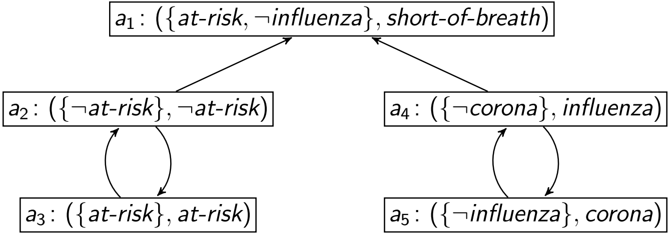
Example

$$fever \mid\!\sim_{\Delta} short\text{-}of\text{-}breath$$



Example

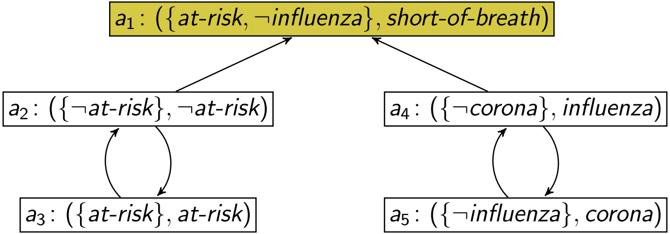
$$\textit{fever} \mid\!\sim_{\Delta} \textit{short-of-breath}$$



Causal Reasoning

Example

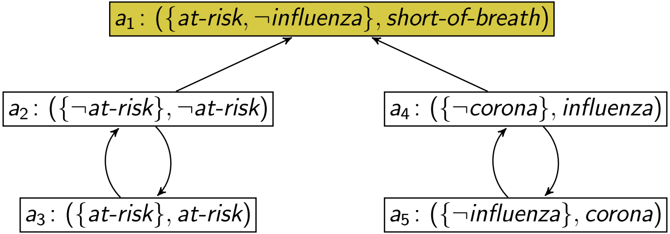
$$fever \mid\!\sim_{\Delta} short\text{-}of\text{-}breath$$



Causal Reasoning

Example

$$fever \mid\!\sim_{\Delta} short\text{-}of\text{-}breath$$



$\Longrightarrow$ *Given fever, shortness of breath is possible but not necessary.*

given ϕ, if v had been x then ψ would be true given ϕ, if v had been x then ψ would be true

Definition

The *twin model* for a causal model K is the causal model K^* defined by

$$K^* = K \cup \{(v^* \leftrightarrow \phi^*) \mid (v \leftrightarrow \phi) \in K\}$$

given ϕ, if *v* had been *x* then ψ would be true

Definition

The *twin model* for a causal model *K* is the causal model K^* defined by

$$K^* = K \cup \{(v^* \leftrightarrow \phi^*) \mid (v \leftrightarrow \phi) \in K\}$$

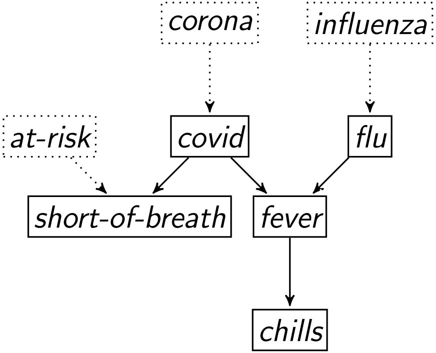
Example

Would the patient have had *fever* if we had administered a covid vaccine (i.e., if *covid* had been false)?

$$fever \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$

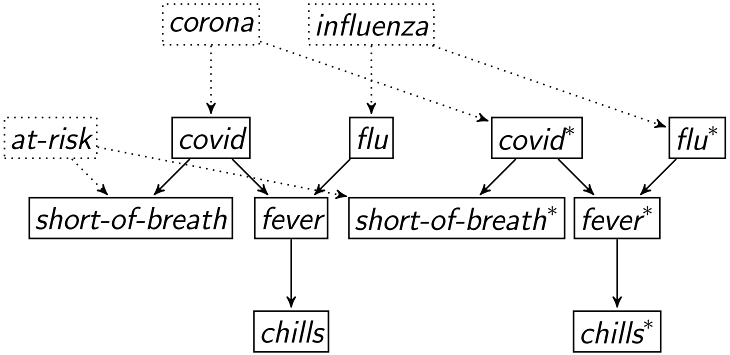
Example

$$fever \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$



Example

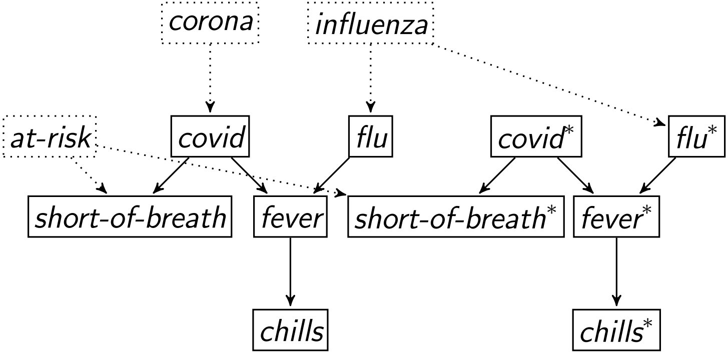
$$fever \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$



Counterfactual Statements I

Example

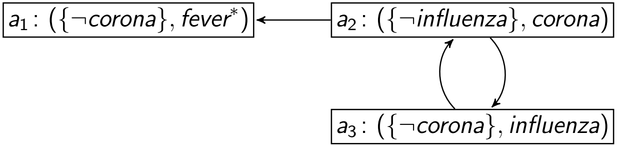
$$fever \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$



Counterfactual Statements II

Example

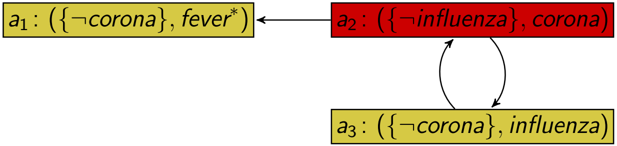
$$fever \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$



Counterfactual Statements II

Example

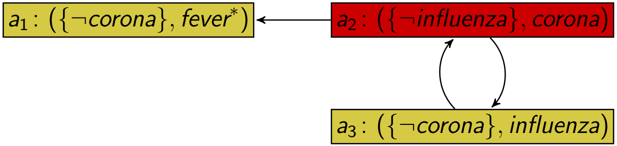
$$fever \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$



Counterfactual Statements II

Example

$$fever \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$



$\implies$ *The patient may or may not have had fever, had we administered a covid vaccine.*

Counterfactual Statements III

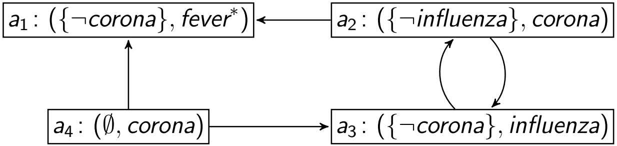
Example

$$\mathit{fever} \land \mathit{short\text{-}of\text{-}breath} \mathrel{|\!\sim}_{\Delta^*_{[\mathit{covid}^* = \bot]}} \mathit{fever}^*$$

Counterfactual Statements III

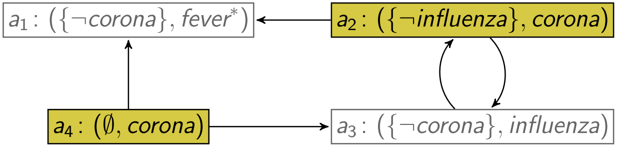
Example

$$\mathit{fever} \wedge \mathit{short\text{-}of\text{-}breath} \mid\!\sim_{\Delta^*_{[\mathit{covid}^*=\bot]}} \mathit{fever}^*$$



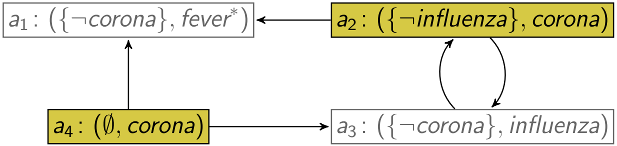
Example

$$fever \wedge short\text{-}of\text{-}breath \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$



Example

$$fever \wedge short\text{-}of\text{-}breath \mid\!\sim_{\Delta^*_{[covid^*=\bot]}} fever^*$$



$\Longrightarrow$ *The patient would not have had fever, had we administered a covid vaccine.*

Conclusion and Future Work

- We defined a model to transform causal knowledge bases into Dung-style AFs.
- The twin model method provides an alternative mechanism for answering counterfactual queries.
- The constructed AF can be used to provide argumentative explanations.

- We defined a model to transform causal knowledge bases into Dung-style AFs.
- The twin model method provides an alternative mechanism for answering counterfactual queries.
- The constructed AF can be used to provide argumentative explanations.

Open Questions

- Representing uncertain causal relations via probabilistic argumentation methods.