Informative Counterfactuals

Adam Bjorndahl (Cornell University, Mathematics) & Todd Snider (Cornell University, Linguistics)

NASSLLI 2014, Maryland

Overview

We use structural equation models (SEMs) to interpret counterfactuals

- SEMs represent dependencies between events
- Formally, counterfactuals denote sets of such dependencies
- Intuitively, these can be thought of as possible explanations
- We classify such explanations into four categories, providing a typology of explanatory strategies

Counterfactuals

- We use counterfactuals to talk about things we know to be false

  (1) If the movie hadn't been so boring, I wouldn't have fallen asleep.

- And to talk about things we're uncertain about

  (2) If Sam were angry, Pat would have been angry, too. (But I don't know if she was.)

- Counterfactuals describe some relationship between the events
- There are many ways for two events to be related

  (3) If Alice had gone to the party, Bob would have stayed home.

- Does Bob try to avoid Alice?
  - Maybe he's shy
  - Maybe he doesn't like her
- Do other circumstances prevent them from attending parties together?
  - Maybe they're a couple on a tight budget
  - Maybe Bob is actually Alice in disguise
- Does Alice try to avoid Bob?
  - Unlike the other scenarios, this one doesn't seem to jive with (3)
- To understand a counterfactual, we have to capture this range of relationships

Modeling Relationships

- To capture relationships between events, we use *structured possible worlds* (Starr 2014)
- Worlds are event variables, their values, and **dependencies** between them
  - Just like truth values, we can use the (non)existence of dependencies to discriminate among worlds
- We model these dependencies using Structural Equation Models as formalized in Pearl 2000
- Nodes as events, arrows as dependencies

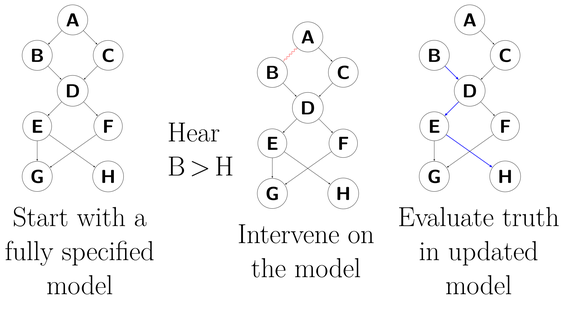
Rejecting Explanations

- There are many reasons to reject an explanation (including the implicated direct dependency)
- It might contradict prior knowledge
- It might violate a law of good explanations
  - e.g. by positing an effect temporally prior to its cause
- It might not satisfy the contextual parameter for specificity

Key Contrast

We think of agents as *building explanations* rather than evaluating truth in a fully specified model. As such, we take the SEM not as a *given* but as a *goal*.

Graph as given

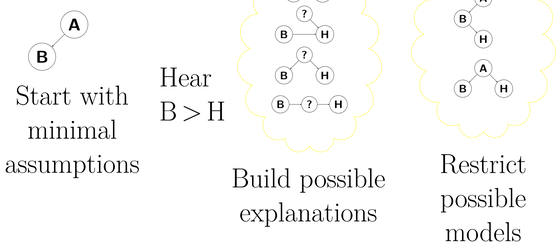


Start with a fully specified model

Intervene on the model

Evaluate truth in updated model

Graph as goal



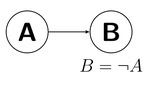
Start with minimal assumptions

Build possible explanations

Restrict possible models

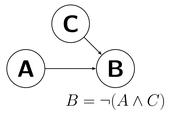
A Typology of Explanatory Strategies

Direct Cause



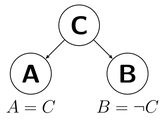
- A simple direct dependency
- The 'default' assumption
- Implicated by a counterfactual, can be canceled or strengthened

Additional Cause



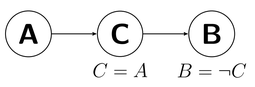
- Positing an additional causal factor
- A & B covary in the right C conditions
- Ex: Bob dislikes Alice

Common Cause



- Positing a shared cause
- No direct relation between A and B
- Ex: Coin flip to determine who attends

Intermediate Cause



- Positing a mediating factor
- A & B related, but not directly
- Ex: Bob is allergic to Alice's cat

Mutual Incompatibility

- Some counterfactuals which are individually felicitous are jointly infelicitous
- Consider a world where Alice and Bob are married, and live with their young son Doug

  (3) If Alice had gone to the party, Bob would have stayed home.

  (4) If Alice had gone to the party, Doug would have been home alone.

- Updating with (3) adds a covariance between A and $\neg B$ to our knowledge base
- Updating with (4) requires that A and B have the same value
- The models compatible with some explanation of (3) are not compatible with any explanation of (4)

Conclusion

- We can use structured possible worlds to model dependencies, and thus counterfactuals
- Doing so provides a natural way to typologize explanatory strategies
- Also yields insight into the mechanism that explains mutually infelicitous counterfactuals

References

Judea Pearl. *Causality: Models, Reasoning and Inference.* Cambridge Univ Press, 2000.

William B. Starr. Structured possible worlds. Ms. Cornell University, 2014.

Acknowledgements

Thanks to Will Starr, Sarah Murray, Christina Bjorndahl, the Cornell Semantics Group, & audiences at PHLINC2 and LGM.

Contact Information

- Web: http://conf.ling.cornell.edu/tsnider
- Email: tns35@cornell.edu