Changing Beliefs about Domain Dynamics in the Situation Calculus

Toryn Q. Klassen[1,2] Sheila A. McIlraith[1,2] Hector J. Levesque[1]

[1]Department of Computer Science, University of Toronto
[2]Vector Institute

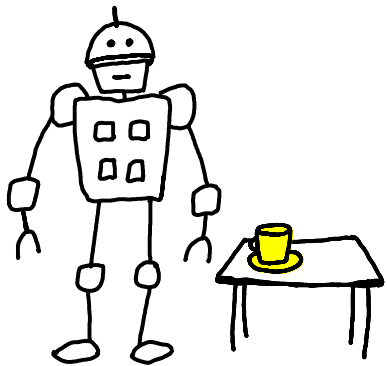
17th International Conference on Principles of Knowledge Representation and Reasoning, September 17, 2020

Introduction

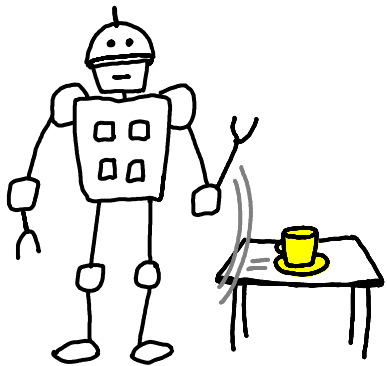
- An agent should be able to **change its beliefs** about the **dynamic** properties of actions
  - effects,
  - preconditions,
  - and sensing

  as a consequence of its observations of the world.
- We propose a way to conveniently **represent** domain dynamics in the situation calculus to support such belief change.
- We focus on how the specification can control how **general** of conclusions an agent draws from observations.

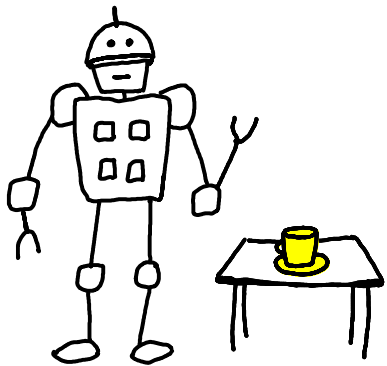
If I (try to) pick up anything, I will be holding it.



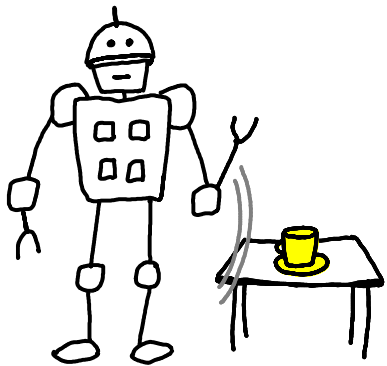
If I (try to) pick up anything, I will be holding it.



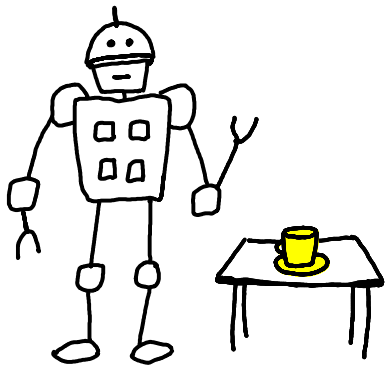
If I pick up anything, I will be holding it – with the exception of that one cup that one time.



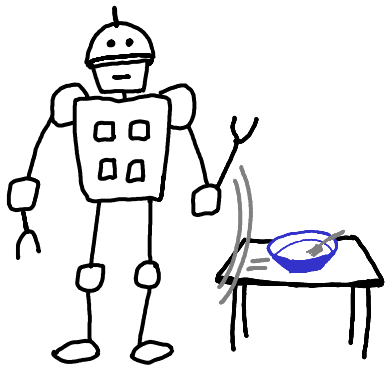
If I pick up anything, I will be holding it – with the exception of that one cup that one time.



If I pick up anything, I will be holding it, unless it's that cup.



If I pick up anything, I will be holding it, unless it's that cup.



If I pick up anything, I will be holding it as long as it's not slippery (and those two objects were slippery).

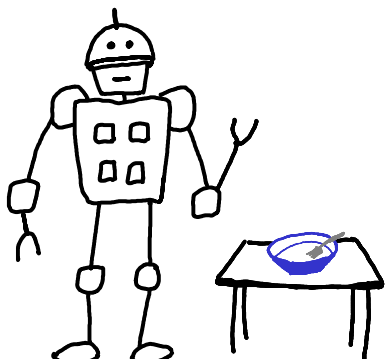


Table of Contents

Background
- The situation calculus
- Belief and plausibility

Beliefs about domain dynamics

Patterns

Conclusion

The situation calculus

- The situation calculus is a language of **second-order logic**.
- Situations represent **histories** of actions. Time is modelled as a **branching** structure.
- Properties that can vary between situations are represented using **fluents**, predicates that take a situation argument.
  - For example, $\text{Holding}(x, s)$ might represent whether an agent is holding object x in situation s.
- An environment can be described in the situation calculus with a set of axioms, an **action theory**.
- Sometimes, an action theory as a whole is taken to represent the knowledge of the agent, but we'll be modelling **beliefs** explicitly.

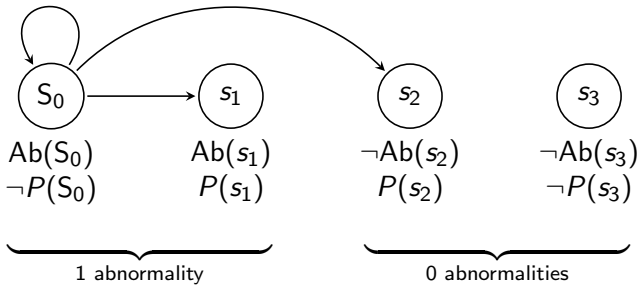
Belief

- The standard way of describing beliefs or knowledge in logic is in terms of **possible worlds**.
- An **accessibility relation** relates world w to world v if in w the agent considers that v may be the actual world.
- An accessibility relation can be encoded in classical logic, using **situations** as the "possible worlds".

Plausibility

- To specify how beliefs can change and be retracted over time, **further structure** beyond the possible worlds model is needed.
- Shapiro et al. (2011) defined belief (in the situation calculus) as what is true in all the **most plausible** accessible situations.
  - **Sensing** cause incompatible situations to become **inaccessible**, potentially replacing the set of most plausible accessible situations.
  - This allows for beliefs to be **revised**.
- In previous work, we measured plausibility by counting the number of **abnormal** atomic formulas true in a situation (Klassen et al., 2018).
  - This is related to **cardinality-based circumscription** (Liberatore and Schaerf, 1997; Sharma and Colomb, 1997; Moinard, 2000).

Example



- The **accessible** situations (from S_0) are those situations s in which $\neg \mathrm{Ab}(s) \supset P(s)$ is true.
- The set of **most plausible accessible** situations is $\{s_2\}$.
- $P(s)$ is true at each most plausible accessible situation s, so P is **believed** by the agent in S_0.

Action theories

- an axiom describing the agent's **initial accessibility relation**
- **initial state axioms**, describing the actual initial situation
- **successor state axioms (SSAs)**, specifying for each fluent how its value in a situation relates to the previous situation
  - for each **abnormality** fluent Ab_i, the SSA is

$$\mathrm{Ab}_i(\vec{x}, \mathrm{do}(a, s)) \equiv \mathrm{Ab}_i(\vec{x}, s)$$

- **precondition axioms**
- **sensing axioms**, describe how the agent can gain information from actions
- and others

Table of Contents

Background
- The situation calculus
- Belief and plausibility

Beliefs about domain dynamics

Patterns

Conclusion

Beliefs about domain dynamics

- A theory's axioms describing dynamics – SSAs, preconditions, and sensing – apply to **all** situations, and so to all accessible situations.
- Therefore, the theory entails that the agent **always believes** them.
- However, the agent **also believes** that other axioms are equivalent to the ones in the theory,
  - and which axioms the agent believes are equivalent may **change** over time.

Example

Suppose an action theory **includes** the SSA

$$\text{Holding}(x, \text{do}(a, s)) \equiv (a = \text{pick}(x) \wedge \neg \text{Ab}_1(s)) \vee \text{Holding}(x, s).$$

Under some conditions, the agent will assume Ab_1 is **false**. Then it **also believes** another SSA,

$$\text{Holding}(x, \text{do}(a, s)) \equiv (a = \text{pick}(x) \wedge \neg \text{False}) \vee \text{Holding}(x, s),$$

which can be simplified to

$$\text{Holding}(x, \text{do}(a, s)) \equiv a = \text{pick} \vee \text{Holding}(x, s).$$

Example continued

If the agent later comes to think that Ab_1 is **actually true**, the agent will **now believe**

$$\mathsf{Holding}(x, \mathsf{do}(a, s)) \equiv (a = \mathsf{pick}(x) \wedge \neg\mathsf{True}) \vee \mathsf{Holding}(x, s)$$

which can be simplified to

$$\mathsf{Holding}(x, \mathsf{do}(a, s)) \equiv \mathsf{Holding}(x, s).$$

Table of Contents

Background
- The situation calculus
- Belief and plausibility

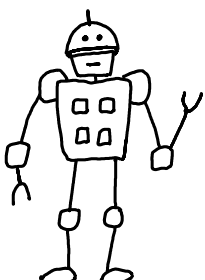
Beliefs about domain dynamics

Patterns

Conclusion

A slightly more complicated example

If I pick up anything,
I will be holding it,
unless it's that cup.



Consider this SSA:

$$\text{Holding}(x, \text{do}(a, s)) \equiv$$
$$(a = \text{pick}(x) \wedge \neg \text{Ab}_2(x, s)) \vee \text{Holding}(x, s)$$

Intuitively:

- Suppose the agent comes to believe that $\text{Ab}_2(\text{c}, \textit{now})$ is true of a **particular object** c
  - (e.g., by observing that Holding does not become true of c when expected).
- Then the agent will conclude that **all actions** will fail to make Holding true of c.

Patterns

Exceptional objects

If I pick up anything, I will be holding it, unless it's that cup.

Exceptional classes

If I pick up anything, I will be holding it as long as it's not slippery.

One-time exceptions

If I pick up anything, I will be holding it – with the exception of that one cup that one time.

Patterns

- Each of the three forms corresponds to a different **pattern** of referring to abnormalities within an SSA.
- Multiple patterns can be **combined** in one SSA, to support iterated belief changes.
- The paper formalizes the robot example using the SSA

$$\begin{aligned}&\mathsf{Holding}(x, \mathsf{do}(a, s)) \equiv \\ &\quad [(a = \mathsf{pick}(x) \wedge \neg \bigvee\nolimits_i \mathsf{Imp}_i(a, x, s)) \vee \mathsf{Holding}(x, s)]\,,\end{aligned}$$

where $\bigvee_i \mathsf{Imp}_i(a, x, s))$ is an abbreviation for

$$\mathsf{Ab}_1^2(\mathsf{history}(s), x, a, s) \vee \mathsf{Ab}_2^3(x, s) \vee [\mathsf{Slippery}(x, s) \wedge \mathsf{Ab}_3^4(s)]$$

(Superscripts are the **weights** given to the abnormalities; see paper for details.)

Table of Contents

Background
- The situation calculus
- Belief and plausibility

Beliefs about domain dynamics

Patterns

Conclusion

Conclusion

We've presented an approach to modelling **changing beliefs about domain dynamics** in the situation calculus, using action theories that assign **plausibility** to situations by **counting abnormalities**.

In the paper, we

- describe the **patterns** for writing SSAs;
- have some more general results about changing beliefs about domain dynamics; and
- describe how to apply **regression rewriting** with our theories, including how to use **beliefs about dynamics** within the regression procedure, and prove its correctness.

References

Toryn Q. Klassen, Sheila A. McIlraith, and Hector J. Levesque. Specifying plausibility levels for iterated belief change in the situation calculus. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018*, pages 257–266, 2018.

Paolo Liberatore and Marco Schaerf. Reducing belief revision to circumscription (and vice versa). *Artificial Intelligence*, 93(1):261–296, 1997. doi: 10.1016/S0004-3702(97)00016-7.

Yves Moinard. Note about cardinality-based circumscription. *Artificial Intelligence*, 119(1):259–273, 2000. doi: 10.1016/S0004-3702(00)00018-7.

Steven Shapiro, Maurice Pagnucco, Yves Lespérance, and Hector J. Levesque. Iterated belief change in the situation calculus. *Artificial Intelligence*, 175 (1):165–192, 2011. doi: 10.1016/j.artint.2010.04.003.

Nirad Sharma and Robert Colomb. Towards an integrated characterisation of model-based diagnosis and configuration through circumscription policies. Technical Report 364, Department of Computer Science, University of Queensland, 1997.