Formalizing Counterfactual and Nondeterministic Actions in First-Order Logic

Charles Elkan
Department of Computer Science and Engineering
University of California, San Diego

elkan@cs.ucsd.edu

ABSTRACT: This short paper outlines some aspects of a method of axiomatizing actions and their effects in standard first-order logic. The method is closely related to other methods using nonmonotonic logics, but the use of first-order logic permits remarkable simplicity. In particular, there is a single, explicit, fixed, and intuitively meaningful frame axiom. Moreover, complex types of inference are possible, including counterfactual reasoning and reasoning about nondeterministic actions.

Introduction

Given the intended audience of this paper, we dispense with a traditional introduction. The context in which we work is the standard situation calculus ontology [McCarthy and Hayes, 1969]. Our axiomatizations use three basic predicates: $holds$, $causes$, and $cancels$. The arguments of $holds$ are a fluent and a state, while the arguments of $causes$ and $cancels$ are an action, a state, and a fluent, as in $causes(stack(a,b), s_0, on(a,b))$. The single, fixed, general frame axiom is

$$\forall a,s,p \quad holds(p, do(a,s)) \leftrightarrow \qquad (1)$$
$$causes(a,s,p) \vee (holds(p,s) \wedge \neg cancels(a,s,p)).$$

This axiom has a commonsense interpretation. It states that a fluent holds in the state resulting from an action if and only if the action "causes" the fluent, or the fluent held before the action, and the action does not "cancel" it. Note that both the $causes$ and $cancels$ predicates have state arguments, so whether or not an action influences a fluent can depend on the state in which the action is taken.

We can use the world of the Yale shooting problem [Hanks and McDermott, 1986] as an example.[1] In this world there are three fluents, $loaded$, $alive$, and $dead$, and three actions, $load$, $shoot$, and $wait$. (Different papers on the Yale shooting problem use slightly different sets of fluents. The ones used here are from the original circumscriptive attempt to solve the problem [Hanks and McDermott, 1986]. Our solution does not depend on this choice of fluents.) The relationships of these fluents and actions can be specified by the following axioms:

$$\forall a,s,p \quad causes(a,s,p) \leftrightarrow \qquad (2)$$
$$(a = load \wedge p = loaded) \vee$$
$$(a = shoot \wedge p = dead \wedge holds(loaded, s))$$
$$\forall a,s,p \quad cancels(a,s,p) \leftrightarrow \qquad (3)$$
$$(a = shoot \wedge p = alive \wedge holds(loaded, s)) \vee$$
$$(a = shoot \wedge p = loaded \wedge holds(loaded, s)).$$

Axioms (1)–(3) describe the Yale shooting world. They can be used to solve many different inference problems, always using the standard semantics of first-order predicate calculus. For example, in all models of sentences (1)–(3) conjoined with $holds(alive, s_0)$, it is the case that

$$holds(dead, do(do(do(s_0, load), wait), shoot))$$

is true. The shooting problem is thus solved.

Backwards and counterfactual reasoning

The so-called murder mystery is a backwards reasoning problem introduced by Baker [1989], The scenario is that Fred is alive initially but not after shooting the gun and then waiting:

$$holds(alive, s_0) \ \wedge$$
$$\neg holds(alive, do(do(s_0, shoot), wait)).$$

The question is to discover when Fred died, and whether the gun was initially loaded. Axioms (1)–(3) and the sentence immediately above entail

$$holds(loaded, s_0) \ \wedge \ cancels(shoot, s_0, alive).$$

The mystery is solved.

Kowalski and Sergot [1986] invented the event calculus in order to handle a type of reasoning about action that they call narrative reasoning. There are three important issues in modeling narrative reasoning: distinguishing between actions that actually occur and actions that are merely envisaged, allowing or precluding in a flexible way the possibility of unknown actions before or after known actions, and representing knowledge about the times and durations of actions. The long version of this paper shows how to solve the first two issues above, and extending the situation calculus to include explicit time has been discussed by Miller and Shanahan [1994] among others. We choose here to discuss counterfactual reasoning, an extension of narrative reasoning whose formalization is called an open problem by Kowalski and Sadri [1994].

Counterfactual reasoning is reasoning about actions that did not actually occur, following the pattern

[1] It is worth stressing that the contribution of this paper is not yet another "solution" to the Yale shooting problem. The contribution is a general method for axiomatizing the effects of actions. The Yale shooting problem is used as an expository example simply because it is well-known. Note also that there is no attempt here to state all claims with the maximum degree of generality.

> Supposing that an event e_1 had happened resulting in outcome p, what would have been the outcome if event e_2 had happened instead?

A question of this form must be refined in two ways for it to be answerable by a standard inference engine using a logical theory. First, the question must be changed from being open-ended into a true-false question: "would the outcome have included the fluent q?". Second, the premise must be made precise as saying "supposing the state of the world were s such that p would hold if e_1 happened in s". Then, the question becomes

$$T \vdash \forall s\ holds(p, do(e_1, s)) \ \rightarrow\ holds(q, do(e_2, s))$$

In words, is it provable that for any state s if e_1 leads to p holding then e_2 leads to q holding? For example, some legal codes make a person guilty of murder only if it is provable that had they not acted, the victim would have stayed alive. In the Yale shooting scenario, suppose there are witnesses to the shooting but not to any previous actions. The shooting is then murder only if

$$\begin{aligned} T \vdash \forall s\ holds(dead, do(shoot, s)) \ \rightarrow& \\ holds(alive, do(wait, s)).& \end{aligned}$$

This is not provable. However, if a witness also observed that Fred was alive before the shooting, then it was murder:

$$\begin{aligned} T \vdash \forall s\ holds(alive, s)\ \wedge\ holds(dead, do(shoot, s)) \ \rightarrow& \\ holds(alive, do(wait, s)).& \end{aligned}$$

Actions with nondeterministic effects

Representing knowledge about actions whose effects are nondeterministic poses a challenge to most formalisms for reasoning about action. In order to illustrate our solution to this problem, we will use a scenario due to Reiter which is discussed by Shanahan [1994]. The knowledge to express formally is that moving an object onto a chessboard nondeterministically causes the object to be either on a white square, on a black square, or on both at once.

From a technical point of view, the problem is to find a refined axiomatization of the $causes$ predicate that allows desired conclusions to be inferred. In particular, we want

$$\begin{aligned} T \vdash causes(move, s, w) \vee causes(move, s, b)& \\ T \not\vdash causes(move, s, w)& \\ T \vdash \neg causes(move, s, red)& \end{aligned}$$

where red is an example of an arbitrary irrelevant fluent.

In order to represent knowledge about nondeterministic effects, we use the notions of "maybe causing" and "actually causing." The idea is that a nondeterministic action only "maybe" causes an effect, and for an effect to be inferrable, the action must also "actually" cause it. This idea is captured formally by adding an alternative to the overall $causes$ axiom:

$$\begin{aligned} \forall a, s, p \quad & causes(a, s, p)\ \leftrightarrow \\ & \qquad \vdots \\ & \vee \Big(mcauses(a, s, p) \wedge acauses(a, s, p) \Big) \end{aligned}$$

The chessboard move action can now be axiomatized as follows:

$$\begin{aligned} \forall a, s, p \quad & mcauses(a, s, p)\ \leftrightarrow \\ & \qquad \vdots \\ & \vee (a = move \wedge p = white) \\ & \vee (a = move \wedge p = black) \\ & \qquad \vdots \\ \forall s \quad & acauses(move, s, white) \vee acauses(move, s, black) \end{aligned}$$

Now the axioms above entail

$$\begin{aligned} \vdash \forall s\ causes(move, s, white) \vee causes(move, s, black)& \\ \vdash \forall s\ \neg causes(move, s, red)& \\ \neg holds(white, s_0)\ \wedge\ holds(white, do(move, s_0))& \\ \vdash acauses(move, s_0, white)& \\ \wedge\ causes(move, s_0, white)& \end{aligned}$$

It is important to note that the the axioms for $causes$ and $mcauses$ are biconditionals, while the axiom for $acauses$ is not. Informally speaking, a biconditional is the completion of a set of logic program definite clauses. A definite clause can be used to generate a conclusion constructively and deterministically from its antecedents. If one of the antecedents of a definite clause is not provable, the clause can still be used if this antecedent is assumed abductively. In logic programming with abduction, integrity constraints restrict what may consistently be assumed [Satoh and Iwayama, 1992]. In the example above, the $acauses$ axiom is a type of constructive integrity constraint: it says that for all s at least one of $acauses(move, s, white)$ and $acauses(move, s, black)$ must be assumed. Alternative integrity constraints could specify that at most one or exactly one of these facts can be true.

Discussion

The sophistication of the ideas for reasoning about action presented above should not be underestimated. The ideas used in this paper appear simple because they are worked out in the context of standard first-order logic. However, many of the ideas are essentially the same conceptually as those used in other recent papers on formalizing commonsense knowledge about actions and their effects. The ideas appear significantly more complicated in other papers because they are implemented using nonmonotonic logics.

An alternative approach to axiomatizing the effects of actions in first-order logic is due to Reiter [1991], building on work by Pednault [1989] and Schubert [1989]. Our method is considerably simpler, for two main reasons. First, Reiter uses multiple frame axioms, each one involving quantification over actions and states. A single frame axiom is sufficient above, because it involves quantification over fluents in addition. Second, no distinction is made in this paper between the preconditions under which an action is possible and the preconditions which must hold for it to have a particular effect. The long version of this paper shows how this distinction can be made if desired without changing the frame axiom (1).

The idea of using a $cancels$ predicate in addition to a $causes$ predicate is not common in work based on the situation calculus, but it is fundamental in the so-called event calculus of Kowalski and Sergot [1986]. In the event calculus $causes$ is called $initiates$ and $cancels$ is called $terminates$. Giving the $causes$ and $cancels$ predicates a state argument means to write $causes(a, s, p)$ rather than $causes(a, p)$. Baker [1991] claims that using a $causes$ predicate severely restricts expressiveness because the context-dependence of effects cannot be captured. In fact this is true only if the $causes$ predicate, as in [Lifschitz, 1987], does not have a state argument. With a state argument one can represent the fact that a given action has a certain effect only in specific contexts.

Perhaps the most basic idea underlying our approach is to use a minimization operator for a limited purpose only, specifically to implement epistemic closed world assumptions. We do not try to use bidirectional implication or any other variety of minimization operator to select desired outcomes of temporal projection, or for any other type of domain inference. Instead, we just use minimization operators to capture assertions that all the instances of particular predicates have been fully stated, i.e., that instances of these predicates not stated to be true are false. In other words, we use minimization operators to implement a communication convention, but we do not use them for any of the other purposes suggested by McCarthy [1980].

Surprisingly there is a close connection between this type of epistemic minimization and the method of Baker [1991]. Following McCarthy and other researchers, Baker asserts the axiom

$$\neg Ab(f, a, s) \ \rightarrow \ (Holds(f, Result(a, s)) \ \equiv \ Holds(f, s)).$$

and uses circumscription to minimize the extent of the Ab predicate while allowing the $Result$ function to vary. Let us call $Ab(f, a, s)$ for a fixed ground f, a, and s an abnormality fact. In general the truth of an abnormality fact can imply that the situation $Result(a, s)$ is identical to another situation, where without the abnormality fact the two situations would be different. Therefore, in some scenarios the truth of an apparently unwarranted abnormality fact can eliminate the need for an apparently needed situation, and hence eliminate the need for an apparently needed abnormality fact.

To ensure that an added abnormality fact can never eliminate the need for another abnormality fact, Baker uses so-called "existence of situations" axioms to ensure that all minimal models of his axioms have the same complete universe of situations. Then circumscribing the Ab predicate has the intended effect of minimizing the set of all (f, a, s) triples where $Holds(f, Result(a, s)) \not\equiv Holds(f, s)$.

The connection between Baker's approach and our approach is that minimizing the set of abnormal state transitions is the same as minimizing the set of $causes$ and $cancels$ facts. It is a theorem that

$$\begin{aligned}\vdash \neg causes(a, s, f) \ \wedge \ \neg cancels(a, s, f) \ \rightarrow \\ (holds(f, do(a, s)) \ \equiv \ holds(f, s)),\end{aligned}$$

so we may identify $Ab(f, a, s)$ and $causes(a, s, f) \vee cancels(a, s, f)$.

Acknowledgement

Many of the ideas here arose in discussions during a visit to Imperial College in 1994.

References

[Baker, 1989] Andrew B. Baker. A simple solution to the Yale shooting problem. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning (KR'89)*, pages 11–20, 1989.

[Baker, 1991] Andrew B. Baker. Nonmonotonic reasoning in the framework of the situation calculus. *Artificial Intelligence*, 49:5–23, 1991.

[Hanks and McDermott, 1986] Steve Hanks and Drew McDermott. Default reasoning, nonmonotonic logics, and the frame problem. In *Proceedings of the National Conference on Artificial Intelligence*, pages 328–333, August 1986.

[Kowalski and Sadri, 1994] Robert Kowalski and Fariba Sadri. The situation calculus and event calculus compared. In *International Logic Programming Symposium*, 1994.

[Kowalski and Sergot, 1986] Robert A. Kowalski and Marek Sergot. A logic-based calculus of events. *New Generation Computing*, 4:67, 1986.

[Lifschitz, 1987] Vladimir Lifschitz. Formal theories of action. In Frank M. Brown, editor, *Proceedings of the Workshop on the Frame Problem in Artificial Intelligence*, pages 35–58, Lawrence, Kansas, 1987. Morgan Kaufmann Publishers, Inc.

[McCarthy and Hayes, 1969] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence*, volume 4, pages 463–502. Edinburgh University Press, 1969.

[McCarthy, 1980] John McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1,2):27–39, April 1980. Special Issue on Non-Monotonic Logic.

[Miller and Shanahan, 1994] Rob Miller and Murray Shanahan. Narratives in the situation calculus. *Journal of Logic and Computation*, 1994. To appear.

[Pednault, 1989] Edwin P. D. Pednault. ADL: Exploring the middle ground between STRIPS and the situation calculus. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning (KR'89)*, pages 324–332. Morgan Kaufmann Publishers, Inc., 1989.

[Reiter, 1991] Raymond Reiter. *The Frame Problem in the Situation Calculus: A Simple Solution (Sometimes) and a Completeness Result for Goal Regression*, pages 359–380. Academic Press, 1991.

[Satoh and Iwayama, 1992] K. Satoh and N Iwayama. A query evaluation method for abductive logic programming. In Krzysztof R. Apt, editor, *Joint International Conference and Symposium on Logic Programming*, pages 671–685. MIT Press, 1992.

[Schubert, 1989] Lenhart K. Schubert. Monotonic solution of the frame problem in the situation calculus: An efficient method for worlds with fully specified actions. In *Knowledge Representation and Defeasible Reasoning*, pages 23–67. Kluwer Academic Publishers, 1989.

[Shanahan, 1994] Murray Shanahan. A circumscriptive calculus of events. To appear, 1994.