Counterfactual Reasoning by Means of a Calculus of Narrative Context

Rolf Nossum[1] and Michael Thielscher[2]

[1] Department of Mathematics
Agder College, Kristiansand (Norway)
[2] Department of Computer Science
Dresden University of Technology (Germany)

Abstract. The basic Event Calculus is extended by a Calculus of Narrative Context, to allow for reasoning about counterfactuals. Different hypothetical courses of events are distinguished by their narrative contexts. A notion of information transfer between contexts provides a basis for drawing conclusions about counterfactual courses of events.

1 Introduction

The Event Calculus [3] is an axiomatization technique for formalizing, and mechanizing, reasoning about narratives that involve actions and events. A typical axiomatization consists of two parts. Firstly, knowledge is provided about the general effects of the actions and events that might occur in a domain. Secondly, a specific narrative is described, to which the general knowledge of effects is applied in order to draw reasonable conclusions about these particular situations.

As regards this second part, plain Event Calculus only allows specifying events that actually occurred and observations that were actually made during any particular course of events. It does not allow for specification and reasoning about counterfactual courses of events. The lack of support for specification of counterfactual developments of the world in plain Event Calculus distinguishes it from its grand old rival, the Situation Calculus.

In this paper, we propose an extension of Event Calculus which allows formalizing and reasoning about hypothetical courses of events. This is accomplished by formally attaching event occurrences and observations to differing contexts, each of which describes alternative evolutions. The resulting theory inherits the representational merits of Event Calculus as regards reasoning about narratives, and combines it with the paradigm of Situation Calculus which supports reasoning about hypothetical sequences of actions.

2 Example: a Shooting Scenario

Let's enlist an old standby as our running example: a variation on the Yale Shooting Scenario [2]. Suppose we know that, in general, shooting at a vase causes it to shatter, provided the gun is loaded. Likewise, shooting at a turkey with the gun loaded always kills it. Suppose further a specific narrative telling us that initially the vase is in one piece and the turkey is alive. The narrative continues with the information that by shooting at the vase it has been destroyed. A reasonable conclusion here would be that had the protagonist shot at the turkey instead, then the bird would not have survived. For the only explanation of the vase being destroyed is that the gun was loaded to begin with. Hence the supposed death of the turkey in the counterfactual course of events.

In a formalization of the scenario we shall introduce two contexts, in one of which the vase is shot at while the turkey is the target in the other context. Enabling conclusions about counterfactual sequences of events requires a suitable transfer between different contexts: The straightforward conclusion that the gun must have been loaded initially in the context where the vase is destroyed, needs to be transferred to the opposing, hypothetical context.

The solution we propose is to stipulate that two contexts share all temporal conclusions up to the first timepoint at which the two courses of events split. In particular, all contexts agree on the information that holds in the initial state. This will support the expected conclusion as regards the above example.

3 Basic Definitions

- We take actions as atomic, e.g.: *shoot-turkey*, *shoot-vase*, etc.
- Time will be linear, and for simplicity we take the natural numbers as our time domain.
- Events are represented using the predicate $Happens(a, t)$, where a and t are an action and a timepoint, respectively.
- Properties are atomic, e.g. *alive*, *loaded*, *broken*, etc.
- Claims about initial truth or falsity of properties take the form $Initially(p)$, resp. $\neg Initially(p)$ for a property p.
- A (Narrative) Context is a set of $Initially(\ldots)$ and $Happens(\ldots, \ldots)$ atoms.
- A Narrative is a narrative context along with a conjunction of *Initiates*, *Terminates*, and *HoldsAt* formulas (see Section 4).

For a given narrative context N, we define initial segments N^t up to given timepoints t as, $N^t \stackrel{\text{def}}{=} N \setminus \{Happens(a, u) : t < u\}$. The latest timepoint at which two narrative contexts N_1 and N_2 agree is,

$$N_1 \uparrow N_2 \stackrel{\text{def}}{=} max(\{t : N_1^t = N_2^t\} \cup \{-1\})$$

4 The Event Calculus

A property f being true (resp. false) at some time t is expressed as $HoldsAt(f, t)$ (resp. $\neg HoldsAt(f, t)$). Axiomatizing the general knowledge as to the effects of actions and events is based on two predicates named $Initiates(a, f, t)$ and

$\mathit{Terminates}(a, f, t)$, indicating that action a occurring at time t causes property f to become true (resp. false). For our running example domain, we have

$$\mathit{HoldsAt}(\mathit{loaded}, t) \rightarrow \mathit{Initiates}(\mathit{shoot\text{-}vase}, \mathit{broken}, t) \tag{1}$$

$$\mathit{HoldsAt}(\mathit{loaded}, t) \rightarrow \mathit{Terminates}(\mathit{shoot\text{-}turkey}, \mathit{alive}, t) \tag{2}$$

In addition, the special predicate $\mathit{Initially}(f)$ serves the purpose of introducing partial information of the initial state, e.g.

$$\neg \mathit{Initially}(\mathit{broken}) \wedge \mathit{Initially}(\mathit{alive}) \tag{3}$$

The following foundational axioms describe the impact of initiating and terminating properties as a consequence of events happening.

$$\mathit{Initially}(f) \wedge \neg \mathit{Clipped}(0, f, t) \rightarrow \mathit{HoldsAt}(f, t) \tag{4}$$

$$\neg \mathit{Initially}(f) \wedge \neg \mathit{Declipped}(0, f, t) \rightarrow \neg \mathit{HoldsAt}(f, t) \tag{5}$$

$$\begin{aligned} &\mathit{Happens}(a, t_1) \wedge \mathit{Initiates}(a, f, t_1) \wedge t_1 < t_2 \wedge \neg \mathit{Clipped}(t_1, f, t_2) \\ &\quad \rightarrow \mathit{HoldsAt}(f, t_2) \end{aligned} \tag{6}$$

$$\begin{aligned} &\mathit{Happens}(a, t_1) \wedge \mathit{Terminates}(a, f, t_1) \wedge t_1 < t_2 \wedge \neg \mathit{Declipped}(t_1, f, t_2) \\ &\quad \rightarrow \neg \mathit{HoldsAt}(f, t_2) \end{aligned} \tag{7}$$

An instance $\mathit{Clipped}(T_1, F, T_2)$ (resp. $\mathit{Declipped}(T_1, F, T_2)$) is true iff an action occurs in the time interval (T_1, T_2) terminating (resp. instantiating) property F. Accordingly these two predicates are defined as follows:

$$\mathit{Clipped}(t_1, f, t_2) \leftrightarrow \exists a, t\, [\, \mathit{Happens}(a, t) \wedge \mathit{Terminates}(a, f, t) \wedge t_1 < t < t_2\,] \tag{8}$$

$$\mathit{Declipped}(t_1, f, t_2) \leftrightarrow \exists a, t\, [\, \mathit{Happens}(a, t) \wedge \mathit{Initiates}(a, f, t) \wedge t_1 < t < t_2\,] \tag{9}$$

Suppose given a narrative specification consisting of a conjunction N of *Happens* and *Initially* atoms along with a set of *HoldsAt* formulas and a conjunction E of *Initiates* and *Terminates* formulas. Then the semantics of this axiomatization is given by circumscribing [4] *Happens* in N and, independently, simultaneously circumscribing *Initiates* and *Terminates* in E. Along with the general axioms plus some suitable unique name assumptions, the resulting classical formula is taken as the meaning of the specification.

As an example, let N consist of the atoms in (3) plus $\mathit{Happens}(\mathit{shoot\text{-}vase}, 2)$. Circumscribing *Happens* in N yields

$$\begin{aligned} &[\, \mathit{Happens}(a, t) \leftrightarrow a = \mathit{shoot\text{-}vase} \wedge t = 2\,] \\ &\wedge \neg \mathit{Initially}(\mathit{broken}) \wedge \mathit{Initially}(\mathit{alive}) \end{aligned} \tag{10}$$

Circumscription of *Initiates* and *Terminates* in $E = \{(1), (2)\}$ yields

$$\begin{aligned} \mathit{Initiates}(a, f, t) &\leftrightarrow a = \mathit{shoot\text{-}vase} \wedge f = \mathit{broken} \wedge \mathit{HoldsAt}(\mathit{loaded}, t) \\ \mathit{Terminates}(a, f, t) &\leftrightarrow a = \mathit{shoot\text{-}turkey} \wedge f = \mathit{alive} \wedge \mathit{HoldsAt}(\mathit{loaded}, t) \end{aligned} \tag{11}$$

Let Σ denote the formulas (10) and (11) along with $\mathit{HoldsAt}(\mathit{broken}, 3)$ and the general axioms (4)–(9). Then Σ entails $\mathit{Declipped}(0, \mathit{broken}, 3)$ according to $\neg\mathit{Initially}(\mathit{broken})$, $\mathit{HoldsAt}(\mathit{broken}, 3)$, and formula (5). Formula (9) in conjunction with (10) then implies $\mathit{Initiates}(\mathit{shoot\text{-}vase}, \mathit{broken}, 2)$. This in turn entails $\mathit{HoldsAt}(\mathit{loaded}, 2)$ following Equation (11). It follows that $\mathit{Initially}(\mathit{loaded})$ according to (5), provided that $\neg\mathit{Declipped}(0, \mathit{loaded}, 2)$. The latter holds according to equation (9) in conjunction with (11).

5 Counterfactuals by Information Transfer

So far so good. Notice, however, that we cannot formalize an alternative course of events with, say, $\mathit{Happens}(\mathit{shoot\text{-}turkey}, 2)$ instead of $\mathit{Happens}(\mathit{shoot\text{-}vase}, 2)$ without losing the implicitly derived information that $\mathit{Initially}(\mathit{loaded})$. Hence conclusions about counterfactual events are not supported by plain Event Calculus. In order to facilitate reasoning about hypothetical courses of events, we amalgamate the Event Calculus with a calculus of Narrative Context.

The basic idea is to consider a particular course of events N as a narrative context. That is, formally a Context is a set of formulas of the form $\mathit{Happens}(a, t)$ or $\mathit{Initially}(f)$. E.g., for our key example we use the two contexts

$$N_1 = \{\neg\mathit{Initially}(\mathit{broken}), \mathit{Initially}(\mathit{alive}), \mathit{Happens}(\mathit{shoot\text{-}vase}, 2)\}$$

$$N_2 = \{\neg\mathit{Initially}(\mathit{broken}), \mathit{Initially}(\mathit{alive}), \mathit{Happens}(\mathit{shoot\text{-}turkey}, 2)\}$$

The notation $\mathbf{ist}(c, \psi)$, with the reading that ψ is true in the context c, was introduced in [1], where the application is to localized contexts in the CYC knowledge base. For our purposes here, c will range over narrative contexts.

All foundational axioms, i.e., (4)–(9) are then universally true in any context. The same applies to effect descriptions as in (11).

To see why information transfer between contexts is necessary, observe that in our example we can derive $\mathbf{ist}(N_1, \mathit{Initially}(\mathit{loaded}))$, as shown above. This, however, does not *per se* imply that $\mathbf{ist}(N_2, \mathit{Initially}(\mathit{loaded}))$. The latter is required in order that the intended conclusion $\mathbf{ist}(N_2, \neg\mathit{HoldsAt}(\mathit{Alive}, 3))$ follows.

The crucial connection is this: we demand that for any two narratives N_1 and N_2 we have

$$t \leq N_1 \uparrow N_2 \rightarrow [\, \mathbf{ist}(N_1, \mathit{HoldsAt}(f, t)) \leftrightarrow \mathbf{ist}(N_2, \mathit{HoldsAt}(f, t)) \,]$$

References

1. R. V. Guha. *Contexts: A Formalization and Some Applications.* PhD thesis, Stanford University, 1991.
2. S. Hanks and D. McDermott. Nonmonotonic logic and temporal projection. *Artificial Intelligence*, 3:379–412, 1987.
3. R. Kowalski and M. Sergot. A logic-based calculus of events. *New Generation Computing*, 4:67–95, 1986.
4. V. Lifschitz. Circumscription. In D. Gabbay etal., editors, *Handbook of Logic in AI and Logic Programming*, volume 3, pages 298–352. Oxford University Press, 1994.