CSE 591 - FALL 04. HANDOUT 7. VERSION 0.4

Chitta Baral

Department of Computer Science and Engineering
Arizona State University
Tempe, AZ 85287-5406 USA
chitta@asu.edu

http://www.public.asu.edu/~cbaral/

November 18, 2004

DELIBERATIVE ARCHITECTURE: ACCOUNTING FOR OBSERVATIONS

Recap

- Domain description: Define a transition function Φ from actions and states to sets of states.
  - Effect of actions: a **causes** f **if** $p_1, \ldots, p_m$
  - Causality: $p_1, \ldots, p_n$ **causes** f
  - Executability conditions: **executable** a **if** $q_1, \ldots, q_r$.
  - $\Phi(a, s) = \emptyset$ means a is not executable in the state s.
- Observations: f **after** $a_1, \ldots, a_n$.
  - These observations (except of the kind **initially** f) are actually hypothetical statements.
  - Together with a domain description (or a Φ) they define possible initial states.
  - Real observations vs hypothetical statements (or oracles)
    - a **causes** f; b **causes** g; f **after** a; g **after** b is consistent.
    - But only one of a and b happened in the initial state!

A language for observations

- Motivation:
  - Initially the car was parked. The next day the car was gone. Conclusion: The car must have been stolen.
  - I am at home, i have a car, and my suitcase is unpacked. I need to go to the airport with a packed suitcase. I make the plan to pack and then drive to the airport. After packing I *observe* that my car is no longer there. I can not just execute the remaining part of my old plan. I need to make a *new plan* from the *current situation* I am in.
  - I left home and my light switch was working. I came back and turned on the light and it does not work. I need to *diagnose* what went wrong; but not just a surface level diagnosis about what component is faulty but also a deeper diagnosis explaining what actions caused the fault.
- Syntax of the language:
  - f **at** s
  - s_1 **preceeds** s_2

– α **occurs_at** a

– α **between** $\mathsf{s}_1, \mathsf{s}_2$

– s and s_is are situations (or time points), α is a sequence of actions.

- Example 1:

  – $O = \{has\_car$ **at** $\mathsf{s}_0; \neg has\_car$ **at** $\mathsf{s}_1; \mathsf{s}_0$ **preceeds** $\mathsf{s}_1\}$

  – $D = \{steal\_car$ **causes** $\neg has\_car\}$

  – From (D, O) we want to conclude that $steal\_car$ **between** $\mathsf{s}_0, \mathsf{s}_1$.

- Example 2:

  – $O_1 = \{at\_home$ **at** $\mathsf{s}_0; has\_car$ **at** $\mathsf{s}_0; \neg packed$ **at** $\mathsf{s}_0; \}$

  – $D_1 = \{pack$ **causes** $packed$; **executable** $pack$ **if** at_home;
  $steal\_car$ **causes** $\neg has\_car$; $rent\_car$ **causes** has_car;
  $drive$ **causes** $at\_airport$; **executable** $drive$ **if** has_car;
  at_home **causes** $\neg at\_airport$; $at\_airport$ **causes** $\neg at\_home\}$

  – From (D_1, O_1) we can conclude that $at\_airport, packed$ **after** $pack, drive$.

  – I do the packing.
  $O_2 = O_1 \cup \{pack$ **occurs_at** $\mathsf{s}_1; \mathsf{s}_0$ **preceeds** $\mathsf{s}_1\}$.

- But then i notice that my car is gone.
  $O_3 = O_2 \cup \{\neg has\_car$ **at** $\mathsf{s}_2; \mathsf{s}_1$ **preceeds** $\mathsf{s}_2\}$
- From (D_1, O_3) I can no longer conclude that $at\_airport, packed$ **after** $drive$.
- But from (D_1, O_3) I should be able to conclude that
  $at\_airport, packed$ **after** $rent\_car; drive$.

- Query language: The language of observations plus
  f **after** $a_1, \ldots, a_n$ **at** s.
  - If $\mathsf{s} = \mathsf{s}_C$, the current situation then we just write
    f **after** $a_1, \ldots, a_n$
  - If $n = 0$ then we write f **at** s.
  - If $n = 0$ and $\mathsf{s} = \mathsf{s}_C$ then we write **currently** f.
  - If $n = 0$ and $\mathsf{s} = \mathsf{s}_0$ then we write **initially** f.
- s_C is a special constant and observations are not allowed with respect to this constant.
- A pair (D, O) will now be referred to as a narrative.

Semantics of narratives

- Defining Φ
  - An action a is said to be possibly executable in a state s w.r.t. a D if there exists **executable** a **if** $q_1, \ldots, q_r$ in D such that $q_1, \ldots, q_r$ hold in s.
  - $\Phi(a, s) = \emptyset$ if a is not possibly executable in s.
  - Else, $\Phi(a, s) = \{s' \ : \ s' = Cn_R((s \cap s') \cup E_a(s))$ and s' is an interpretation $\}$.
- *Valid states* (w.r.t. D): States that are closed under the static causal rules in D.
- Semantics is defined in terms of a *causal model* Ψ and a *situation assignment* Σ.
- A *causal model* Ψ of a domain description D is a function from action sequences to valid states such that
  - $\Psi([\,])$ is a valid state.
  - If $\Phi(a, \Psi(\alpha)) \neq \emptyset$ then $\Psi(\alpha \circ a) \in \Phi(a, \Psi(\alpha))$ else $\Psi(\alpha \circ a)$ is undefined.

  Intuitively Ψ selects one of the possible trajectories defined by Φ as a possible evolution of the world.

- A *situation assignment* with respect to D is a mapping Σ from situations into the set of action sequences of D that satisfy the following properties:
  - $\Sigma(\mathsf{s}_0) = []$;
  - for every $\mathsf{s} \in \mathbf{S}$, $\Sigma(\mathsf{s})$ is a prefix of $\Sigma(\mathsf{s}_C)$.
- An *interpretation* M of (D, O) is a pair (Ψ, Σ), where Ψ is a causal model of D, Σ is a situation assignment, and $\Sigma(\mathsf{s}_C)$ is defined.
- For an interpretation $M = (\Psi, \Sigma)$ of (D, O):
  - α **occurs_at** s is true in M if the sequence $\Sigma(\mathsf{s}) \circ \alpha$ is a prefix of $\Sigma(\mathsf{s}_C)$;
  - α **between** $\mathsf{s}_1, \mathsf{s}_2$ is true in M if $\Sigma(\mathsf{s}_1) \circ \alpha = \Sigma(\mathsf{s}_2)$;
  - f **at** s is true in M if f holds in $\Psi(\Sigma(\mathsf{s}))$;
  - s_1 **preceeds** s_2 is true in M if $\Sigma(\mathsf{s}_1)$ is a prefix of $\Sigma(\mathsf{s}_2)$.
- An interpretation $M = (\Psi, \Sigma)$ is a *model* of a narrative (D, O) if:

  (i) observations in O are true in M;

  (ii) there is no other interpretation $M' = (\Psi, \Sigma')$ such that M' satisfies condition (i) above and $\Sigma'(\mathsf{s}_C)$ is a subsequence of $\Sigma(\mathsf{s}_C)$.

Observe that these models are minimal in the sense that they exclude extraneous actions.

For a model $M = (\Psi, \Sigma)$, $\Sigma(\mathsf{s}_C)$ is referred to as the *actual line* w.r.t. M. It encodes how the world has progressed till the current situation.

- A narrative is *consistent* if it has a model. Otherwise, it is *inconsistent*.
- A query of the form φ **after** α **at** s is true in a model $M = (\Psi, \Sigma)$ if φ is true in $\Psi(\Sigma(\mathsf{s}) \circ \alpha)$.
- We already defined when other kind of queries (i.e., observations) are true in a model.
- A query q is entailed by a narrative (D, O), denoted by $(D, O) \models q$, if q is true in every model of (D, O).

The stolen car example revisited

- $D = \{steal\_car$ **causes** $\neg has\_car\}$
- $O = \{has\_car$ **at** $\mathsf{s}_0; \neg has\_car$ **at** $\mathsf{s}_1; \mathsf{s}_0$ **preceeds** $\mathsf{s}_1\}$
- The semantics:
  - States: $\emptyset$, $\{has\_car\}$.
  - $\Phi(steal\_car, \emptyset) = \emptyset$; $\Phi(steal\_car, \{has\_car\}) = \emptyset$.
  - $\Psi([]) = \{has\_car\}$; $\Psi(\alpha \circ steal\_car) = \emptyset$.
  - $\Sigma(\mathsf{s}_0) = []$; $\Sigma(\mathsf{s}_1) = [steal\_car]$; $\Sigma(\mathsf{s}_C) = [steal\_car]$.
  - (D, O) has only one model $M = (\Psi, \Sigma)$.
  - Why is (Ψ, Σ') not a model?
    (where $\Sigma'(\mathsf{s}_0) = []$; $\Sigma'(\mathsf{s}_1) = [steal\_car, steal\_car]$;
    $\Sigma'(\mathsf{s}_C) = [steal\_car, steal\_car]$.)
- $D, O \models steal\_car$ **between** $\mathsf{s}_0, \mathsf{s}_1$
  because $steal\_car$ **between** $\mathsf{s}_0, \mathsf{s}_1$ is true in the only model M of (D, O)
  because $\Sigma(\mathsf{s}_1) \circ steal\_car = \Sigma(\mathsf{s}_0)$.

An example with multiple models.

- $O = \{\neg f, \neg g$ **at** s_0; f **at** s_1; s_0 **preceeds** s_1; a_4 **occurs_at** $\mathsf{s}_1\}$
- $D = a_1$ **causes** g; a_2 **causes** f **if** g; a_3 **causes** g; a_4 **causes** $\neg g$ **if** $f\}$
- What are the models of (D, O)? What are the actual lines? Does $(D, O) \models$ **currently** $\neg g$?
- Φ and Ψ are straightforward.

actions	a_1	a_2	a_3	a_4
states				
$s_1 = \{\neg f, \neg g\}$ $s_2 = \{f, \neg g\}$ $s_3 = \{\neg f, g\}$ $s_4 = \{f, g\}$				

- $\Sigma_1(\mathsf{s}_0) = []$, $\Sigma_1(\mathsf{s}_1) = [a_1, a_2]$, $\Sigma_1(\mathsf{s}_C) = [a_1, a_2, a_4]$.
- $\Sigma_2(\mathsf{s}_0) = []$, $\Sigma_2(\mathsf{s}_1) = [a_1, a_2, a_3]$, $\Sigma_1(\mathsf{s}_C) = [a_1, a_2, a_3, a_4]$.
- $\Sigma_3(\mathsf{s}_0) = []$, $\Sigma_3(\mathsf{s}_1) = [a_3, a_2]$, $\Sigma_3(\mathsf{s}_C) = [a_3, a_2, a_4]$.