Reasoning about agents intentions

Justin Blount and Michael Gelfond

October 17, 2011

What

We further develop the formal theory of intentions suggested in "Reasoning about intended actions" (2005) by C. Baral and M. Gelfond to include reasoning about the goals of actions sequences (plans) and a rational agents response to the failure of a plan to achieve its goal.

Why

Previous work defined intention as a relation between an agent and an action sequence and formalized the following properties of intentions,

1) persistence – I will not usually give on my intentions

2) non-procrastination – If I intend to execute an action and I can, I will

Viewing intention as a relation between an agent and an action sequence can be an oversimplification. The goal of an action sequence is needed if we want to reason about domains where:

1) success may be caused unexpectedly by an exogenous action

2) failure may occur

Outline of how

• We extend the notion of what is intended from action sequence to *activities*, defined as a plan aimed at achieving a goal.

• We extend the notion of state of the domain by combining the physical state of the environment with the *mental state* of the agent.

Before we formally describe *activities* and the axioms of the *Theory of intentions* that describe the *mental state* of an agent, we will first give example scenarios of the types of reasoning we wish to formalize.

Scenario 1 - Expected success

Consider a row of four rooms, r_1,r_2,r_3,r_4 connected by doorways, such that an agent may move along the row from one room to the next. We say that two people *meet* if they are located in the same room. Assume that initially our agent Bob is in r_1 and his goal is to meet with John who, as Bob knows, is in room r_3. Bob's plan is to move from r_1 to r_2 and then to r_3. If John remains in r_3 then, after the plan intended by Bob is executed, he and John meet.

Scenario 2 - Unexpected success

Now suppose that, as Bob is moving from r_1 to r_2, John moves from r_3 to r_2. In this case it will be rational for Bob to recognize the unexpected achievement of his goal and not continue moving to r_3.

Scenario 3 - Failure

Suppose now that Bob moved from the $r1$ to $r2$ and then to $r3$, but John was not there. Bob must recognize that his plan failed. The further analysis however can allow Bob to conclude that, while he was executing his plan, John must have moved to $r4$. After that Bob will come up with a new plan to meet John and form and execute new intended activity designed for this purpose (in our case simply move to room 4).

Activity

- **An *activity* is a triple $\langle m, \alpha, g \rangle$, where m is the activity's name, α is a sequence $[a_1, \ldots, a_n]$ where a_i's are either names of agent actions or activities, and g is a goal expected to be achieved as a result of executing α.**

- **To execute m the agent will attempt to execute sequence $[a_1, \ldots, a_n]$. The execution can be interrupted or accompanied by exogenous actions.**

- **The execution of m ends in a *finishing state* where if the goal is achieved it is a *success* and a *failure* if its goal is not achieved.**

Activity - example

Bob(b) is going to meet with John(j) by moving from r1 to r2, and then to r3.

$activity(m).$

$goal(m, meet(b, j)).$

$component(move(b, r1, r2), 1, m).$

$component(move(b, r2, r3), 2, m).$

Length of an activity

$$\begin{aligned} has\_comp(M, K) &:- component(PA, K, M). \\ has\_comp(M, K) &:- component(M1, K, M). \end{aligned} \tag{1}$$

$$\begin{aligned} length(M, K) &:- has\_comp(M, K), \\ &\quad not\ has\_comp(M, K + 1). \end{aligned} \tag{2}$$

Mental state - the idea

Primarily described by inertial fluent $status(activities, index)$**.**

The *status* **of the execution of an activity** m **when** m **is being executed is the component of** m **that was most recently executed, and is** -1 **when** m **is not being executed.**

The execution of an activity m **is initiated and terminated by mental actions** *start(m)* **and** *stop(m)*, **respectively.**

Now we will show several example trajectories of the execution of activities. To simplify the examples, only positive fluent literals are shown.

Trajectories

The trajectories we will show correspond to answer sets of a ASP program that includes a history, definitions of activities, a description of the agent's environment, and the Theory of intentions.

Uninterrupted execution of activity m

State	*Action*
$\{status(m,-1), in(b,r1), in(j,r3)\}$	$(start(m))$
$\{status(m,0), in(b,r1), in(j,r3)\}$	$(move(b,r1,r2))$
$\{status(m,1), in(b,r2), in(j,r3)\}$	$(move(b,r2,r3)$
$\{status(m,2), in(b,r3), in(j,r3),$ $meet(b,j), success(m)\}$	$(stop(m))$
$\{status(m,-1), in(b,r3), in(j,r3), meet(b,j)\}$	

Unexpected success of activity m

State	*Action*
$\{status(m,-1), in(b,r1), in(j,r3)\}$	$(start(m))$
$\{status(m,0), in(b,r1), in(j,r3)\}$	$(move(b,r1,r2), move(j,r3,r2))$
$\{status(m,1), in(b,r2), in(j,r2),$	
$\quad meet(b,j), success(m)\}$	$(stop(m))$
$\{status(m,-1), in(b,r3), in(j,r3), meet(b,j)\}$	

Failure of activity m

State	*Action*
$\{status(m,-1), in(b,r1), in(j,r3)\}$	$(start(m))$
$\{status(m,0), in(b,r1), in(j,r3)\}$	$(move(b,r1,r2), move(j,r3,r4))$
$\{status(m,1), in(b,r2), in(j,r4)\}$	$(move(b,r2,r3)$
$\{status(m,2), in(b,r3), in(j,r4), failure(m)\}$	$(stop(m))$
$\{status(m,-1), in(b,r3), in(j,r3), meet(b,j)\}$	

Response to failure – The idea

If an activity fails and we know why, we should find a new plan to achieve the goal.

We create and execute a new activity with the same goal as the failed activity.

We call this new activity created because the failure of m the replan of m.

Failure of activity m with replanning

State	*Action*
$\{status(m,-1), in(b,r1), in(j,r3)\}$	$(start(m))$
$\{status(m,0), in(b,r1), in(j,r3)\}$	$(move(b,r1,r2)$
$\{status(m,1), in(b,r2), in(j,r4)\}$	$(move(b,r2,r3), move(j,r3,r4))$
$\{status(m,2), in(b,r3), in(j,r4),$ $failure(m), replanning(m)\}$	$(stop(m), start(replan(m)))$
$\{status(m,-1), status(replan(m),0),$ $in(b,r3), in(j,r4)\}$	$(move(b,r3,r4))$
$\{status(m,-1), status(replan(m),1),$ $in(b,r4), in(j,r4), success(replan(m))\}$	$(stop(replan(m))$

Response to inconsistent observations – The idea

What if we have an observation that is inconsistent with what we expect?

Find an explanation – diagnosis

Intuitively an explanation is an occurrence of an exogenous action in the past that restores consistency to what we expect and what we observe.

Failure of activity m with diagnosis

State	*Action*
$\{status(m,-1), in(b,r1), in(j,r3)\}$	$(start(m))$
$\{status(m,0), in(b,r1), in(j,r3)\}$	$(move(b,r1,r2), \underline{move(j,r3,r4)})$
$\{status(m,1), in(b,r2), \underline{in(j,r4)}\}$	$(move(b,r2,r3)$
$\{status(m,2), in(b,r3), in(j,r4),$	
$\underline{obs(in(j,r3),false)}, failure(m)\}$	$(stop(m))$
$\{status(m,-1), in(b,r3), in(j,r4)\}$	

In the third state we observe that John is not in $r3$. The underlined action is one of three explanations. John must have moved to $r4$ at step 0,1,or 2. The underlined fluent is in the state as a result of the explanantion.

Failure of activity m with diagnosis and planning

State	*Action*
$\{status(m,-1), in(b,r1), in(j,r3)\}$	$(start(m))$
$\{status(m,0), in(b,r1), in(j,r3)\}$	$(move(b,r1,r2), move(j,r3,r4))$
$\{status(m,1), in(b,r2), in(j,r4)\}$	$(move(b,r2,r3)$
$\{status(m,2), in(b,r3), in(j,r4), failure(m),$ $obs(in(j,r3), false), replan(m)\}$	$(stop(m), start(r(m))$
$\{status(m,-1), status(r(m),0),$ $in(b,r3), in(j,r4)\}$	$(move(b,r3,r4))$
$\{status(m,-1), status(r(m),1),$ $in(b,r4), in(j,r4), success(r(m)\}$	$(stop(r(m))$

Theory of Intentions – Describing the mental state

Now we will describe the fluents and actions of the Theory of Intentions.

Mental fluents

- $status(M, K)$ **Inertial fluent**

which is true if

1. K **is the position of the last component of** M **executed by the agent when** $K \in [0..l]$ **or**

2. M **is dormant when** $K = -1$**.**

Fluent $status(M, K)$ **is a function of** M**.**

$$\begin{aligned} \textbf{-}h(status(M, K1), I) \ :- \ & h(status(M, K2), I) \\ & K1 != K2. \end{aligned} \tag{3}$$

Mental fluents

An activity is $awake$ **when its status is not** -1**.**

- $awake(M)$ **Defined fluent**

$$h(awake(M), I) \ :- \ \textbf{-}h(status(M, -1), I). \tag{4}$$

Mental Actions - Start

To awaken an activity, action $start$ **sets the status to** 0**. An agent cannot start an activity that is already awake.**

$$h(status(M, 0), I + 1) \ :- \ o(start(M), I). \tag{5}$$

$$\text{-}o(start(M), I) \ :- \ h(awake(M), I). \tag{6}$$

Mental Actions - Stop

To return an activity to a dormant state, action *stop* sets the status to -1. An agent cannot stop and activity that is already dormant.

$$h(status(M, -1), I + 1) \ :- \ o(stop(M), I). \tag{7}$$

$$\text{-}o(stop(M), I) \ :- \ \text{-}h(awake(M), I). \tag{8}$$

Stopping subactitivies

- $subactivity(M1, M)$ **defined fluent**

Is true when $M1$ is the next component of an activity M.

$$\begin{aligned} h(subactivity(M1, M), I) &:- component(M1, K+1, M), \\ &\quad h(status(M, K), I). \\ h(subactivity(M1, M), I) &:- h(subactivity(M2, M), I), \\ &\quad h(subactivity(M1, M2), I). \end{aligned} \tag{9}$$

Stopping an activity stops all of its subactivities.

$$\begin{aligned} h(status(M1, -1), I+1) &:- o(stop(M), I), \\ &\quad h(subactivity(M1, M), \\ &\quad h(awake(M1), I). \end{aligned} \tag{10}$$

Success

- $success(M)$ **Defined fluent**

which is true when the goal of an activity M is achieved.

$$\begin{aligned} h(success(M), I) \ :- \ & h(awake(M), I), \\ & goal(M, F), \\ & h(F, I). \end{aligned} \tag{11}$$

Failure

- $failure(M)$ **Defined fluent**

which is true when

the execution of an activity M does not achieve its goal.

$$\begin{aligned} h(failure(M), I) \ :- \ & length(M, K), \\ & h(status(M, K), I), \\ & \text{-}h(success(M), I). \end{aligned} \tag{12}$$

Propagation of failure

An activity M is a failure if it contains an activity $M1$ that is a failure.

$$\begin{aligned} h(failure(M), I) \quad :- \quad & h(state(M, K), I), \\ & component(M1, K+1, M), \\ & h(failure(M1), I), \\ & \text{-}h(success(M), I). \end{aligned} \tag{13}$$

In progress

- $in\_progress(M)$ **Defined fluent**

which is true when an activity M is awake but is neither a success or failure.

$$\begin{aligned} h(in\_progress(M,I)) \quad :- \quad & h(awake(M),I), \\ & \text{-}h(success(M),I), \\ & \text{-}h(failure(M),I). \end{aligned} \tag{14}$$

Intended action

- $intended\_action(M, E)$ **Defined fluent**

defines the action E of activity M that is intended for execution.

Action E may be a physical or mental action.

Intended Actions - Physical

Action PA is a physical action in the next position of an intended activity M.

$$\begin{aligned} h(intended\_action(M, PA), I) \ :- \ & h(state(M, K), I), \\ & component(PA, K+1, M), \\ & h(in\_progress(M), I). \end{aligned} \tag{15}$$

Intended Action - Physical Action

The execution of such an intended action increments the status of the activity of which it is a part.

$$\begin{aligned} h(state(M, K+1), I+1) \ :- \ & o(PA, I), \\ & h(intended\_action(M, PA), I), \\ & h(state(M, K), I). \end{aligned} \tag{16}$$

Intended Action - Start

An activity whose execution is intended and is dormant should be started.

$$\begin{aligned} h(intended\_action(M, start(M1)), I) \ :- \ & h(state(M, K), I), \\ & component(M1, K+1, M), \\ & h(in\_progress(M), I), \\ & \text{-}h(awake(M1), I). \end{aligned} \tag{17}$$

Intended actions in a finishing state

Activities that are a success or failure should be stopped.

$$h(intended\_action(M, stop(M)), I) \ :- \ h(failure(M), I). \tag{18}$$

$$h(intended\_action(M, stop(M)), I) \ :- \ h(success(M), I). \tag{19}$$

The ending of an nested activity $M1$ increments the status of the intended activity M of which it is an part.

$$\begin{aligned} h(status(M, K+1), I+1) \ :- \ & o(stop(M1), I), \\ & h(status(M, K), I), \\ & component(M1, K+1, M), \\ & \text{-}h(failure(M1), I), \\ & h(in\_progress(M), I). \end{aligned} \tag{20}$$

Axioms of non procrastination

Normally an agent executes his intended actions.

$$\begin{aligned} o(E, I) \ :- \ & h(intended\_action(M, E), I), \\ & not \ \text{-}o(E, I), \\ & I < n. \end{aligned} \tag{21}$$

Normally if an agent is executing an activity, he does not execute actions that are not intended actions.

$$intended(PA, I) \ :- \ h(intended\_action(M, PA), I). \tag{22}$$

$$\begin{aligned} \text{-}o(PA, I) \ :- \ & h(awake(M), I), \\ & not \ intended(PA, I). \end{aligned} \tag{23}$$

Modeling behavior in response to failure – planning

When an agent's activity fails we expect him to find another plan to achieve the goal.

The agent will create and execute a new activity, $replan(M)$**.**

What is the goal of $replan(M)$**?**

What are the components of $replan(M)$**?**

What property must $replan(M)$ **satisfy?**

Main activities

• $minor(M)$ **Defined fluent**

An activity $m1$ **that is a subactivity of an activity** m **is a** *minor* **activity.**

$$h(minor(M1), I) \ :- \ h(subactivity(M1, M), I). \tag{24}$$

An activity m **that is being executed and is not minor activity is a** *main* **activity.**

• $main(M)$ **Defined fluent**

$$\begin{aligned} h(main(M), I) \ :- \ & h(awake(M), I), \\ & \text{-}h(minor(M), I). \end{aligned} \tag{25}$$

Planning

Replanning is done when the main activity is a failure.

$$h(replanning(M), I) \ :- \ h(main(M), I), \\ h(failure(M), I). \tag{26}$$

Activity $replan(M)$ **should be initiated.**

$$h(intended\_action(M, start(replan(M))), I) \ :- \ h(replanning(M), I). \tag{27}$$

Determining the goal and components of $replan(M)$

The goal of $replan(M)$ is the same as M.

$$goal(replan(M), F) \ :- \ h(replanning(M), I), \\ goal(M, F). \tag{28}$$

We determine the components of $replan(M)$ by finding a sequence of actions that achieve the goal of the main activity.

$$replan(PA, K, I) : component(PA, K, replan(M)) \ +- \ h(replanning(M), I), \\ 0 < K. \tag{29}$$

Property that $replan(M)$ must satisfy

These axioms ensure that $replan(M)$ will achieve the goal of the main activity.

$$\begin{aligned} g(I) \ :- \ & h(main(M), I), \\ & goal(M, F), \\ & I < I1, \\ & h(F, I1). \end{aligned} \tag{30}$$

$$\begin{aligned} g \ :- \ & g(I). \\ :- \ & not \ g. \end{aligned} \tag{31}$$

Finding Explanations

The reality check axioms ensure that observations are consistent.

$$\begin{aligned} &:- \ obs(F, false, I), \\ &\quad\ h(F, I). \\ &:- \ obs(F, true, I), \\ &\quad\ \text{-}h(F, I). \end{aligned} \tag{32}$$

$$\begin{aligned} obs(I) \ &:- \ obs(F, true, I). \\ obs(I) \ &:- \ obs(F, false, I). \end{aligned} \tag{33}$$

Finding an explanation is done by considering occurrence of exogenous actions that restore consistency.

$$diag(EA, I1, I2) : o(EA, I1) \ +- \ obs(I2), \quad I1 < I2. \tag{34}$$

conclusions

In this work we present a *Theory of Intentions* that allows for the

• reasoning about plans and goals

• reasoning about an agent's response to failure and inconsistent observations.

In the later we use cr-prolog to form new activities in response to failure and to find explanations for inconsistent observations.

The END