Reasoning about Intended Actions

Chitta Baral and Michael Gelfond

To appear in AAAI05

Intentions in Natural Language

A mathematical analysis of a notion of "intention" may help to better understand a large number of natural language constructs.

Sentences

- **Mike went to school where he was planning to meet John.**
- **The train is scheduled to leave at 10:00.**
- **Mary decided to go to Paris as soon as possible.**

can all be analyzed in terms of actors (Mike, train, and Mary) having an intention to perform an action (meet, leave, go).

The Starting Point

- **Psychological and philosophical studies of the role of intentions in rational human behavior cataloged different functions of intentions.**
- **The BDI architecture for rational agents concentrated on the intentions ability to trigger or initiate the intended action and sustain it until completion.**
- **Logical analysis formalized some of the properties of intentions in modal logics.**

The Problem

- **Axiomatic systems which include formalization of intentions are rather complex and not very elaboration tolerant.**
- **Mathematical properties of these systems are not well understood.**
- **There is a substantial distance between the logical theory and implementations of the corresponding architecture.**

Intended Actions in LP: Basic Axioms

1. Normally intended actions are executed the moment such execution becomes possible.

```
occurs(A,I) :-
       intend(A,I),
       not -occurs(A,I).
```

2. Normally failure does not cause the abandonment of intended actions:

```
intend(A,I+1) :-
            intend(A,I),
            -occurs(A,I),
            not -intend(A,I+1).
```

Intended Actions in LP: Example

(i). Consider theory T_1 consisting of axioms (1) and (2) and a statement $intend(a, 1)$ - "the agent intends to execute action a at time-step 1**".**

$$T_1 \models occurs(a, 1).$$

(ii). Now let $T_2 = T_1 \cup \{\neg occurs(a, 1)\}$. This time

$$T_2 \models occurs(a, 2).$$

This is possible only because LP theories are non-monotonic and hence elaboration tolerant.

Intended Sequences of Actions

Intention to execute a sequence $s = \langle a_1, \ldots, a_n \rangle$ of actions at time-step i consists of the intention to execute a_1 at i followed by the intention to execute $a_2, \ldots, a_n$ at time-step at which the execution of a_1 is completed.

Execution of the intended sequence can be interrupted by exogenous events at any time but the axioms for intentions should allow the agent to complete the execution of s if at all possible.

More Axioms

Axioms (1) and (2) together with axioms

```
3.  intend(A,I) :-
            intend([A | S],I).
4.  intend(S,I1) :-
            intend([A | S],I),
            ends(A,I1).
5.  ends(A,I+1) :-
            occurs(A,I).
```

initiate a sequence $[A|S]$ of actions and sustain it until completion.

Example

Theory T_1 **consisting of axioms (1)–(5) and a statement** $intend([a_1, a_2, a_3], 1)$ **entails**

$occurs(a_1, 1)$

$occurs(a_2, 2)$

$occurs(a_3, 3)$

while $T_2 = T_1 \cup \{\neg occurs(a_1, 1), \neg occurs(a_2, 3)\}$ **entails**

$occurs(a_1, 2)$

$occurs(a_2, 4)$

$occurs(a_3, 5)$

Limitations

The above axiomatization has a number of limitations:

- **Axioms (3)–(5) are limited to sequences of distinct actions. We may need more complexity, e.g. sequences of sequences.**
- **Use of lists make the universe of our theory infinite which prevents us from combining reasoning about intentions with answer set programming solutions of planning and diagnostic problems.**

Making the Universe Finite

A sequence $s = \langle a_0, \ldots, a_n \rangle$ will be recorded as:

$component(a_1, 1, s)$

$\ldots$

$component(a_n, n, s)$**.**

$length(s, n)$**.**

New Axioms

To accommodate new representation we replace axioms (3) and (4) by

```
3a.  intend(A,I) :-
            intend(S,I),
            component(A,1,S).
4a.  intend(A2,I) :-
            intend(S,I),
            component(A1,K,S),
            component(A2,K+1,S),
            ends(A1,I).
```

Models of a new theory are always finite.

Correctness and Completeness

• **Given:**

(i) A transition diagram T describing all possible trajectories of a domain given by action theory T_a.

(ii) A history H of the domain consisting of statements

$happend(s, i)$

$observe(l, i)$

$intend(s, i)$

• **Define trajectories of T satisfying H. Show that there is one-to-one correspondence between these trajectories and answer sets of a program $T_a \cup H \cup T_i$ where T_i consists of the above axioms.**

Conclusions

We developed and investigated a theory of intentions which:

- **Declarative;**
- **Is well integrated with general theories of dynamic domains;**
- **Allows query answering by general Answer Set Programming algorithms**
- **The preliminary experience shows that use of theory of intentions simplifies formalization of many domains and will allow development of more elegant and efficient query-answering systems.**

Future Work

- **Axiomatize delayed and conditional intentions (”Michael plans to go to Arlington on June 6th, if he’ll be able to find the money for the trip”)**
- **Better understand the methodology of using the theories of intentions. (Build and compare micro-theories with and without intentions).**
- **Integrate intentions to execute action a with deliberate planning to make a executable.**