Electronic News Journal on Reasoning about Actions and Change

Editor: Erik Sandewall

Volume 2, Number 11-12

Contributions during November-December, 1998

ISSN, print: 1403 - 2031
on-line: 1403 - 204X

Linköping University Electronic Press
http://www.ep.liu.se/ej/enrac/1998/11/

Publication Trail and Update Rules

The Electronic News Journals (ENJ) are a medium for exchange of scientific information and debate. In particular, they serve as the forum where articles received by the ETAI (Electronic Transactions on Artificial Intelligence) are discussed publicly for review, in separate News Journals for each ETAI area.

The Electronic News Journal on Reasoning about Actions and Change (ENRAC), in particular, uses a publication trail where contributed information is first distributed on a daily basis as *Newsletters* using electronic mail and web pages in HTML. Then, contributions during the same month are compiled into an HTML version of the monthly *News Journal*. Finally, the same information is converted via Latex to a postscript version that is suitable for printing on paper and reading off-line. It is formally published and archived by the Linköping University Electronic Press in both its electronic and its paper form, as described on page (iv).

Offprints of individual segments. Each News Journal issue consists of a sequence of *segments* relating to specific topics. "Offprints" of these can be obtained from the URL:s indicated at the head of the first page of each segment, or from the ENRAC service page specified below. It is intended to maintain these URL:s and their contents for the foreseeable future. Misprints of minor significance are corrected in these offprints (but not in the E-Press version since it is considered archival).

Corresponding HTML edition with dynamic links. The HTML edition contains considerable numbers of links to other pages and structures on the net: articles that are available on-line, home pages of conferences and of individual researchers, links to other part of the ETAI structure, and so on. Due to the natural limitations of the paper medium, only some of the WWW links have been retained here as footnotes. In order to retrieve articles and other information that are referenced in the present issue, it is recommended to look up the corresponding issue in HTML and to use its link. The HTML issues of the News Journals on Reasoning about Actions and Change can be found at the ENRAC *service page,*

`http://www.etaij.org/rac/nj/`

It is intended that the HTML issues will by updated continously to the largest extent possible, for example by replacing URL links to the home pages of authors that have changed to another site.

Table of Contents

Dates of Publication

Since the date of publication may be understood either as the date of first public appearance, or as the day of reproduction on paper in many copies, and since both of these definitions may be difficult to apply in the case of electronic publication, we make the following clarifying statement.

The contents of the present issue were put on-line in their original, versions as Newsletters in plaintext, HTML, and latex/postscript during the month of November-December, 1998. Then the contents were edited and formatted, resulting in the present, formatted version which was published on 16.7.1999, in two concurrent editions: an on-line edition and a paper edition. The on-line edition was timestamped electronically and put on-line by Linköping University Electronic Press at the URL specified on page (i). The paper edition was obtained by printing the on-line edition on a standard computer printer. It was reproduced in 25 copies, legally archived, and made available for distribution.

Electronic News Journal on Reasoning about Actions and Change 357
Vol. 2: *357-357*, 31.12.1998
http://www.etaij.org/rac/nj/9811/

The Months of November-December

Summary of News Journals Contents

During December we had the pleasure to announce the acceptance to the ETAI of the article "The Complexity of Model Checking in Modal Event Calculi with Quantifiers", by Iliano Cervesato, Massimo Franceschet, and Angelo Montanari.

One of the discussion sessions at the AAAI Fall symposium on Cognitive robotics concerned the design of robot competitions. The starting point for the discussion was that success in these competitions does not seem to be obtained by making the robot "intelligent", but rather by engineering its more mundane properties. While respecting the importance of low-level robot control, the discussion addressed the question how the competitions ought to be designed in order to emphasize (or at least not de-emphasize) those aspects of the overall design that are relevant for A.I. - Yves Lespérance has contributed a summary of the discussion at the symposium, and the ENRAC has now opened a Newsletter discussion on the same topic, with Yves' report as the initiating statement.

We received one more reference article, this one about the Toronto variant of the situation calculus, and we feature continued discussions on the submitted articles by Poole and by Baral et al.

Electronic News Journal on Reasoning about Actions and Change 358
Vol. 2: *358-365*, 31.12.1998
http://www.etaij.org/ra/rac/008/

Protocol of on-line discussion during November-December, 1998 about the following research article:

David Poole

Decision Theory, the Situation Calculus and Conditional Plans

Q3. **Hector Geffner** (19.11)

David,

I've enjoyed reading your paper "Decision Theory, the Situation Calculus and Conditional Plans". I find it hard to disagree on your general point that a general model of action must combine elements from logic and probability/utility theory, so my questions and comments are addressed more to the question of how this combination is achieved in your proposal. I'll try to comment first on the more general issues.

1. In page 24, you say "The representation in this paper can be seen as a representation for POMDPs".

If so, I wonder, why not present the representation in that way from the very beginning? Wouldn't that make things much clearer?

Namely, a POMDP is described by (see [1]):

A - state space
B - actions
C - transition probabilities
D - cost function
E - observations
F - observation probabilities

Then the constructions in your language could be seen as a convenient/concise/modular way for specifying these components.

Actually, once the mapping from the language to these components is determined, no further semantics would be needed (i.e., the probabilitities of trajectories, the expected utility of policies, etc., are all defined as usual in POMDPs). This is actually the way we do things in [2].

2. I can see some reasons why not to present your representation as a "front end" to POMDPs.

- you want to be more general; e.g., be able to deal with cost functions that depend on the system history. If this is the case, I'd suggest to introduce "generalized" POMDPS where (cumulative) cost functions do not have to be additive (btw, in such POMDPS belief states are not necessarily sufficient statistics ..)

- you want to accommodate multiple agents. Yet this is not done in this paper, but even in that case, multi-agent POMDPs could be defined as well, and have been defined in the completely observable setting (e.g., see Littman).
- you are not interested in determining an optimal or near-optimal solution of the POMDP but are interested in analyzing the expected cost of a policy supplied by a user in the form of a contingent plan. Again, this is no problem from the point of view of POMDPs, as long as the contingent plan determines a (stationary or non-stationary) *function*
- *of belief states into actions (see [4]). Indeed, you require* more
- than this when you demand (page 18), that the tests in conditions of the plan, be *directly observable* at the time when the conditions need to be evaluted. This is not required in POMDPs [4] or in possible world accounts [3], where the test may be known indirectly through other observations.

  3. A final comment about a non-trivial problem that results from the decision of removing all probabilities from state transitions transferring them into suitable priors in "choice space".

  Consider the effect of an action *turn_off* on a single boolean fluent "on" with transition probabilities:

$$P(on = false \mid on = true \; ; \; a = turn\_off) = .5$$
$$P(on = false \mid on = false \; ; \; a = turn\_off) = 1$$

Let's say that initially $on = true$ and then we perform a sequence of N consecutive *turn_off*s. Then the Probability of $on = false$ is 0.5^N, which clearly depends on N.

I don't see how this behavior could be captured by priors over choice space. This seems to be a strong limitation. It looks as if probabilistic actions cannot be handled after all. Is this right?

References

[1] L. Kaebling, M. Littman, and T. Cassandra. Planning and Acting in Partially Observable Stochastic Domains, AIJ 1998
[2] H. Geffner, B. Bonet. High level planning with POMDPs. Proc. 1998 AAAI Fall Symp on Cognitive Robotics (www.ldc.usb.ve/~hector)
[3] H. Levesque. What's planning in the presence of sensing? AAAI96
[4] H. Geffner and J. Wainer. A model of action, knowledge and control, Proc. ECAI 98 (www.ldc.usb.ve/~hector)

A3. **David Poole** (24.11)

I've enjoyed reading your paper "Decision Theory, the Situation Calculus and Conditional Plans". I find it hard to disagree on your general point that a general model of action must combine elements from logic and probability/utility theory, so my questions and comments are addressed more to the question of how this combination is achieved in your proposal. I'll try to comment first on the more general issues.

1. In page 24, you say "The representation in this paper can be seen as a representation for POMDPs".

If so, I wonder, why not present the representation in that way from the very beginning? Wouldn't that make things much clearer?

One of the problems with writing a paper that is trying to bridge different areas is to try to explain it for all readers (or to keep them all equally unhappy). I suppose I have succeeded when POMDP researchers say "this is just a representation for POMDPs" and when situation calculus researchers claim "this is just the situation calculus with some some probability added" or those studying Bayesian networks claim that "this is just a rule-based representation for dynamic Bayesian networks with actions and observables". Describing it explicitly in terms of POMDPs may have made it easier for POMDP researchers, but not necessarily for everyone else.

Namely, a POMDP is described by (see [1]):

A - state space
B - actions
C - transition probabilities
D - cost function
E - observations
F - observation probabilities

Then the constructions in your language could be seen as a convenient/concise/modular way for specifying these components.

I thought I did do this. (At least I was trying to do this, and I knew when I was finished when I had defined all of the components of a POMDP). The actions and observables of Definition 2.6 are the same as B and E. I specify C, D & F using rules. In particular I represent the cost function in terms of rules that imply the *utility* predicate (Section 2.6), and observation probabilities in terms of rules that imply the *sense* predicate (Section 2.7).

Actually, once the mapping from the language to these components is determined, no further semantics would be needed (i.e., the probabilities of trajectories, the expected utility of policies, etc., are all defined as usual in POMDPs). This is actually the way we do things in [2].

That is essentially what I did. There are some problems with just relying on the definition of a POMDP. This has to do with what is a state. In the formal work, the notion of state is taken as primitive. However, for a well-defined semantics we need to be clearer. A state could either mean:

(a) everything that is true at a stage. This would include the actual time (if it is relevant), the values of sensors, the accumulated reward, etc.

(b) a sufficient statistic about the history to render the future independent of the past given the state (which, by definition is Markovian). This typically doesn't include the time, the output of the sensors, or the accumulated reward.

I have taken the approach of (a). Most POMDP researchers take the approach of (b) because their algorithms reason in terms of the state space. They therefore want to keep the state space as small as possible. I don't want to reason at the level of the state space, but instead at the level of the propositions.

> *2. I can see some reasons why not to present your representation as a "front end" to POMDPs.*

In general I don't want to do this, because I don't think that reasoning in terms of the state space is the right thing to do. What we can learn from AI is that we want to reason in terms of propositions or random variables, not in terms of the state space. One of the themes of my work is to find compact representations and to exploit the compactness for computational gains. Thus, I don't want to present this as a "front end" to POMDPs. If you mean following the semantic intuitions of POMDPS, then I think I do. But POMDPs are just one of the formalisms I am building on. They have nothing to say about useful independence assumptions or about concise descriptions of actions (state transition functions), which is what this paper is mainly about.

> *you want to be more general; e.g., be able to deal with cost functions that depend on the system history. If this is the case, I'd suggest to introduce "generalized" POMDPS where (cumulative) cost functions do not have to be additive (btw, in such POMDPS belief states are not necessarily sufficient statistics ..)*

(unless you include the accumulated reward in the state.)

> *you want to accommodate multiple agents. Yet this is not done in this paper, but even in that case, multi-agent POMDPs could be defined as well, and have been defined in the completely observable setting (e.g., see Littman).*

This was done in my AIJ paper, "The Independent Choice Logic for Modelling Multiple Agents Under Uncertainty". I actually don't think that the ICL_SC formulation is as good for this as the formalism in that paper.

> *you are not interested in determining an optimal or near-optimal solution of the POMDP but are interested in analyzing the expected cost of a policy supplied by a user in the form of a contingent plan. Again, this is no problem from the point of view of POMDPs, as long as the contingent plan determines a (stationary or non-stationary)* function *of belief states into actions (see [4]).*

First let me point out that POMDPs are not defined in terms of belief states. It is a theorem about POMDPs (Astrom 1965) that says that an optimal policy can be represented in terms of a function from belief states (probability distributions over ordinary states) into actions. However, "optimal" here doesn't take into account computation time. It is pretty obvious that this isn't true for bounded optimal agents (where the computation time and the space are taken into account), which is where I would like to take this.

I *am* interested in determining an optimal or near-optimal solution (but one that takes computation time and space onto account). I want to be able to define the expected utility of any strategy the agent may be carrying out, whether it maintains a belief state or not. No matter what program the agent is following, it can be described in terms of a conditional plan that describes what the agent will do based on the alternate sequences of observations (i.e., its function from observation sequences into actions). Note that this doesn't refer to anything inside the agent. The agent could be reasoning with a probability distribution over states, remembering a few previous observations, just reacting to its current observations, or actually following a conditional plan. The important thing is what it does is based on what it observes.

What I am advocating is good hacks; the best agent won't do exact probabilistic reasoning. We need good strategies, perhaps like your RTDP-BEL. However, it is not obvious (to me anyway) that the bounded optimal solution will necessarily be an approximation to the (unbounded) optimal solution. The best real (bounded time and bounded space) agent may do something quite different than approximating belief states. Unless we have some representation that lets us model the expected utility of an agent following its function from observation and action history into actions (i.e., its transduction), we won't be able to specify when one agent is better than another. The conditional plans (policy trees) let us model this, without any commitment to the internal representation of the agent.

> *Indeed, you require* more *than this when you demand (page 18), that the tests in conditions of the plan, be* directly observable *at the time when the conditions need to be evaluated. This is not required in POMDPs [4] or in possible world accounts [3], where the test may be known indirectly through other observations.*

NO! I mean that the conditions are those things that are observable. I mean whatever evidence (the "other observations") that the POMDPs or possible worlds accounts take into account. This is a standard technique in Bayesian networks/ influence diagrams; we create a variable that represents the actual value sensed. I am quite happy to have this as a (state) variable. You don't want it as a part of your state because it isn't necessary to make the system Markovian. This is fine, but I can model these "other observations" using the "sense" predicate. Is the sense predicate part of the state? I don't know; that is one reason why I didn't blindly accept the notion of state.

> *3. A final comment about a non-trivial problem that results from the decision of removing all probabilities from state transitions transferring them into suitable priors in "choice space".*
>
> *Consider the effect of an action "turn_off" on a single boolean fluent "on" with transition probabilities:*
>
> $$P(on = false \mid on = true \; ; \; a = turn\_off) = .5$$
> $$P(on = false \mid on = false \; ; \; a = turn\_off) = 1$$
>
> *Let's say that initially $on = true$ and then we perform a sequence of N consecutive turn_off's. Then the Probability of $on = false$ is 0.5^N, which clearly depends on N.*
>
> *I don't see how this behavior could be captured by priors over choice space. This seems to be a strong limitation. It looks as if probabilistic actions cannot be handled after all. Is this right?*

Here is how I would represent this: Facts:

$on(do(turn\_off, S)) \rightarrow on(S) \wedge off\_failed(S).$

$on(init).$

C_0 *contains* $\{off\_failed(S), off\_succeeded(S)\}$ *for all situations* S.

$P_0(off\_failed(S)) = 0.5$ *for all situations* S.

[Note that the completion of the facts for *on* gives your two conditional probabilities.]

We can then derive

$P(on(do(turn\_off, do(turn\_off, do(turn\_off, init))))) = 0.5^3$

just as you want (similarly for any N). We need an assumption for each time step. The minimal explanation for

$on(do(turn\_off, do(turn\_off, do(turn\_off, init))))$

is the set of the following:

$off\_failed(init)$

$off\_failed(do(turn\_off, init))$

$off\_failed(do(turn\_off, do(turn\_off, init)))$

In all worlds with this explanation true, the light is on in the state after three *turn_offs*. The probability of these worlds sum to 0.5^3.

Any probabilistic actions that you can represent in a dynamic Bayesian network, I can represent in the ICL_SC. Moreover I can represent it at least as succinctly as the dynamic Bayesian network (up to a constant factor) and sometimes exponentially (in the number of state variables) more succinctly (when there is context-specific independence). The mapping is local and modular.

Thanks for taking the time to read and comment on the paper. I hope my comments help makes the paper and what I am trying to do clearer. Let me stress that this paper has nothing to say about computation. Whereas your work was motivated by being a front end to an actual POMDP algorithm; the motivation here relies on being able to deliver on POMDP algorithms that can exploit the conditional and contextual independencies of the representation. Unfortunately, I can't show you such algorithms now. But we are working on it.

David

C3-1. **Hector Geffner** (26.11)

David,

thanks for your answers; they helped me a lot. Just a brief follow up on the status of your proposed framework: a new model? a new language? a new algorithm? all of them? ...

Best regards. -hector

> 1. In page 24, you say "The representation in this paper can be seen as a representation for POMDPs".
>
> If so, I wonder, why not present the representation in that way from the very beginning? Wouldn't that make things much clearer?

One of the problems with writing a paper that is trying to bridge different areas is to try to explain it for all readers (or to keep them all equally unhappy). I suppose I have succeeded when POMDP researchers say "this is just a representation for POMDPs" and when *situation calculus researchers claim "this is just the situation calculus with some some probability added" or those studying Bayesian networks claim that "this is just a rule-based representation for dynamic Bayesian networks with actions and observables". Describing it explicitly in terms of POMDPs may have made it easier for POMDP researchers, but not necessarily for everyone else.*

I think I don't agree with this. I believe that POMDPs are a general and natural *model* for sequential decision problems that involve sensing. They are not for POMDPs researchers only; in the same sense, that logic is not only for logicians. And they are simple too (unlike some of the POMDP *algorithms* that are indeed complex).

I think that anyone dealing with *sequential decision problems that involve sensing* should know about POMDPs, whether or not they appeal to POMDP algorithms for solving them, and whether or not they deal with probabilities. I know this sounds dogmatic ... but it is the truth!!! :-)

No, really, POMDPs are very useful for understanding and identifying the different dimensions of a decision problem: transitions, costs, information; and at the same time they have little to do with probabilities (namely, even if all probabilities become zero or one, POMDPs are still very meaningful and don't reduce to anything else as 0-1 probabilities that would reduce for instance to logic).

On a more general note, I think there are three essential aspects to the work in planning and control in AI:

- models
- languages
- algorithms

Models are about the mathematics of the task: what is a problem? what is a solution? what is an optimal solution? Languages are for describing these models in a convenient way. Algorithms are for computing the solutions.

I think these three ingredients are always present in approaches to planning and control in AI, and I believe it is useful to make to them explicit; even when the algorithms may take advantage of the particular language in which the model is represented (as in Strips planning).

I understand that you expect your language to be useful not only for specifying POMDPs in a convenient way, but also for solving them conveniently. That would be great. Yet even in that case, people using different POMDP algorithms could in principle benefit from your language for setting up their POMDP models. This could also be important, and indeed, in the recent AAAI Symposium on POMDPs, the need for good languages for building POMDPs for particular applications and for exchanging benchmarks was emphasized. May be your language, as well as other action languages suitably extended, could fill up that need. As you know, we have also been doing work in that direction.

Best regards.

- hector

C3-2. **Chitta Baral** (22.12)

Hi Hector and David:

I have read your exchanges in the ENRAC about the importance of POMDPs in *decision problems involving sensing.* I have the following two questions to Hector.

(i) I think POMDPs are very good for this particular type of problems. However, I am not sure they scale up in the following sense: If we elaborate *sensing* as obtaining knowledge then when we consider obtaining knowledge of the form $(KKKf)$ where K is the modal opeartor 'knows'. I am not sure if a POMDP based approach can scale upto this.

On the other hand it seems that the approach in Bachus, Levesque and Halpern (IJCAI 95) can scale up to this and they can be shown equivalent to a POMDP based approach in the special case where we only deal with one level of knowledge. (eg: S5 guarantees that.)

(ii) I don't understand why (perhaps you can help me here), in a POMDP formulation (for example, in Hector's AAAI fall symposium paper this year) we have the action a mentioned in $P_a(o \mid s)$. I feel it should not matter how s was reached, and we should just have $P(o \mid s)$. Otherwise observations are in some sense non-Markovian, as they not only depend on the state of the world, but also how that state was reached. (This is not that big a deal though.)

Best regards

Chitta

Electronic News Journal on Reasoning about Actions and Change 366
Vol. 2: *366-367*, 31.12.1998
http://www.etaij.org/ra/rac/011/

Protocol of on-line discussion during November-December, 1998 about the following research article:

Chitta Baral and Son Cao Tran

Relating Theories of Actions and Reactive Control

A1. **The Authors** (23.11)

Dear Erik:

We really appreciate your comments. Following is a point by point response to your comments. In addition, we plan to revise our paper to take into account some of your suggestions and our responses to it.

Best regards
Chitta and Son

1. We agree that our current formulation in terms of maintaining goals is weaker than the formulation where the control never allows the agent to get into an unacceptable state, regardless of exogenous actions. The later approach is used in (among other approaches) policy determination algorithms for Markov Decision Process based approaches [DKKN95,KBSD97]. (As we point out in the paper, the algorithms in [DKKN95,KBSD97] consider the control module as a whole and do not have the sufficiency condition for correctness of individual goals.)

We feel that the later approach is too strong and often there may not be any control module satisfying the stronger criteria.

Nevertheless, studying a middle ground between the two is important.

Following your suggestions, one middle ground can be based on using a maximum number (say m) of exogenous actions that are considered (or allowed) between two successive agent actions.

Let G be the set of goal states. We can define $core(G, m)$ as a subset of G, from where application of any m exogenous actions will result in a state in G, and there exists a sequence of exogenous actions of length $m + 1$ that will take us to a state outside G.

We can then use this parameter m in defining satisfaction of maintenance goals, by modifying Defn 3.6 in the paper, to have $Closure(S, M, A) \subseteq G$ and for all states s in $Closure(S, M, A)$, each step of executing M takes us to a state in $Core(G, m)$.

We do not forsee any problem with concurrency with this approach. But the definition of achievement and maintainance will no longer be uniform, and as a result the definition for mixed control modules will get a little complicated.

We will try to pursue this in greater detail (or at least mention this) in the revised version of our paper.

2. As you mention, our paper can be formulated in terms of states, state spaces and transition functions without talking about action theory.

We felt that by doing that, the connection between an action specification and the states and transition function it defines in a succinct and elaboration tolerant (often) way is pushed into the background, and we wanted to stress that states and transition functions are not often directly given but are rather specified by an action description. For that reason, we bring in action theories.

We will point this out in our revised version.

3.

(a) Our results rely on earlier theorems about existence of least fixpoints obtained through iterating a monotonic function.

(b) We have two kind of theorems in our paper: correctness based on sufficiency condition (Theorem 4.1 and 7.1) and correctness of automatic construction algorithms (Proposition 4.1 and 7.3). The later propositions are of course important, as any algorithms should have a result about its correctness. The former theorems are important because they are used in developing the algorithm, and are crucial in proving the correctness of the algorithms.

4. We appreciate your observation. We had not considered this connection.

One difference with respect to your KR 96 approach is that in your approach the first step of your "closure" is by applying an action and the subsequent steps are due to a *successor* relation, while in our approach the subsequent steps are due to exogenous actions.

Regardless of this difference, we will point out the connection in our revised version.

5. Calling natural events as actions may sound odd some time. The reason we chose to call everything actions are several.

(i) The exogenous actions are not just natural actions, but are actions that are beyond the control of the agents. They could be the action of another agent in a multi-agent world. Thus they are a super set of what I would normally think of as "natural actions".

(ii) By just saying actions, we avoid talking separately about a theory of events. Otherwise every time we talk about effect of actions we would have to talk about effect of events, and their will be a lot of duplication.

Electronic News Journal on Reasoning about Actions and Change
Vol. 2: *368-368*, 31.12.1998
http://www.etaij.org/ra/rac/017/

Hector Levesque, Fiora Pirri, and Ray Reiter:

Foundations for the Situation Calculus

Abstract of the reference article

This article gives the logical foundations for the situations-as-histories variant of the situation calculus, focusing on the following items:

- The language of the situation calculus.
- Foundational axioms for the domain of situations.
- Axioms for an underlying domain theory.
- The syntax and semantics of the logic programming language GOLOG.
- Axioms for knowledge and sensing actions.
- Essential metatheoretic results about the situation calculus.