In Lester, J. (Ed.) (2001). *IUI2001: International Conference on Intelligent User Interfaces.* New York: ACM.

When Policies Are Better Than Plans: Decision-Theoretic Planning of Recommendation Sequences

Thorsten Bohnenberger Anthony Jameson
Department of Computer Science, University of Saarbrücken
P.O. Box 15 11 50, 66041 Saarbrücken, Germany
{bohnenberger | jameson}@cs.uni-sb.de, http://w5.cs.uni-sb.de/~ready/

ABSTRACT

An intelligent user interface sometimes needs to present a sequence of related recommendations to a user, in spite of being uncertain in advance as to whether (and with what success) the user will follow each recommendation. There are potential advantages to the use of decision-theoretic planning methods which yield an optimal *policy* for the situation-dependent presentation of recommendations. This approach is discussed with reference to an example involving route instructions given by an airport assistance system.

KEYWORDS

Intelligent user interfaces, Decision-theoretic planning, Recommendations, Route planning

INTRODUCTION

One type of task that is sometimes faced by an intelligent user interface (IUI) is that of presenting a sequence of related recommendations to a user:

1. The system ($\mathcal{S}$) may give the user ($\mathcal{U}$) directions as to how to get from Location *A* to Location *B* (see, e.g., [6]).
2. $\mathcal{S}$ may give $\mathcal{U}$ instructions for operating a technical device or a software application (see, e.g., [7, 3]).
3. $\mathcal{S}$ may give $\mathcal{U}$ hints on how to navigate in a web site while searching for interesting information (see, e.g., [4], especially Section 4.2).

Previous IUI research on such tasks has focused mainly on the question of how the form and content of a recommendation should depend on factors such as $\mathcal{U}$'s preferences and knowledge. The present paper focuses on a less familiar—but likewise important—aspect of tasks like these: There is often significant uncertainty about what will happen when $\mathcal{U}$ has been given a recommendation. For example, $\mathcal{U}$ may choose not to follow the recommendation; and if she does follow it, her action may not bring about the intended result (e.g., finding the desired piece of information).

In some cases, it is feasible and desirable for $\mathcal{S}$ to plan the entire recommendation sequence in such a way as to take this uncertainty into account. The result will in general be not a *plan* but a *policy* that determines which recommendation should be made in each possible future situation.

Attempts to plan recommendation sequences in this way can draw on a large body of research on methods for decision-theoretic planning (see, e.g., [1, 2]). The present paper describes work in progress on applying these methods to the class of problems just introduced.

EXAMPLE DOMAIN

We consider as an example a hand-held system that gives recommendations (and other information) to a visitor $\mathcal{U}$ in an airport terminal. Suppose that $\mathcal{U}$ wants to go from her present location to a particular departure gate. Instead of giving $\mathcal{U}$ an entire set of recommendations at once, $\mathcal{S}$ beams a single recommendation to her whenever she approaches one of 16 transmitters mounted at various points in the terminal.

Suppose that $\mathcal{U}$ would like to pick up a gift on the way to the gate but would also like to reach the gate relatively soon (e.g., in order to get a good seat assignment). Then $\mathcal{S}$'s navigation recommendations should take into account both (a) the time it will take $\mathcal{U}$ to follow the recommendations and (b) the likelihood that $\mathcal{U}$ will see a suitable gift in one of the shops along the way.

Figure 1 depicts a situation of this sort. There are two possible departure gates on the right, as well as a number of shops. In the center of the terminal, there are four regions of high pedestrian traffic, where walking from one location to the next has a relatively high time cost. Moreover, there are two locations (indicated with larger circles) at which there is a significant danger that $\mathcal{U}$ might make a wrong turn: If $\mathcal{U}$ is instructed to go east or west at these places, with some probability $\mathcal{U}$ will end up going north or south.

$\mathcal{S}$ can choose between two presentation modes for its recom-

IUI'01 January 14-17, 2001, Sante Fe, New Mexico.

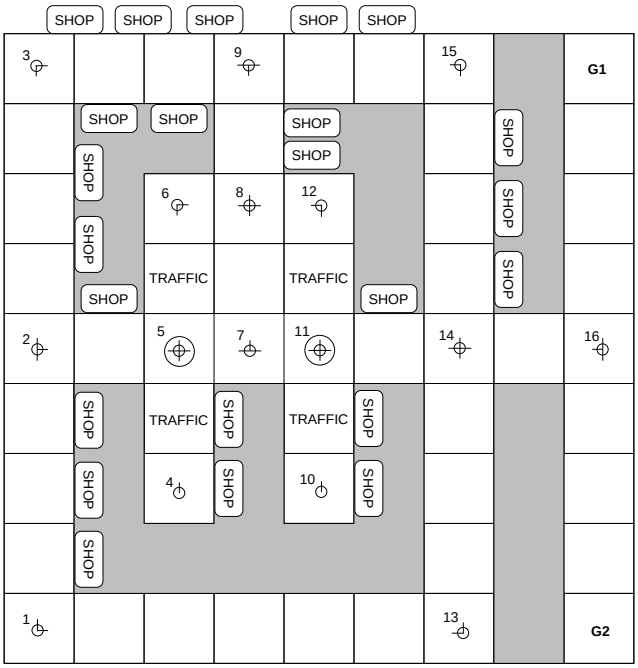

Figure 1. Topology of the airport terminal referred to in the example.

(Each numbered circle represents a transmitter. G1 *and* G2 *are departure gates. At locations 5 and 11, there is a danger that the user $\mathcal{U}$ will take a wrong turn.)*

mendations: (a) *speech mode*, in which $\mathcal{S}$ presents a recommendation in simple speech; and (b) *map mode*, in which $\mathcal{S}$ displays a map of the area around $\mathcal{U}$'s current location. Each map shows the recommended next destination, but it also offers supplementary information about the location and nature of the shops in the area. Speech mode will tend to make $\mathcal{U}$ move faster, but map mode will increase the likelihood that $\mathcal{U}$ will find a gift.

Finally, we assume that, whenever $\mathcal{U}$ is near a transmitter, $\mathcal{S}$ is aware of both $\mathcal{U}$'s current location and whether $\mathcal{U}$ has found a gift (the latter piece of information perhaps being supplied explicitly by $\mathcal{U}$).

MODELING

The situation just described can be modeled straightforwardly in terms of a fully observable Markov decision process (see, e.g., [1, 2]).[1] Each state in the transition model has two *features*: the current location of $\mathcal{U}$ and the information as to whether $\mathcal{U}$ has bought a gift or not ("+" or "–"). So for each of the 16 locations of the transmitters, there are two states.

Each recommendation action comprises a subsequent destination and a presentation mode. As Figure 2 illustrates, each action in a given state leads to one or more possible transitions to a following state. The cost of a given transition is

[1]Readers unfamiliar with the technical concepts involved should still be able to follow the description that follows.

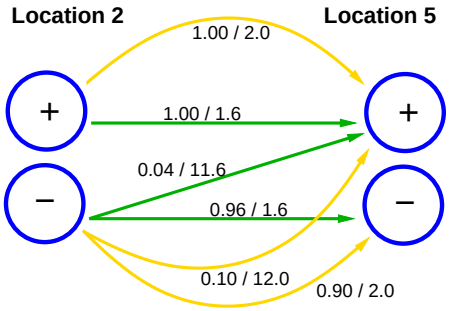

Figure 2. State transitions, probabilities, and costs associated with a recommendation that $\mathcal{U}$ go from Location 2 to Location 5.

(Dark and light arrows represent possible transitions when $\mathcal{S}$ uses speech mode and map mode, respectively. Each pair of numbers next to an arrow gives (a) the probability that the corresponding transition will be made if the recommendation is given; and (b) the cost of making that transition.)

the time that $\mathcal{U}$ will require to reach the corresponding subsequent destination.

The reward for reaching the gate without a gift is 0; and the reward for reaching it with a gift is some nonnegative number that reflects the importance that $\mathcal{U}$ attaches to the goal of finding a gift, relative to the goal of arriving as early as possible at the gate.[2]

The reward and the time costs together determine the overall utility of any given trip by $\mathcal{U}$ to the gate. $\mathcal{S}$'s job is to compute the *policy*—the mapping of states onto recommendation actions—that has the highest expected utility. Our example system uses one of the standard algorithms, *value iteration*, to compute its policies.

SYSTEM BEHAVIOR

Figure 3 shows the policies that were computed for eight situations that differ along two dimensions: $\mathcal{U}$'s "reliability"—i.e., her ability to avoid making wrong turns—and the importance to $\mathcal{U}$ of finding a gift.

1. The two graphs at the top show the recommendations that $\mathcal{S}$ will make in the case where $\mathcal{U}$ attaches absolutely no importance to a gift—or if she has already found a gift (see below). The reliable $\mathcal{U}$ is steered toward the gate as quickly as possible, even when this means that she will pass through one of the locations where other users might make a wrong turn. The less reliable $\mathcal{U}$, by contrast, is drawn away from the central area of the terminal whenever this is possible, because of the danger that she might make a wrong turn and waste time in one of the congested areas. For both users, all of the recommendations are presented in speech mode; the slower map mode can only be worthwhile if there is some value to finding a gift.

2. The graphs in the second row show policies that are suit-

[2]To take into account, for example, the possibility that $\mathcal{U}$ will miss her plane if she arrives after a certain point in time, a more complex modeling of time would be required.

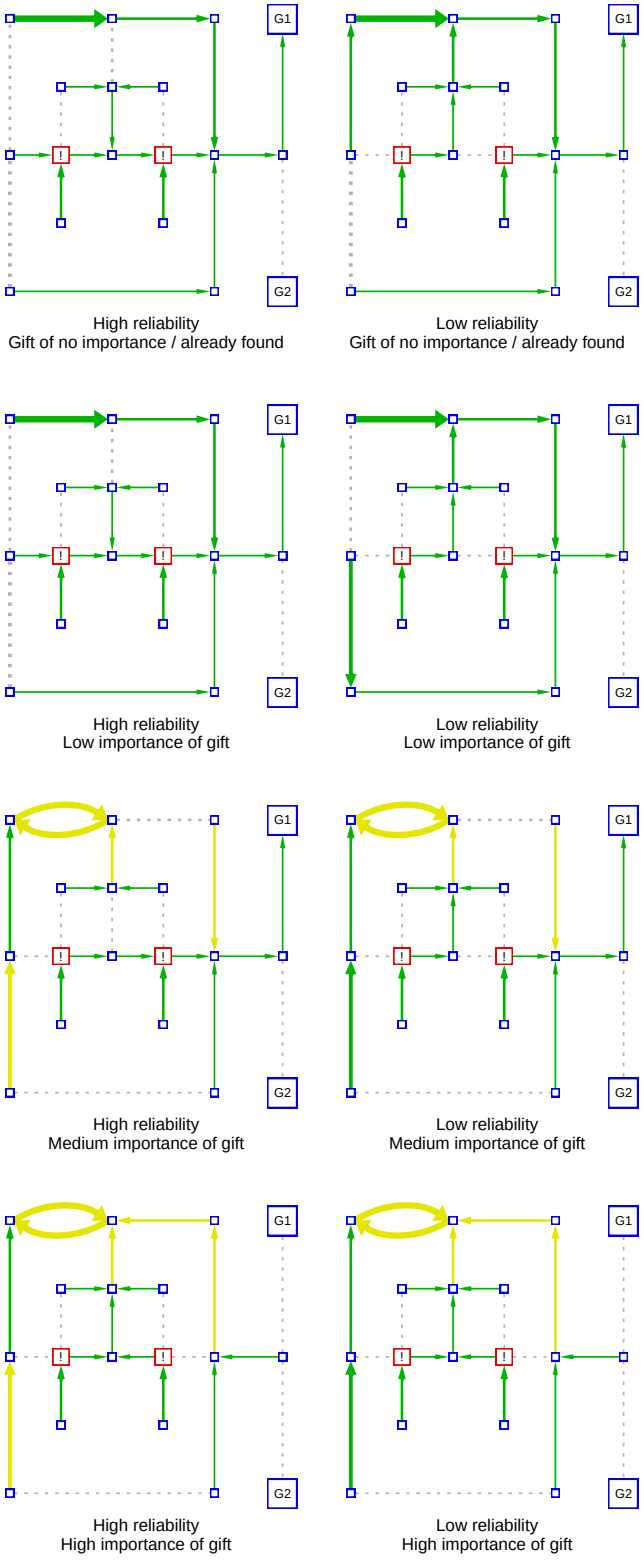

Figure 3. Generated recommendation policies as a function of $\mathcal{U}$'s ability to follow directions and $\mathcal{U}$'s desire to find a present.

(Dark and light arrows represent recommendations given in speech mode and map mode, respectively. Dashed lines show where movements are possible but not recommended. The width of an arrow or a line between two locations reflects the number of shops between them.)

able when $\mathcal{U}$ attaches some (low) importance to finding a gift. (As soon as $\mathcal{U}$ has found a gift, $\mathcal{S}$ will switch to the corresponding recommendations shown in the top two graphs; for clarity, these recommendations are not included in the other graphs.) The reliable $\mathcal{U}$ is still directed to the gate along the fastest possible route. The unreliable $\mathcal{U}$ is still steered away from the dangerous locations, but an interesting change occurs when she is at the middle location on the left-hand side: Instead of directing $\mathcal{U}$ northward, through the area with the most stores, $\mathcal{S}$ sends her southward. The reason is that, if $\mathcal{U}$ passed through the store-rich area, she might stop to buy a gift, although according to her expressed preferences it would not be worthwhile for her to do so.[3]

3. If $\mathcal{U}$ has expressed moderate interest in a gift, she is given some recommendations in map mode (lighter arrows). From some locations, both users are led to the region in the upper left-hand corner with the highest shop density. Here, $\mathcal{S}$ will advise $\mathcal{U}$ to go back and forth, past the shops, until she finds a gift, at which point $\mathcal{S}$ will lead her to the gate in the way shown in the top two graphs. According to the probabilities specified in the model, $\mathcal{U}$ is likely to find a gift quite quickly in this area, so there is no danger of an infinite loop.

4. Finally, if $\mathcal{U}$ has expressed a sufficiently strong interest in finding a gift, $\mathcal{S}$ will direct her toward the area with the highest density of shops, even when she is already very close to her destination G1. In this situation, the time costs of walking around the airport are dominated by the perceived benefits of finding a gift.

DISCUSSION

Ways of Conveying Recommendations

In mobile computing scenarios such as the one considered in our example, the mobile device will often not have enough computing power to perform decision-theoretic planning, and the planning will have to be done on a central computer. There are then three basic ways of conveying the recommendations to $\mathcal{U}$:

1. Each recommendation can be sent when it is needed from the central computer to $\mathcal{U}$'s mobile device. The fact that $\mathcal{S}$ has computed the entire policy in advance will not be directly recognizable to $\mathcal{U}$; but the recommendations will tend to be more useful than those derived with less sophisticated methods.
2. The central computer sends the entire policy to $\mathcal{U}$'s device before $\mathcal{U}$ begins following any recommendations. $\mathcal{U}$'s device performs the relatively easy task of presenting the recommendations specified by the policy. $\mathcal{U}$ can then benefit from sophisticated, situation-dependent recommendations even while she lacks access to any significant computing power.
3. $\mathcal{S}$ transmits a visualization of the entire policy, which $\mathcal{U}$ then applies herself. The question of how best to visualize

[3] The philosophical question of whether it is desirable for a system to lead its user away from temptation exceeds the scope of this paper.

such a policy is a challenging problem of information design. The graphs in Figure 3 show that compact, domain-specific solutions are sometimes available. (In our example, each $\mathcal{U}$ would require two graphs that were somewhat more self-explanatory than the ones shown in the figure.)

Alternative Planning Methods

Besides Markov decision processes, there are other approaches to planning that can deal with some of the issues discussed here (for comparative overviews, see, e.g., [1, 2]). Within the basic classical planning framework, *conditional planners* generate contingency plans that specify different actions for different possible future states (see, e.g., [5]). There is typically less emphasis on taking into account the probabilities and utilities of the various possible outcomes—a central feature of the example discussed above.

Another way of dealing with uncertainty about future states is to interleave planning and execution. For instance, our example system might initially plan a recommendation sequence that assumed that $\mathcal{U}$ would make no wrong turns and then replan the rest of the sequence whenever $\mathcal{U}$ did make a wrong turn. But this approach can yield sequences of events that are highly undesirable and that in principle could be avoided. By contrast, our example system plans so as to avoid undesirable states in the first place (e.g., locations at which $\mathcal{U}$ would be likely to make a costly wrong turn).

Complexity Considerations

Fully observable Markov decision processes of the sort considered here can be solved in time polynomial in the size of the state space and the number of available actions. In our example, doubling the number of locations or adding a new binary feature would double the size of the state space. So although the computing time required for each policy in Figure 3 was on the order of 1 second, more complex models could lead to computation times that were problematic for an interactive system. One general approach to these complexity issues is to look for ways of exploiting the structure of the problem to simplify the planning process (see, e.g., [2]).

In some settings it may be feasible to precompute recommendation policies for a large number of specific situations, so that significant real-time computation is not required.

Different Degrees of Observability

The method illustrated above is applicable only when $\mathcal{S}$ (perhaps with help from $\mathcal{U}$) obtains definite feedback about what happens after $\mathcal{S}$ has given a recommendation. Decision-theoretic planning methods are also straightforwardly applicable if $\mathcal{S}$ receives no feedback at all (see [3] for an example involving instruction sequences). The planning process then takes into account the various things that might happen during the execution of the recommendation sequence, but it must yield a single best plan, since there will be no opportunity for $\mathcal{S}$ to adjust the sequence while presenting it.

The most challenging case is the one where $\mathcal{S}$ receives feedback that does not completely determine the current state—quite a likely scenario in the context of recommendation sequences. The resulting *partially observable Markov decision processes* raise especially serious complexity problems.

CONCLUDING REMARKS

The main contributions of this short paper have been

1. to call attention to the issues and potential advantages associated with the application of decision-theoretic planning methods to the generation of recommendation sequences; and
2. to illustrate these points with an example of an application that provides a useful service that could not easily be provided by other means.

The next steps are

1. to extend the approach along some of the dimensions mentioned above; and
2. to compare this approach systematically, using both theoretical and empirical methods, with alternative ways of dealing with the same type of problem.

ACKNOWLEDGMENTS

This research was supported by the German Science Foundation (DFG) in its Collaborative Research Center on Resource-Adaptive Cognitive Processes, SFB 378, Project B2 (READY). The example application discussed is being worked out in collaboration with Andreas Butz and Antonio Krüger of the project REAL in the same program. We thank the anonymous reviewers for their helpful comments.

REFERENCES

1. J. Blythe. Decision-theoretic planning. *AI Magazine*, 20(2):37–54, 1999.
2. C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94, 1999.
3. A. Jameson, B. Großmann-Hutter, L. March, R. Rummer, T. Bohnenberger, and F. Wittig. When actions have consequences: Empirically based decision making for intelligent user interfaces. *Knowledge-Based Systems*, 13, 2000. In press.
4. T. Joachims, D. Freitag, and T. Mitchell. WebWatcher: A tour guide for the World Wide Web. In M. E. Pollack, editor, *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, pages 770–777. Morgan Kaufmann, San Francisco, CA, 1997.
5. L. Pryor and G. Collins. Planning for contingencies: A decision-based approach. *Journal of Artificial Intelligence Research*, 4:287–339, 1996.
6. S. Rogers, C. Fiechter, and P. Langley. An adaptive interactive agent for route advice. In *Proceedings of the Third International Conference on Autonomous Agents*, Seattle, WA, 1999.
7. W. Wahlster, E. André, W. Finkler, H.-J. Profitlich, and T. Rist. Plan-based integration of natural language and graphics generation. *Artificial Intelligence*, 63:387–427, 1993.