Factored MDP Elicitation and Plan Display

Krol Kevin Mathias, Casey Lengacher, Derek Williams, Austin Cornett
Alex Dekhtyar and **Judy Goldsmith**
{kevin9,clleng0,dwwill0,amcorn3}@uky.edu, {dekhtyar,goldsmit}@cs.uky.edu
Department of Computer Science
University of Kentucky, Lexington, KY

Abstract

The software suite we will demonstrate at AAAI '06 was designed around planning with factored Markov decision processes (MDPs). It is a user-friendly suite that facilitates domain elicitation, preference elicitation, planning, and MDP policy display. The demo will concentrate on user interactions for domain experts and those for whom plans are made.

Introduction

There are two aspects of our software suite that are, as separate components, truly novel. The first is the plan displayer that allows a user to walk through possible state/action sequences using the plan's actions or others. We know of no other displayer of factored Markov decision process policies that are designed with a nontechnical user in mind, yet cover and convey the multitude of possible futures inherent in such a plan.

The second truly novel contribution is the new "shape" for Bayesian network fragment elicitation, called the *bowtie* (Mathias *et al.* 2006). These fragments explicitly represent an action's *success* as a stochastic function of its predictors, and explicitly represent the effects of the action's success.

The other major contribution, which is the main thrust of the demo, is the integration of multiple functionalities for both user and domain expert. The user can specify preferences, walk through plans, and move with ease between these modes. Furthermore, if she has domain-expert permissions, she can edit the preference or planning domains. All of the elicitation functions, both preference and domain, are available qualitatively, although the models are all converted to quantitative representations. The die-hard Bayesian can, of course, input her model as a quantitative Bayesian network.

We have developed a software suite to accompany stochastic planners (Figure 1). The software suite contains elicitation tools, databases, and display tools.

The Domain Instance Elicitation tool (DIET) helps a domain expert describe the possible preference attributes in their domain. DIET elicits preference variables and "archetypes," which are hand-built templates of preferences made up of different attributes and their values, that are used by the Preference Elicitation tool (POET). A user provides her preferences to POET by adjusting values for the attributes of her selected archetype. The Bayes Net Builder tool (BNB) helps a domain expert build Bayesian network domain models. The Probability Elicitation tool (PET) assists a domain expert in inputting conditional probabilities for nodes in the Bayesian network. The BNB and PET together provide a quantitative method for domain elicitation. The High Level Elicitor (HELL) (Mathias *et al.* 2006), developed for our welfare-to-work application, provides a qualitative method for modelling an action as a "bowtie fragment" (see Figure 2) instead of a dynamic Bayesian network (DBN) representation of factored MDPs (Boutilier, Dean, & Hanks 1999). A factored-MDP planner uses the information provided by the elicitation tools to generate a plan. The plan is provided as input to PlanIt, the plan displayer (Williams 2006). PlanIt allows a user to step through the plan, as described in the next section.

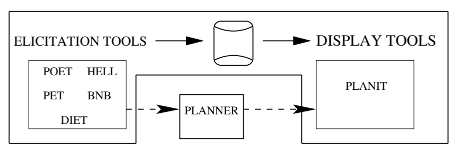

Figure 1: The planning suite.

Copyright © 2006, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

Overview of the Demo

In our demonstration we will provide AAAI/IAAI participants with advice on attending sessions at the conference. In particular, we consider the following MDP model. The actions are attending a particular parallel session or taking a walk instead. The attendee's utility is expressed in terms of topics covered by the talks she hears, and the times of day.

We assume that a talk covers topics $K_1, \ldots, K_j$ stochastically. The probability of a session covering a topic is computed according to the topics listed for the talks in that session.

The state variables represent the sessions (S_i), the saturation for each topic (K_j), and the time (T_k). At time $T = t$, only finitely many actions are enabled; those actions rely on T_{t-1} having been incremented by the plan. (Any action at time t has as one of its effects that T_t changes value from 0 to 1.) There is a variable, “coverage”, that depends on the *number* of topic variables that reach saturation.

The attendee expresses interest levels (capacities) in topics. The attendees also express different habitual preferences. Thus, an attendee chooses an “archetype” preference description, and then edits the suggested preferences as needed. For instance, the archetype SOCIAL might specify a preference for taking walks during all the morning sessions, and on coverage being high. The archetype MODELLING might put high preferences on topics such as preferences, knowledge elicitation, and Bayesian networks.

The planning domain assumes a fixed probability that the user will leave any session before saturation occurs (i.e., the action of attending that session fails); assumes *a priori* that she will attend nearby talks with higher probability than talks that are farther away.

The planner uses factored MDPs. A Markov decision process is a 4-tuple $M = \langle S, A, t, r\rangle$ where S is a finite set of *states*, A a finite set of *actions*, $t : S \times A \times S \rightarrow [0, 1]$, and $r : S \rightarrow \mathcal{R}$. The function $t(s, a, s')$ specifies the probability of ending up in state s', given that you take action a in state s. The function $r(s)$ gives the (positive or negative) reward for being in state s. A factored MDP has a state space expressed as the product of finitely many variables, so a state corresponds to a setting for each variable.

Each action in the planning domain is modelled as a bowtie. For example, the action “attending a session” has a probability of success associated with it. The probability of success is determined by the session characteristics and the attendee’s state. If the action is successful, there is a stochastic increase in coverage of topics. On failure of the action, there is no change in the coverage of topics.

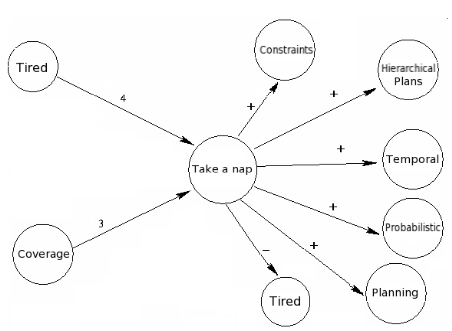

Figure 2: The bowtie model for action “Take a nap”.

The bowtie model consists of a set of input nodes linked to a central action success node. The success node, in return, is linked to the set of output nodes. The bowtie model is similar to the QPN model(Wellman 1990) except that in the bowtie model, a weight is associated with each of the influence factors.

The user will walk through our preference elicitation process by choosing a preference archetype and then editing the suggested preferences as needed in POET. If the user is satisfied with her preferences, she can request a plan. Once the plan is computed, the user can step through her plan in PlanIt. At each step, she can see a representation of her current state, a set of boxes representing different outcomes, and the most likely outcome. The boxes are sized according to probability and colored according to her specified utility. She can click on a box to continue the walkthrough, or can choose a different, unrecommended action to explore.

If she is not happy with the suggested plan, she has many options: to update her preferences and request a replan, or to assume the role of domain expert and edit either the preference domain or the planning domain.

If she chooses to edit the planning domain, for instance by adding a “take a nap” action (Figure 2), we have available the BNB and PET tools for entering or editing dependencies and conditional probabilities for an action. She can also opt for building a qualitative Bayesian network fragments using the HELL tool. This allows her to specify predictors of an action’s success and their relevance (“weight”) and the attributes that are likely to change as a result of the action’s success or failure, and the direction of that change (positive or negative—it is assumed that all domains are ranked). For instance, the success of “take a nap” is likely affected by the attendee’s level of tiredness, and by her overwhelmedness, as measured by the “coverage” variable. If she sleeps, then information from the talks may be transferred from short-term to long-term memory, thus probabilistically increasing any non-zero K_i values. HELL uses this qualitative elicited information to compute a consistent quantitative dynamic Bayesian network.

In summary, ours is the only integrated tool for domain elicitation, Bayesian network management, planning with factored MDPs, preference elicitation, and plan display that is designed to be accessible to noncomputer-scientist users and domain experts. The domain elicitation has as an option of a qualitative-to-quantitative process based on a highly intuitive and accessible model of action effects called the bowtie model. The plan displayer makes use of an innovative method of walking a user through a plan with many possible outcomes at each step.

References

Boutilier, C.; Dean, T.; and Hanks, S. 1999. Decision-theoretic planning: Structural assumptions and computational leverage. *J. Artif. Intell. Res. (JAIR)* 11:1–94.

Mathias, K. K.; Isenhour, C.; Dekhtyar, A.; Goldsmith, J.; and Goldstein, B. 2006. Eliciting and combining influence diagrams: Tying many bowties together. Technical Report 453-06, University of Kentucky Dept. of Computer Science.

Wellman, M. P. 1990. Fundamental concepts of qualitative probabilistic networks. *Artif. Intell.* 44(3):257–303.

Williams, D. 2006. Interactive preference elicitation and plan visualization. Master’s thesis, University of Kentucky.