A Probabilistic Model of Plan Recognition*

Eugene Charniak
Department of Computer Science
Brown University
Providence RI 02912-1910
ec@cs.brown.edu

Robert Goldman
Department of Computer Science
Tulane University
New Orleans, LA 70118-5698
rpg@cs.tulane.edu

Abstract

Plan-recognition requires the construction of possible plans which could explain a set of observed actions, and then selecting one or more of them as providing the *best* explanation. In this paper we present a formal model of the latter process based upon probability theory. Our model consists of a knowledge-base of facts about the world expressed in a first-order language, and rules for using that knowledge-base to construct a Bayesian network. The network is then evaluated to find the plans with the highest probability.

Introduction

Plan recognition is the problem of inferring an agent's plans from observations. Typically the observations are actions performed by the agent, and previous work has limited itself to this case. We will do so as well, although the approach we suggest generalizes to other types of observations (states of affairs, actions of others, etc.). A plan-recognition system must be able to retrieve, or construct, possible explanatory plans, and decide to what degree the evidence supports any particular plan hypothesis. In what follows we will concentrate exclusively on the second of these tasks.

Probably the best-known work in the area is that of Kautz [8,9]. Kautz provides a formal basis for the problem in terms of minimizing (in the circumscriptive sense) the number of "top-level plans." The idea is that every observed action is part of one or more top-level plans. With this restriction plan-recognition becomes a task of nonmonotonic deduction. For example, if action A_1 can only be part of P_1 or P_2, while A_2 can only be part of P_2 or P_3, after asserting that there is only one top-level plan, it follows deductively that A_1 and A_2 are both part of P_2. While this work is justly noted for the clarity with which it lays out its logical and algorithmic basis, as a theory of plan-recognition it suffers from three major flaws.

First, because this approach is, essentially, minimal set covering, it cannot decide that a particular plan, no matter how likely, explains a set of actions, as long as there is another plan, no matter how *unlikely*, which could also explain the observed actions. Consider

> Jack packed a bag. He went to the airport.

Any normal reader would assume that Jack is taking a plane-trip. But Kautz's plan-recognition system would not be able to decide between plane-trip, and air-terrorist-bombing, since the latter also has the terrorist packing a bag (with a bomb) and going to the airport (to get the bag on the plane). Bayesians (such as your authors) would say that Kautz's program ignores the priors on the plans, and thus cannot make the right decision. Nor can one simply say that the program should defer the decision. There will be cases where a decision must be made. Imagine the following:

> Person1: Is Jack around?
> Person2: I saw him leave for the airport with his bag this morning.
> Person1: Yes, but is he around?

Person1 must ask this question because Jack might be bombing the plane, rather than flying on it, and if he is bombing the plane he will not have boarded, so Jack might still be in town.

Second, the distinction between top-level plans, which are minimized, and the rest, which are not, is problematic. While most of the time walking is in service of a higher plan (getting some place), occasionally we walk "because we feel like it." So sometimes walking is a top-level plan, and sometimes it is not. Nor can Kautz do away with the distinction. If one minimized over all actions, then one would never postulate a top-level plan at all.

Finally, set minimization as a principle for abduction (reasoning from effects to causes) is simply wrong. In medicine it may be correct to diagnose two common ailments rather than a single uncommon one, particularly if the symptoms are strongly associated with the common ailments, but not with the uncommon one.

Indeed, because of such objections, and particularly because of the need for priors, we doubt that any model which does not incorporate some theory of reasoning under uncertainty can be adequate. The only other

*This work has been supported by the National Science Foundation under grant IRI-8911122 and by the Office of Naval Research, under contract N00014-88-K-0589.

such model that we are aware of is Carberry's [1], which uses Dempster-Shafer (D-S) belief functions. We applaud her decision to confront the issues of reasoning under uncertainty and her use of numerical measures for belief, and thus the criticisms we have of her work are of a much narrower sort. These criticisms have to do with our preference for probability theory over D-S for this problem, and the architecture of her approach.

We prefer Bayesian probability theory to D-S on general principles. It is instructive to note that many attempts to give a semantics to D-S belief functions do so by relating them to probabilities, thus admitting that probability is on a much firmer footing than "beliefs." Given this, it seems to us that it is incumbent on anyone using D-S to justify why probability theory would not do. Carberry has not.

Carberry makes two arguments against probability theory. First she claims that probability theory requires "a great deal of time-consuming and complicated computation." In response we note that while evaluating Bayesian networks is NP hard, so are D-S calculations. In fact, since point-valued probability theory is a limit case of the D-S calculus, D-S calculations are *at least as* expensive as probability updating. Furthermore, Carberry's updating problems involve extensive independence assumptions, and she assumes that an action is part of only one plan. If these computations were rephrased in terms of Bayesian probability, they could be computed in linear time.

Carberry's second complaint is that "probability computations are extremely difficult to explain and justify to a lay person." We have yet to see an argument that Dempster's rule of combination is more intuitively comprehensible than Bayes' rule.[1] In any case, one does not explain the computations, one explains their *results*. This is what Carberry does with her D-S computation, and we would do likewise.

Lastly, we have subsumed more of our system's reasoning within our uncertainty calculus than Carberry has done with hers. In particular, several of Carberry's heuristic rules are not needed in our system.[2]

Our work on plan recognition is in the context of Wimp3, our Bayesian story understander. Earlier publications have discussed its knowledge representation [4], problems assessing its probabilities [3], and rule-based network construction [6]. Here we concentrate on plan recognition. However, our scheme interacts with other language tasks like pronoun resolution, word-sense disambiguation, etc., something the above researchers have yet to attack.

Our Plan Representation

Our model of plan recognition consists of a knowledge-base K of facts about the world, which is used in the production of a set of Bayesian networks, $\mathcal{P}$, called "plan-recognition Bayesian networks." K is expressed in a first-order notation. The language consists of terms denoting actions and physical objects (which will have names consisting of a type name, followed by a distinguishing number e.g., buy43, milkshake2), terms denoting sets of action and object natural kinds (type name followed by a hyphen, e.g. buy-, milkshake-), functions from terms to other terms, typically used to indicate roles, (either common case names e.g., agent, patient, or ending with "-of", e.g., straw-of, so we might have (agent buy43) denoting the agent of the event buy43), the three predicates ==, inst, and isa; and the usual truth-functional predicates.

The predicate inst (set membership), as in (inst buy43 buy-) says that the first term is a member of the set of buying events buy-. The predicate isa, (subset) as in (isa buy- obtain-) says that the first term is a subset of the second. The predicate == (the better name relation) as in (== (agent buy43) jack2) says that the second argument is a better name for the first. The better name relationship is used for noun-phrase reference, as in (== it22 milk8) would be the proposition that milk-8 is a better name for it22. This would be true if the pronoun "it" which gave rise to it22 referred to the milk in question. We abbreviate == to = when space is important.

We make no distinction between plans and actions. Plans are complicated actions — those with sub-actions. The sub-actions of a plan relate to the plan in the same way that roles do, as functions from the plan-instance to the sub-action. So (== (pay-stp shop55) pay76) indicates that the paying event, pay76, was the "pay step" of the shopping event, shop55. Thus, in this notation, recognizing a plan p from sub-actions $a_1 \ldots a_n$ is inferring the following:

$$\begin{array}{ll} (\text{inst } p \textit{ plan-type}) & (== (stp_1\ p)\ a_1) \\ \ldots & (== (stp_n\ p)\ a_n) \end{array}$$

Making such inferences requires generic information about plans and their parts. This information can be represented using the above predicates, functions and terms. For example, to state that actions of type A can fill the A-*of* slot in of plans of type P, and that if it does then the A_1-*of* slot of A must be filled by the entity which fills the P_1-*of* slot of P, and the A_2-*of* by P_2-*of* etc would come out as this:

$$\begin{array}{l} (\text{inst ?P } P) \rightarrow (\text{and } (\text{inst } (A\text{-}of\ ?\text{P})\ A) \\ \qquad\qquad\qquad\qquad (== (A_1\text{-}of\ (A\text{-}of\ ?\text{P})) \\ \qquad\qquad\qquad\qquad\quad (P_1\text{-}of\ ?\text{P})) \ldots) \end{array}$$

For example:

```
(inst ?shop shopping-) →
    (and (inst (go-stp ?shop) go-)
         (== (agent (go-stp ?shop)) (agent ?shop))
         (== (destination (go-stp ?shop))
             (store-of ?shop)))
```

[1] In fact, there has recently been very promising work on explaining probability updating in networks [7].

[2] Rules D4,D5, and D6, for those familiar with her paper.

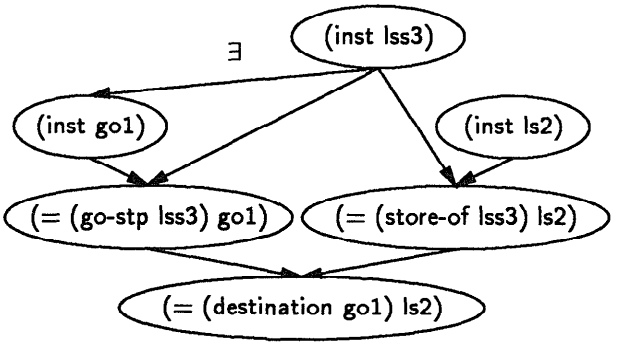

Figure 1: A Bayesian network for plan-recognition

The Probabilistic Model

Given this model of plan schemas (as actions which can be broken down into distinct sub-actions), and plan recognition (finding the plan instances for which the given actions are sub-actions) we can now talk about how a probabilistic model can make choices between competing hypotheses.

Figure 1 shows a Bayesian network for plan recognition. In this section we first informally explain what is going on in this network, and then formally describe the class of Bayesian networks which we use. We assume that the reader is already familiar with the details of Bayesian networks.

The random variables in our network are of two sorts. Most denote propositions, i.e., boolean variables. The only exception are random variables which indicate types of things (e.g., (inst go1) in Figure 1). The sample space for this type of variable is the set of primitive types (types with no sub-types) in the isa hierarchy. Thus the possibility that, go1 is a driving event would be expressed as (inst go1) $\in$ drive-.

Figure 1 is a the network constructed to represent the possibility that someone's going to a liquor store (ls2) is part of a liquor-store shopping event (lss3). At the top of the network is the hypothesized high-level plan itself. Our belief in this plan-recognition hypothesis will be expressed by the probability that the value of this random variable is liquor-store-shopping-. The hypothesized plans are typically root nodes in the Bayesian network so we need to specify their priors. This is done according to the semantics outlined in [4] by which lss3 is a random variable (in particular, a Bernoulli trial) with a sample space consisting of all entities in the world. We assume a large, but finite domain D of equiprobable elements. Thus the prior of (inst i) = *type* is $\frac{|type|}{|D|}$. If an inst is not a root (e.g., (inst go1) in Figure 1) its distribution is given by the "existential" rules described below.

Below the node denoting the plan are propositions describing the entities in the input which are to be "explained" by fitting them into the high-level plan as slot fillers. In Figure 1 these are

(inst go1), (== (go-stp lss3) go1), (inst ls2), (== (store-of lss3) ls2).

That is, we have two entities, go1 and ls2, which the two slot-filler statements (the two equality statements) fit into the lss3 plan. As we noted earlier in our discussion of the basic model, the first of these, (== (go-stp lss3) go1) constitutes the *direct* explanation of the going event in terms of the higher-level plan of shopping at a liquor store.[3]

Next, note the arcs from the inst statements to the slot-filler statements involving the instances (e.g., the two arcs into (== (store-of lss3) ls2)). These arcs are there to capture the fact that the probability that two entities ((store-of lss3) and ls2) are the same is zero if they are of different types, and $\frac{1}{|t|}$ ($= \frac{P(=)}{P(t)}$) if they are both of type t. (Remember that we are assuming a set of equiprobable elements.)

There is an extra arc (marked with a $\exists$ in Figure 1) from the plan to the inst node of one of its slot-fillers, go1. We call this the existential slot-filler. It is distinguished from the rest of the plan's slot fillers by having a different probability distribution at its inst node and slot-filler proposition, (== (go-stp lss3) go1). To see the need for this, consider the difference between how this going event fits into lss3, and how, say a subsequent paying event might fit into it. When we see the going event we have no shopping hypothesis. Thus when we create it, the going event serves to "define" it in the sense that if there is a shopping event here, it is the one into which this going event fits (as the go-stp). The random variable lss3 is akin to an existential variable scoped within go1. Thus, the probability of the inst being the appropriate type, and filling the slot, given that lss3 is a shopping event is 1: the probability that go1 fills the go-step slot of the liquor-shopping plan which explains go1. We call this the ***existential*** slot filler because of the ***exists*** in the interpretation, and the arc is called an "up existential" arc. (We will shortly encounter "down existentials " as well.)

On the other hand, consider the possibility that a subsequently mentioned paying event fills the pay-stp in lss3. Here lss3 already exists, and thus cannot be interpreted as, by definition, the shopping plan into which the paying fits. Rather we must interpret this as a "random" paying event which might not fit into lss3, even if lss3 is of the correct type. Thus the probability that the paying event fills the (pay-stp lss3) is not the probability that such a paying event exists (given lss3 is a shopping), but rather the probability that it exists, ***and that the one mentioned is the one for*** lss3. We did not have this second part for go1 because it was true by definition of lss3.

We would like to point out two complications. First, we do not allow up-existentials from objects to

[3]This direct explanation might be part of a more complicated explanation such as the agent's plan to buy refreshments for a party.

actions. For example, we could not define lss3 as the liquor-store shopping which occurred at ls2 since there are many such events. Second, in some more complicated cases we also allow "down existential" arcs from a plan down to an otherwise undefined sub-part. This *may* be between actions and objects, since (in our axiomatization) there is only one filler for each slot.

Now let us consider evidence other than the presence of objects. Items of evidence for a hypothesis are facts in the input which are (probabilistically) implied by the hypothesis.[4] For example, in Figure 1 the statement (== (destination go1) ls2) is evidence for the shopping hypothesis. It is evidence because if one is shopping, then one will go to the store at which one is shopping. Note that this fact would also constitute evidence in favor of any other hypothesis which predicted going to a liquor-store. Later we will address the issue of how the model handles competing hypotheses.

We will now give a formal definition of our networks. In this description we will use e_b^a to denote the edge from a down to b. In what follows we will only give the probabilities for the case where there is only a single proposed plan in the network. In the case of more than one the joint distribution will be that specified by the "noisy or" gate [10]. (Those unfamiliar with this can simply think of it as normal or-gate, i.e., the probability is 1 iff at least one of the plans is true.)

The set of "Plan-recognition Bayesian networks" ($\mathcal{P}$) for a knowledge-base K, and a set of instances I, are Bayesian networks (which are pairs (N, E) of nodes and edges) defined as follows:

1. (Basis) $(\{\}, \{\}) \in \mathcal{P}$
2. (Object Evidence) Let $(N, E) \in \mathcal{P}$, and $b = (\text{inst } j)$ $(j \in I)$, then $(\{b\} \cup N, \{E\}) \in \mathcal{P}$. Any formula introduced this way is part of the evidence.
3. (Up-Existential) Let $(N, E) \in \mathcal{P}$. If $b = (\text{inst } j) \in N$, $(\text{inst } ?\text{i } t_1) \rightarrow (\text{inst } (\textit{slot } ?\text{i}) \; t_2) \in K$, $i, j \in I$, and either t_2 is an event or t_1 is an object, then $(N \cup \{a, c\}, E \cup \{e_c^a, e_c^b, e_b^a\}) \in \mathcal{P}$, where $a = (\text{inst } i)$, $c = (= (\textit{slot } i) \; j)$, and $a, c \notin N$. The probability of $b = t_2$ given that $a \subset t_1$ is 1. The probability of $b = t_2$ given that $a \not\subset t_1$ is $\mathrm{P}(t_2) - \mathrm{P}(t_2 | a \subset t_1) \cdot \mathrm{P}(a \subset t_1)$. The probability of c given that $a \subset t_1$ and $b \subset t_2$, is 1, else 0.
4. (Down-Existential) Let $(N, E) \in \mathcal{P}$. If $a = (\text{inst } i) \in N$ $(\text{inst } ?\text{i } t_1) \rightarrow (\text{inst } (\textit{slot } ?\text{i}) \; t_2) \in K$, and $i, j \in I$, then $(N \cup \{b, c\}, E \cup \{e_c^a, e_c^b, e_b^a\}) \in \mathcal{P}$, where $b = (\text{inst } j)$, $c = (= (\textit{slot } i) \; j)$, and $a, c \notin N$. The probabilities are defined as in the up-existential case.
5. (Slot-filler) Let $(N, E) \in \mathcal{P}$. If $a = (\text{inst } i)$, and $b = (\text{inst } j) \in N$, $(\text{inst } ?\text{i } t_1) \rightarrow (\text{inst } (\textit{slot } ?\text{i}) \; t_2) \in K$, and $i, j \in I$, then $(N \cup \{c\}, E \cup \{e_c^a, e_c^b\}) \in \mathcal{P}$, where $c = (= (\textit{slot } i) \; j)$. The probability of c given $a \subset t_1, b \subset t_2 = \frac{P(=)}{P(t_2)}$, else 0.

[4] I.e., the propositions E such that $P(E|\textit{hypothesis}) > P(E)$.

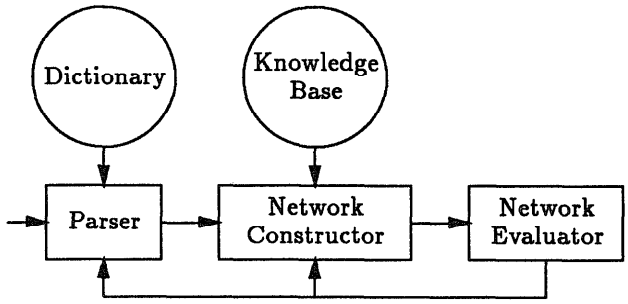

Figure 2: Structure of Wimp3

6. (Other Evidence) if $(N, E) \in \mathcal{P}$, and $\{a_1, \ldots, a_i, =_1, \ldots, =_n\} \subset N - \mathcal{E}$ where $\mathcal{E}$ is the evidence introduced by rule 2, all $=_i$ are equality statements, and $A_1, \ldots, A_i \rightarrow C$ is a member of K, such that after using the equality statements, $a_1, \ldots, a_i$ unifies with the antecedent of the rule, then the network created by adding C plus edges from all of $a_1, \ldots, a_i, =_1, \ldots, =_n$ to C is a member of $\mathcal{P}$, provided that if C is a member of N, then it was introduced by this rule. The probabilities are given by the rule.
7. Nothing else is in $\mathcal{P}$.

Creating Hypotheses

We have built a program, Wimp3, which automatically creates the networks we have described. Wimp3 deals with more than just plan recognition. It produces such networks directly from English narratives, and thus must (and can) deal with other problems such as pronoun reference, word-sense ambiguity, case ambiguity, and syntactic ambiguity. In what follows we will concentrate on how Wimp3 goes about plan recognition and ignore these other issues. It should be noted, however, that the machinery we describe for plan recognition is also used in these other tasks.

The structure of Wimp3 is shown in Figure 2. Words from a narrative are fed to the parser one at a time, and a syntactic parse is produced (as a series of first-order formulas) which are then fed to "Frail," our inference engine. Frail produces the Bayesian network. Directly responsible for this are a set of forward chaining rules which produce not simply propositions, but Bayesian networks [6].

The forward chaining rules pretty much follow the formal description of $\mathcal{P}$.[5] The major difference between the forward-chaining rules which *construct* $\mathcal{P}$s and the rules which *define* them is that the former contain restrictions on when to apply them. These restrictions are there to avoid the construction of useless sections of the network.

[5] Although currently Wimp3 does not have down-existentials, they would be easy to add.

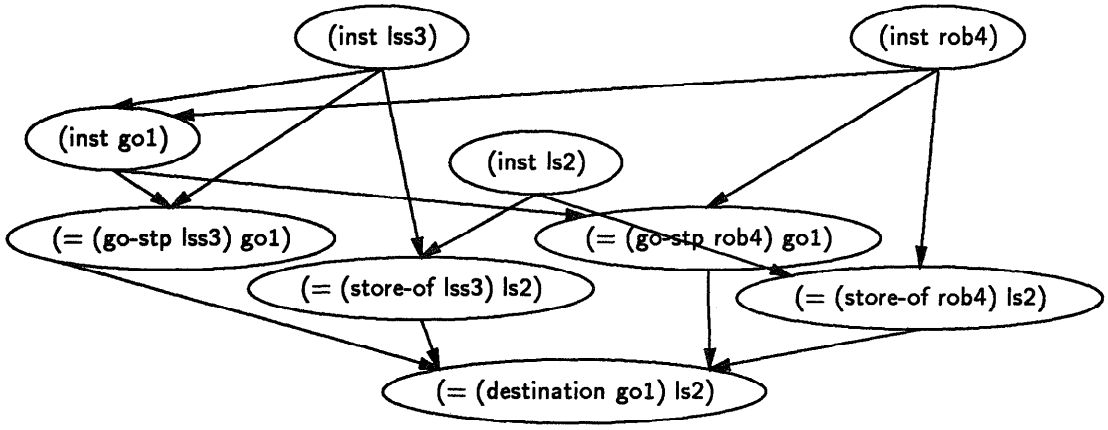

Figure 3: Competition between shopping and robbing

For example, consider the rule for up-existentials. The description of $\mathcal{P}$ says that if we have a token j, and a rule stating that things of type t_2 can fill slots in things of type t_1, then propose a thing of type t_1 of which j fills the appropriate slot. The corresponding rule also requires a) some suggestion that j is, in fact, of type t_2, and b) the approval of a marker passer [2] which is responsible for proposing hypotheses.

Competing Hypotheses

In a story like "Jack went to the liquor store. He pointed a gun at the owner," Wimp will believe with high probability that Jack is shopping at a liquor store after reading just the first line.[6] It does this even though it has in its knowledge-base a plan for robbing stores. After reading the second sentence, Wimp will "change its mind," and decide that the robbery hypothesis is more likely. It is examples such as this that to us most decidedly argue against a non-monotonic logic approach, and for a system which uses numbers to encode the prior probabilities of events. In this section we will examine this example in more detail, to show how Wimp "changes its mind."

Given this story, and a knowledge-base containing both liquor-shop and rob- the competition between them would be expressed by the Bayesian network shown in Figure 3. The two explanations, lss3 and rob4, compete for belief because they both serve to explain the same evidence, (= (destination go1) ls2). The node for this latter proposition is, as we have said above, a noisy-or node. That is to say, this proposition can be explained either by the hypothesis lss3 or[7] the hypothesis rob4. Evidence at or-nodes in Bayesian networks gives support to all their possible causes roughly in proportion to how probable these explanations are based on any other evidence observed.[8] In the case at hand this means that the evidence (= (destination go1) ls2) will support lss3 and rob4 roughly in proportion to the prior probabilities of shopping and robbing liquor stores, since there is no other evidence. Since shopping is much more common, we find that the probability assigned to it is quite high (about .8 in our system) while that for rob4 is only slightly above its prior probability (currently about 10^{-6}).

Actually, Wimp does not construct the network shown in Figure 3 after seeing "Jack went to a liquor-store." Rather the network it constructs is the simpler one shown in Figure 1, without an explicit robbery hypothesis at all. The reason for this is that Wimp uses a marker passing scheme to find potential explanations, and the scheme weeds out the robbery hypothesis. In [2] it is shown that the marker passer is sound with respect to the probabilistic evaluation mechanism in that the number it computes as a cut-off mechanism for its search is an upper bound on the probability of the hypothesis (given certain reasonable assumptions above the evidence). Thus while the marker passer will sometimes suggest an hypothesis which does not pan out, if it does not find an hypothesis, it means that the hypothesis would have been rejected even it if had been proposed. This is what happens to robbery.

The situation, however, is quite different after reading "He pointed a gun at the owner." Figure 4 shows a part of the network which Wimp constructs after "pointed a gun."[9] From an intuitive point of view, the fact that Jack pointed a gun suggests robbery, and does so even more strongly after we learn that he is pointing it at the owner of the liquor store. This tips the balance in the competition over the evidence that (= (destination go1) ls2) so that now this evidence is seen

[6] We apologize for the anthropomorphic terminology; it is simply more succinct.

[7] *inclusive*

[8] This follows simply as a consequence of Bayes' law.

[9] We have omitted the section of the network concerned with the pronoun reference of "He" and the evidence provided by the fact that Jack is the agent of both activities.

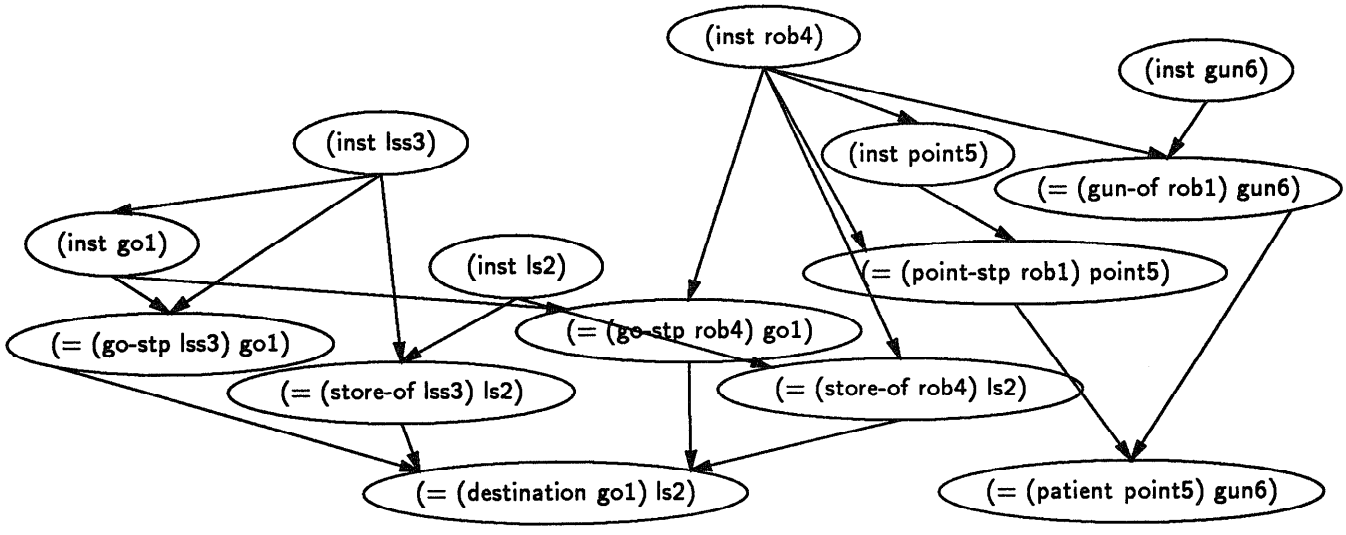

Figure 4: Competition after pointing a gun

to support rob4, not lss3, and the posterior probability of lss3 now goes down.

Conclusion

A crucial component of any plan-recognition system is the ability to decide between competing hypotheses. We have presented a model of this process in which a Bayesian network evaluates the conditional probability of the competing hypotheses given the evidence. This model has been implemented within the Wimp3 story understanding system and it is completely compatible with the other decision processes of Wimp3, such as pronoun referent resolution, word-sense disambiguation, case-determination, and syntactic ambiguity resolution. Wimp3 as been subjected to a single-blind test in which it had to pair-up stories (which had the same plans) after being debugged on half of each pair. The results are reported in [5]. To summarize these results, the program correctly paired 19 out of 25 after 3 lexical items were added to its lexicon, and 24 out of 25 after the further addition of 4 formulas to the knowledge base. The remaining example generated a network too large to evaluate, pointing to what remains the most important impediment to the scheme we have proposed.

References

1. CARBERRY, S. Incorporating default inferences into plan recognition. Presented at *Proceedings of the Eighth National Conference on Artificial Intelligence* (1990).
2. CHARNIAK, E. AND CARROLL, G. A Probabilistic Analysis of Marker-Passing Techniques for Plan-Recognition. Department of Computer Science, Brown University, Technical Report, 1991.
3. CHARNIAK, E. AND GOLDMAN, R. Plan recognition in stories and in life. Presented at *Workshop on Uncertainty in Artificial Intelligence* (1989).
4. CHARNIAK, E. AND GOLDMAN, R. P. A semantics for probabilistic quantifier-free first-order lanuages, with particular application to story understanding. Presented at *IJCAI-89* (1989).
5. GOLDMAN, R. A Probabilistic Approach to Language Understanding. Department of Computer Science, Brown University, Technical Report, 1991.
6. GOLDMAN, R. AND CHARNIAK, E. Dynamic construction of belief networks. Presented at *Proceedings of the Conference on Uncertainty in Artificial Intelligence* (1990).
7. HENRION, M. AND DRUZDEL, M. J. Qualitative propagation and scenario-based approaches to explanation of probabilistic reasoning. Presented at *Proceedings ofl the Sixth Conference on Uncertainty in Artificial Intelligence* (1990).
8. KAUTZ, H. A Formal Theory of Plan Recognition. University of Rochester, Technical report, Rochester N.Y., 1987.
9. KAUTZ, H. AND ALLEN, J. Generalized plan recognition. Presented at *AAAI-86* (1986).
10. PEARL, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, Los Altos, Calf., 1988.