A Bias towards Relevance: Recognizing plans where goal minimization fails

Abigail S. Gertner
Learning Research & Development Center
University of Pittsburgh
Pittsburgh PA 15260
gertner+@pitt.edu

Bonnie L. Webber
Department of Computer & Information Science
University of Pennsylvania
Philadelphia PA 19104
bonnie@linc.cis.upenn.edu

Abstract

Domains such as multiple trauma management, in which there are multiple interacting goals that change over time, are ones in which plan recognition's standard inductive bias towards a single explanatory goal is inappropriate. In this paper we define and argue for an alternative bias based on identifying contextually "relevant" goals. We support this claim by showing how a complementary planning system in TraumAID 2.0, a decision-support system for the management of multiple trauma, allows us to define a four-level scale of relevance and therefore, of measurable deviations from relevance. This in turn allows definition of a bias towards relevance in the incremental recognition of physician plans by TraumAID's critiquing interface, TraumaTIQ.

Introduction

Domains such as multiple trauma management, in which there are multiple interacting goals that change over time, are ones in which plan recognition's standard inductive bias towards a single explanatory goal is inappropriate. Yet some kind of bias is nevertheless necessary if plan recognition is to identify a best explanation for observed actions. In this paper, we describe how plans produced by a complementary planning system allow us to define an alternative bias towards contextually relevant goals, along with a four-level scale for relevance, which is used in the incremental recognition and evaluation of physician plans. These functions are carried out by TraumAID's interface, TraumaTIQ, which uses them to produce *critiques* of physician orders in only those cases where it could make a clinically significant difference.

The task of TraumaTIQ's plan recognizer is to build incrementally a model of the physician's plan based on the actions she has ordered. TraumaTIQ then evaluates that plan and compares it to TraumAID's plan in order to determine potential errors to comment on in the critique. The plan evaluation and critique generation components will not be described in this paper. They are discussed in detail in (Gertner 1995).

In the next section, we introduce TraumAID 2.0 and describe the representation of planning knowledge and the process by which it generates plans. We then describe the plan recognition algorithm used by TraumAID's critiquing module, TraumaTIQ, and show how the planning knowledge in TraumAID provides a recognition bias based on relevance. We conclude with a discussion of an evaluation performed on TraumaTIQ's plan recognition algorithm and its implications for further system development.

An Overview of TraumAID 2.0

The TraumAID system is a tool for assisting physicians during the initial definitive management of patients with multiple trauma (Rymon 1993; Webber, Rymon, & Clarke 1992). During this phase of patient care, which often requires urgent action, preliminary diagnoses are pursued and initial treatments are carried out. The current system, TraumAID 2.0, embodies a goal-directed approach to patient management. The system architecture links a rule-based reasoner that derives conclusions and goals from the evidence currently available about the patient, and a planner that constructs a (partially ordered) plan for how best to address the currently relevant goals.

TraumAID 2.0's management plans have been retrospectively validated by a panel of three experienced trauma surgeons in a blinded comparison with actual care. Panel members preferred TraumAID's plans over actual care to a statistically significant extent (Clarke *et al.* 1993; Gertner, Webber, & Clarke 1996). This result suggests that such plans could provide a valid basis for producing critiques of physician plans which could lead to improvements in patient care.

To understand how general knowledge and patient-specific information in TraumAID's planner allow us to define and use an inductive bias towards contextually relevant goals in TraumaTIQ's plan recognition, it is important to understand how TraumAID forms goals and clinically appropriate plans for addressing them.

When a new piece of evidence is entered, TraumAID's reasoner is triggered, forward chaining through its entire set of rules and generating a list of active goals. When rule activity ceases, the planner is invoked to determine how best to satisfy the current combination of management goals and address the competing diagnostic and therapeutic needs arising from multiple injuries.

TraumAID's plans are constructed out of three types of objects: *goals*, *procedures*, and *actions* (see Figure 1).

Part of TraumAID's general knowledge of goals consists of *goal-procedure mappings* – disjunctive lists of procedures for addressing each goal. Procedures in a mapping are ordered preferentially by their cost, effectiveness, invasiveness, etc. For example, the goal NEED ACCESS CHEST CAVITY can be addressed by either PERFORM THORACOTOMY or PERFORM BILATERAL THORACOTOMY WITH TRANSVERSE STERNOTOMY, but the former is preferred.

Given a set of goals, TraumAID's planner selects one procedure for each goal from its goal-procedure mapping. Selection depends on both the *a priori* preference ordering and a more global need to address multiple goals efficiently, since one procedure can sometimes be used to address more than one goal.

A procedure comprises an ordered sequence of actions and/or sub-goals, stored in a *procedure-action mapping*. The use of sub-goals allows TraumAID's planner to *delay* certain decisions about how to address top-level goals. For example, if TraumAID is planning to address the goal TREAT UPPER THORACIC ESOPHAGEAL INJURY by performing PERFORM UPPER ESOPHAGUS REPAIR, it can commit early on to its specific component actions, GIVE ANTIBIOTICS and ESOPHAGUS REPAIR AND DRAIN, while basing its choice of how to address NEED ACCESS CHEST CAVITY on the other currently relevant goals.

Another feature of TraumAID's goal posting and planning is that its reasoner embeds a conservative, *staged* strategy for selecting diagnosis and treatment goals (Rymon 1993): goals whose satisfaction requires expensive and/or risky procedures are not included in a plan until they are justified by less costly tests or observations. These strategies appear in the knowledge base as implicitly related management goals, such as a DIAGNOSE HEMATURIA (blood in the urine), which if present, triggers DIAGNOSE BLADDER INJURY, which in turn can lead to a goal TREAT BLADDER INJURY.

Using context to bias search in plan recognition

Intelligent interaction with another agent often depends on understanding the agent's underlying *mental states* that lead her to act as she does. These mental states include *beliefs* about the world, *desires* for the future state of the world, and *intentions* to act in certain ways. The process of inferring these mental states is generally referred to as plan recognition.

The importance of plan recognition for automated decision support has been recognized by both ourselves and Shahar and Musen (Shahar & Musen 1995). In connection with automated decision support, plan recognition can support several elements of critiquing, including flexible plan evaluation, explanation of critiques, and proposing alternative actions and goals.

Since there are theoretically many possible explanations for any set or sequence of observations, plan recognition requires an inductive *bias*. Previous plan recognition algorithms, most notably (Kautz 1990), incorporated a bias towards minimizing the number of goals used to explain the observed actions. Such a bias is inappropriate in a domain such as multiple trauma management where, as discussed in the preceding section, a range of independent diagnostic and therapeutic goals may be active simultaneously.

Other factors also constrain the kind of bias that can be used: physician orders (which serve the role of observed actions) are not necessarily given and entered in the order in which they are intended to be performed. TraumaTIQ therefore cannot assume that consecutive orders address the same or similar goals. In addition, physicians' plans are not always correct. Since the set of incorrect plans is too large to encode *a priori*, a bias is needed that will still allow interpretation of orders that do not correspond exactly with its knowledge of clinically appropriate plans.

Given these constraints, TraumaTIQ's plan recognizer employs a bias towards *relevance*, attempting to explain physician orders as closely as possible in conformance with the principles of trauma care encoded in TraumAID. TraumAID's current goals and plan then provide a standard of relevance, with ways of interpreting deviations from relevance following from TraumAID's extensive general knowledge base of conclusions, goals, and actions in the domain.

Several researchers have pointed out the advantages of using contextual knowledge and basic domain principles to bias the search for an explanatory plan (Huff & Lesser 1993; Hill & Johnson 1995; London & Clancey 1982). The basic idea is that the plan recognizer can use its knowledge of what actions are appropriate in the current situation to reduce am-

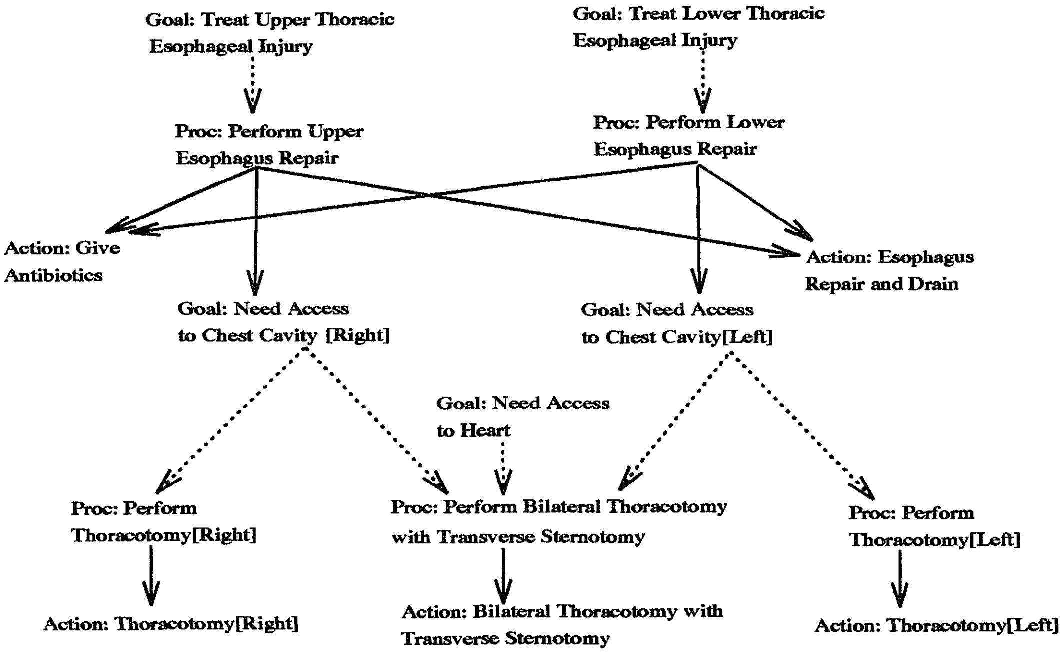

Figure 1: An example plan graph. Dotted arrows indicate disjunctive *goal-procedure mappings*, while solid arrows indicate conjunctive *procedure-action mappings.*

biguities in interpreting observed actions. We believe this is an appropriate bias to use in TraumaTIQ because we can assume:

- The head of the trauma team will have training and experience, and will usually develop plans that are similar to TraumAID's.
- The head of the trauma team is more likely to have appropriate goals but be addressing them in a sub-optimal way, than to be pursuing the wrong goals altogether.
- While TraumAID follows a conservative strategy for pursuing diagnosis and treatment from observations, the head of the trauma team may proceed more rapidly, pursuing a goal for which TraumAID does not yet have enough evidence to conclude its relevance.

The first two assumptions motivate a policy of giving the physician "the benefit of the doubt": if an ordered action can be explained in terms of TraumAID's current goal set, the physician will be assumed to be pursuing the explanatory goal(s). An ordered action can be explained if it appears in TraumAID's plan for addressing a goal in the goal set, *or* if TraumAID has chosen a different action to address this goal.

The third assumption allows the plan recognizer to interpret actions that could be justified by more evidence. Using knowledge about the strategic relationships between goals, TraumaTIQ can identify when the physician's orders may be motivated by a goal that is partially but not yet completely supported by the evidence.

The Plan Recognition algorithm

Informally, our plan recognition algorithm works by first enumerating the set of possible *explanations* for all actions that have been ordered. Each explanation consists of a path in the plan graph from the ordered action to a *procedure* in which the action plays a part, back to a top level *goal.* The path may pass through a series of sub-goals and procedures before reaching a top level goal. Since the same goal may be addressed by more than one procedure, an action may be explained by one goal in the context of two different procedures.

The possible explanations are evaluated in two phases. The first phase considers the top level *goals.* These are sorted according to their *relevance* in the current situation, and the most relevant ones are selected as candidate explanations. The plan recognizer categorizes potential explanatory goals on a 4-level scale:

1. Relevant goals: goals in TraumAID's set of goals to be pursued.
2. Potentially relevant goals: goals that are part of a currently active *diagnostic strategy*, as described ear-

lier. For example, if the goal of diagnosing a fractured rib is currently relevant, then the goal of treating a fractured rib is potentially relevant, depending on the result of the diagnostic test.

3. Previously relevant goals: goals that were once relevant but are no longer so, because either already addressed or ruled out by new evidence.

4. Irrelevant goals: all other goals.

The bias embodied in this phase of plan recognition is that the higher a goal is on this scale, the more likely the physician is considered to be pursuing it.

Formally, phase one of the algorithm can be specified as follows:

1. For each action α ordered, TraumaTIQ's plan recognizer extracts from TraumAID's knowledge base a set of *explanatory procedure-goal chains*, PG_α, that could explain the presence of that action:

$$PG_\alpha = \{\langle P \ldots G \rangle_1, \ldots, \langle P \ldots G \rangle_n\}$$

where P is a procedure containing α in its decomposition, and $\langle P \ldots G \rangle_i$ is a backward path through the plan graph ending with the goal G.

2. Now consider the set $\Gamma = \{G_i\}$ where G_i is the top level goal ending $\langle P \ldots G \rangle_i$. In rank order, Γ consists of: Γ_1 the relevant goals, Γ_2 the potentially relevant goals, Γ_3 the previously relevant goals, and Γ_4 all other goals. Let $\Gamma' = \{G_j\}$ be the highest ranking non-empty subset of Γ. If Γ' is the set of irrelevant goals, halt here and add α to the plan with no explanatory procedure-goal chains.

The second phase considers the procedures in the remaining explanations. These are evaluated according to how strongly the physician's other actions/orders provide additional evidence for them. All procedures in the highest non-empty category are accepted as explanations for the action. For simplicity, the procedures are actually sorted according to a four-level scale of evidence:

1. Completed procedures: procedures for which all the actions have been ordered by the physician.

2. Partially completed procedures: procedures for which some of the actions have been ordered.

3. Relevant procedures: procedures that are currently in TraumAID's plan. This means that if an action could address a goal by playing a role in two different procedures, the one in TraumAID's plan is preferred as the explanation for the physician's action.

4. All other procedures.

Formally, phase two of the algorithm can be specified as follows:

3. Let $\mathcal{P} = \{P_j\}$ where P_j is the procedure that is the child of G_j in PG_α. In rank order, $\mathcal{P}$ consists of: $\mathcal{P}_1$, procedures for which all the actions have been ordered, $\mathcal{P}_2$, procedures for which some actions have been ordered, $\mathcal{P}_3$, procedures currently in TraumAID's plan, and $\mathcal{P}_4$, all other procedures. Let $\mathcal{P}'$ be the highest ranking non-empty subset of $\mathcal{P}$.

4. Select the paths $PG' \subseteq PG$ such that PG' contains all paths ending with goals in Γ' with children in $\mathcal{P}'$.

Finally, the explanations with the most relevant top-level goals and the highest level of evidence (i.e., the paths in PG') are ascribed to the physician and incorporated into TraumaTIQ's model of the physician's plan. Incorporating a new explanation into the plan involves adding new procedures and goals if they are not already present, and adding links between items that are not already connected.

Note that there may be more than one explanation for a given action, if the explanatory goals are equally relevant and the procedures equally manifested. For example, TREAT UPPER THORACIC ESOPHAGEAL INJURY and TREAT LOWER THORACIC ESOPHAGEAL INJURY might be accepted as explanatory goals for the action ESOPHAGUS REPAIR AND DRAIN, provided that both goals are in the same category of relevance.

An example of TraumaTIQ's plan recognition process

The use of context to bias the search for explanatory goals means that TraumaTIQ's plan recognizer can distinguish between goals that are otherwise equally good explanations of the observed actions. Continuing the example from Figure 1, suppose that TREAT UPPER THORACIC ESOPHAGEAL INJURY is currently the only goal in TraumAID's relevant goal set, but the physician is erroneously pursuing the goal of treating an *lower* thoracic esophageal injury. If the physician first orders ANTIBIOTICS, TraumaTIQ will infer that they are being given as part of the procedure to treat the upper esophageal injury, even though antibiotics may play a role in a number of other procedures, including treating a lower thoracic esophageal injury.

Next, if the physician orders a BILATERAL THORACOTOMY in order to get access to the left chest, this action will also be inferred as part of TREAT UPPER THORACIC ESOPHAGEAL INJURY. However, since this is the less preferred procedure for addressing that goal, a comment will be produced to the effect that "Doing a right thoracotomy is preferred over doing a bilateral thoracotomy with a transverse sternotomy to get access to the right chest cavity." Note that such a comment leaves it to the physician to determine that the correct sub-goal is getting access to the right half of the chest in order to treat the upper esophagus.

If, on the other hand, the physician orders a LEFT THORACOTOMY, this action is inconsistent with the goal of treating an upper esophageal injury, and so TraumaTIQ infers that it is being done for some reason that TraumAID does not currently consider relevant. This will result in the comment, "Doing a left thoracotomy is not justified at this time. Please reconsider this order or provide justification." Furthermore, since the physician has failed to order a procedure to get access to the right chest cavity, TraumaTIQ will also produce the comment, "Please consider doing a right thoracotomy and repairing and draining the esophagus in order to treat the upper thoracic esophageal injury." Such a comment would make explicit a possible discrepancy in goals between the physician and TraumAID.

Analysis of the plan recognition algorithm

A serious criticism of previous approaches to plan recognition is that they are computationally intractable (Charniak & Goldman 1993; Goodman & Litman 1992). In a time-critical domain like trauma management, it is essential for TraumaTIQ to respond quickly. The complexity of the algorithm is not really a problem in the current implementation of TraumaTIQ because of the limited size and complexity of the plans generated by TraumAID 2.0. To demonstrate how fast the implementation actually is in practice: TraumaTIQ's plan recognizer, implemented in Lucid Common Lisp and compiled on a Sun 4 processed 584 actions in an average of 0.023 cpu seconds per action.

The problem arises when we consider extending the system to cover other areas of the body and/or blunt injury, increasing the number of procedures and goals that might explain an action in the knowledge base. To allow for the growth of the system, it is important that the plan recognition algorithm scale up efficiently.

As Rymon (1993) points out, plan recognition can be formalized as a set-covering problem in which two sets of observations, symptoms and actions, are mapped onto a set of goals which covers both of them: every symptom motivates some goal and every action is motivated by some goal in the covering set (Figure 2). The covering set is optimized according to some cost function, such as set minimization. Since the set covering problem in general is NP-hard, so is this formalization of plan recognition.

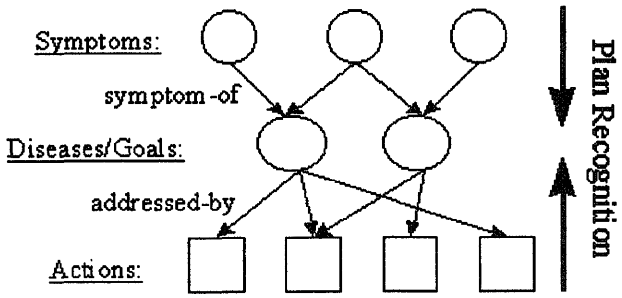

Figure 2: Plan Recognition as a set covering problem

In general, any plan recognition algorithm that considers all possible combinations of explanatory goals for the observed actions is going to grow exponentially with the number of actions. The algorithm we present here avoids the need for an exponential search by grouping the potential explanations according to *relevance* and then greedily accepting all the explanations in the most relevant group. One way to look at this is that rather than trying to optimize the covering goal set according to a cost function, we simply choose to maximize the number of relevant goals in the covering set.

In doing this, for each ordered action α, the algorithm only has to consider $|\Gamma|$ goals, where Γ is the set of possible explanatory goals for α, and $\sum_{|\Gamma'|} |\mathcal{P}_{\Gamma_j}|$ procedures, where Γ' is the most relevant non-empty subset of Γ, and $\mathcal{P}_{\Gamma_j}$ are the procedures linked to each goal Γ_j in Γ'. For each procedure, it has to look at $|\mathcal{A}_{\mathcal{P}}|$ actions in the procedure, and compare them with at most all of the actions that have been ordered. So the total cost of inferring a plan from a set of orders, $\mathcal{A}$, is at most

$$|\mathcal{A}| * (|\Gamma| + (\sum_{|\Gamma'|} |\mathcal{P}_{\Gamma_j}| * |\mathcal{A}_{\mathcal{P}_j}| * |\mathcal{A}|))$$

Thus, this algorithm is polynomial in the number of ordered actions, and linear in the number of possible goals per action, the number of goals in the most relevant goals set, and the number of possible procedures per action.

Evaluation and Discussion

We evaluated the performance of the plan recognition algorithm by applying it to the management plans from the 97 actual trauma cases from the Medical College of Pennsylvania used in the retrospective validation of TraumAID (Clarke *et al.* 1993; Gertner, Webber, & Clarke 1996).

Out of 584 actions, 234 of them were not also part of TraumAID's plan at the time that they were ordered.

Of these 234, 15 of them could be explained by a goal that was currently in TraumAID's relevant goal set. Of the remaining 219, 69 could be explained by a goal that was considered to be potentially relevant, given TraumAID's current knowledge about the state of the patient. The plan recognizer failed to explain the remaining 148 actions in terms of relevant or potentially relevant goals.

Many of the actions that TraumaTIQ fails to infer a goal for are broad diagnostic tests that can be used to look for a number of conditions, and the physician may not actually have a specific goal in mind when ordering them. To understand physicians' plans in such cases it is necessary to have a more complete abstraction hierarchy for goals than is currently available in TraumAID 2.0. Since the knowledge base was implemented in support of plan *generation* rather than plan *recognition*, only goals that could be directly operationalized as actions were included.

Second, some goals that physicians may pursue in these cases are not included in TraumAID's knowledge base because its designers opted not to pursue these goals under any circumstances relevant to the current domain of the system. To have a complete plan recognition system, it will be necessary to include such goals in the knowledge base.

Summary and Conclusion

In this paper we have pointed out the weakness of standard inductive biases, such as goal minimization in domains where agents can have multiple independent goals. We have further argued that the goals and plans that a decision support system would adopt under the circumstances can provide a workable inductive bias. To show this, we have described how TraumAID's planner provides a standard of relevance and of measurable deviations from relevance, providing in turn a context for the incremental recognition of physician plans by TraumaTIQ. The approach to plan recognition presented here is computationally efficient and can be applied in any domain where the user's behavior can be predicted on the basis of contextual information.

Acknowledgments

This work has been supported in part by the National Library of Medicine under grants R01 LM05217-03 and R01 LM05764-01 and the Agency for Health Care Policy and Research under grant RO1 HS06740.

References

Charniak, E., and Goldman, R. P. 1993. A Bayesian model of plan recognition. *Artificial Intelligence* 64:53–79.

Clarke, J. R.; Rymon, R.; Webber, B.; Hayward, C.; Santora, T.; Wagner, D.; and Ruffin, A. 1993. The importance of planning in the provision of medical care. *Medical Decision Making* 13(4):383. abstract.

Gertner, A.; Webber, B.; and Clarke, J. 1996. On-line quality assurance in the initial definitive management of multiple trauma. *submitted to Artificial Intelligence in Medicine.*

Gertner, A. S. 1995. *Critiquing: Effective Decision Support in Time-Critical Domains.* Ph.D. Dissertation, University of Pennsylvania, Philadelphia, Pennsylvania.

Goodman, B. A., and Litman, D. J. 1992. On the interaction between plan recognition and intelligent interfaces. *User Modeling and User-Adapted Interaction* 2(1-2):55–82.

Hill, R. W., and Johnson, W. L. 1995. Situated plan attribution. *Journal of Artificial Intelligence in Education.* to appear.

Huff, K. E., and Lesser, V. R. 1993. Integrating plausible reasoning in an incremental plan recognizer. Technical Report 93-72, University of Massachusetts, Amherst.

Kautz, H. 1990. A circumscriptive theory of plan recognition. In Philip R. Cohen, J. M., and Pollack, M. E., eds., *Intentions in Communication.* Bradford Books.

London, R., and Clancey, W. J. 1982. Plan recognition strategies in student modelling: prediction and description. In *Proceedings of the American Association for Artificial Intelligence*, 335–338.

Rymon, R. 1993. *Diagnostic Reasoning and Planning in Exploratory-Corrective Domains.* Ph.D. Dissertation, Department of Computer & Information Science, University of Pennsylvania. Appears as Technical Report MS-CIS-93-84.

Shahar, Y., and Musen, M. A. 1995. Plan recognition and revision in support of guideline-based care. In *Working notes of the AAAI Spring Symposium on Representing Mental States and Mechanisms*, 118–126.

Webber, B. L.; Rymon, R.; and Clarke, J. R. 1992. Flexible support for trauma management through goal-directed reasoning and planning. *Artificial Intelligence in Medicine* 4(2):145–163.