TAILORING EXPLANATIONS FOR THE USER

Kathleen R. McKeown
Myron Wish
Kevin Matthews

CUCS-172-85
Tailoring Explanations for the User

Kathleen R. McKeown
Dept. of Computer Science
Columbia University
New York NY 10027
212-280-8194
McKEOWN@COLUMBIA-20

Myron Wish
AT&T Bell Laboratories
600 Mountain Ave
Murray Hill, N.J. 07974
201-582-7630

Kevin Matthews
Dept. of Computer Science
Columbia University
New York, NY 10027
212-280-8190
MATT@COLUMBIA-20

Abstract

In order for an expert system to provide the most effective explanations, it should be able to tailor its response to the concerns of the user. One way in which explanations may be tailored is by point of view. A method is presented for representing the knowledge to support different points of view in the current domain. In addition, a method is presented for determining the point of view to take by inferring the user's goals within a brief discourse segment. The advising system's response to the derived goal depends on the strength of its belief in the inference, for which a method of determination is also provided. This information enables the system to decide what answer to give to a question, which kind of justification is relevant, and when to provide it. Some details of the current implementation are included.

1 Introduction

While research on explanation for expert systems has addressed some important issues in identifying the kind of knowledge needed to provide acceptable explanations (e.g., Swartout '81, Clancey '79), one main problem with existing systems is their inability to tailor an explanation adequately to the needs or perspective of a particular user. In this paper, we show how information about the current user can and should influence the type of explanation provided.

In past artificial intelligence research, there have been two main approaches to user modeling: classifying users according to a priori types often by direct interrogation (e.g., Rich '79, Swartout '81, Wallis '82) or deriving information about the current state of the user's goals, beliefs, and desires from the ongoing discourse itself (e.g., Allen and Pernoll '80, Carberry '83). Our work draws from the second of these two main approaches. Recent research (e.g., Allen and Pernoll '80) has emphasized the derivation of a user's goal in order to interpret an utterance correctly. We are interested in making use of derived goals to generate appropriate explanations. This difference in emphasis has required the development of techniques for handling four specific tasks: representing different points of view in a knowledge base; support different explanations; identifying which of several possible goals underlying the current discourse should be addressed, determining when the derived goal should be taken into account, and specifying how a generation system can relate the derived goal to different points of view to determine explanation content. This extends Allen and Pernoll's (80) approach by showing how a goal can be derived to represent a sequence of utterances as opposed to a single utterance, and goes beyond Carberry's (83) approach by showing how a system can decide to respond to such goals.

This work is being done within the context of an ongoing project to develop a dialogue facility for a computer-aided problem solving. A student advising system is being developed which can provide information about courses and advice about whether a student can or should take a course, in particular course. The system is currently structured as a question-answering system which involves an underlying expert system on receiving "can" questions (e.g., "Can I take natural language this semester?") and "should" question (e.g., "Should I take data structures?"). This production system uses its rule base to determine the advice provided (i.e., yes or no) and the trace of rule invocations is used to provide a supporting explanation of the advice.

The Advisor system consists of an ATN parser (Woods '70), a K-LONE knowledge base (Brachman '79) with access functions to a student's underlying production system, and a surface generator to produce responses and explanations in natural language (Derr and McKeown '84). Currently the system can produce responses to information questions by accessing the knowledge base and to "can" questions by invoking the underlying production system. Certain aspects of response generation and inferring for "should" questions have been implemented.

In the following sections, we first show the different types of explanations required and then describe in some detail the techniques we have developed.

2 Different Explanations

In this paper, we focus on how the content of an explanation must vary according to the perspective or point of view taken on the underlying problem domain. For example, in the student advising domain, there are a number of points of view the student can adopt for selecting courses. It can be viewed among others, the process of meeting requirements (i.e., how do courses fit into schedule slots?) as a state model process (i.e., what should be completed at each state in the process?), as a semester scheduling process (i.e., how can courses fit into schedule slots?), or as a process of maximizing personal interest (as in "how will courses help me learn more about AI?"). Given these different points of view, alternative explanations of the same piece of advice (i.e., yes) can be generated in the response to the question...

1 Requirements: Yes, data structures is a requirement for all later Computer Science courses and discrete math is a corequisite for data structures.
2 State Models: Yes, you usually take them both first semester sophomore year.
3 Semester Scheduling: Yes, they're offered next semester, but not in the spring.

1The work described in this paper was partially supported by ONR grant N00014-82-K-0255 and by AT&T Bell Laboratories.
and you need to get them out of the way as soon as possible.

4 Personal Interests (e.g., AI) Yes if you take data structures this semester you can take Introduction to AI next semester, and you must take discrete math at the same time as data structures.

One of these explanations may be more appropriate than others depending upon the user's goal in pursuing the dialogue. For example, we might supply explanation (1) above if the user's goal were to complete requirements as soon as possible and explanation (2) if the user's goals were to keep pace with the normal rate of progress. Thus to address the problem of selecting a perspective to use in an explanation, we must develop techniques that allow a system to infer a user goal from a discourse segment as well as techniques that can indicate information that is relevant for any given perspective.

3 Knowledge Representation

In order to identify information that is relevant to a user's goal, we are using intersecting multiple hierarchies to represent different points of view in the underlying knowledge base. The hierarchies are cross-linked by entities or processes (often courses in the student advising domain) which can be viewed from different perspectives (and thus occur in more than one hierarchy). Hence to construct the context for explanation (1) above, the system would extract information about the relation between data structures and discrete math from the requirements hierarchy, and for explanation (2) extracts information from the state model hierarchy. A diagram of a portion of these two hierarchies containing information for the two points of view is shown in Figure 1 below.

![Figure 1: Representing Points of View](image)

The partitioning of the knowledge base by intersecting hierarchies allows the generation system to distinguish between different types of information that support the same fact. From this partitioning, the system can select the portion that contains the information relevant to the current request and user goal.

4 Deriving the User Goal

The system must also be able to reason about the appropriateness of one perspective versus others. Since the perspective taken is related to the user's goal in pursuing the dialogue, the large body of work on goal inference techniques (Allen and Perrault 80), Litman and Allen 84), is applicable for deriving the user's goal. We have drawn heavily from Allen and Perrault's (80) work, making use of their plausible inference rules, representation of domain plans, and representation of speech acts as plans. While their work has been extremely useful it falls short for our purposes in several ways. For example, their inferring procedure derives a plausible goal for a user based on a single utterance, while we are interested in deriving a goal based on the current sequence of utterances.

Consider the discourse shown in (6) below. Assuming that a data of domain plans common to the student advising domain is maintained, Allen and Perrault's techniques could be used to derive the domain goal shown following each question. But the explanation shown in (6c) addresses not the derived goal of (6c), nor any of the derived goals of the previous utterances; instead addresses the higher level goal indicated by the derived goals of (6a) and (6b). The problem for responding to such goals in an explanation then is to be able to derive a higher level goal relating the goals of individual utterances.

6a S: I've read about the field of AI and I'm interested in learning more about it eventually. Is natural language offered next semester?

Plausible goal = take natural language

A: Yes

b S: Who is teaching artificial intelligence?

Plausible goal = take AI

A: Lebowitz this semester

c S: I haven't taken data structures yet. Should I take it this semester?

Plausible goal = take data structures

A: Yes if you take data structures this semester you can take AI next semester which is necessary for all later AI courses

We use Allen and Perrault's rules to derive the domain goal of each individual utterance which we term the current goal. We also identify a goal representing the discourse sequence which we term the

---

2 In this work we restrict ourselves to a discourse segment that deals with a single or related set of goals. Over a longer sequence of discourse, topics may shift and the user may express very different goals across such boundaries. Detecting topic shifts and natural changes in goals is a difficult problem that we are not addressing.
relevant goal since it will be used to generate later explanations. Intuitively the relevant goal is a higher level goal if there is one, relating the goals of several utterances.

The process of determining the relevant goal involves the following steps: The current goal is first derived from the initial utterance. All higher level domain goals are then derived from the current goal using Allen and Perrault’s body-action inference rule (i.e. if the user wants a step in the body of a plan to hold, it is plausible that s/he wants the action to hold). Any one of these is a candidate for the relevant plan. A derivation of the higher level plans for the utterance “Is natural language offered next semester?” is shown in Figure 2. Note that the action take natural language is a step in two separate plans: concentrate-on-AI and fulfill electives, and thus two parent paths are formed.

1. “Is natural language offered next semester?”
current goal = take natural language
2. “Who is teaching artificial intelligence?”
current goal = take AI
relevant goal = concentrate-on-AI

Figure 3: Relevant Goal for Utterances 1 and 2

When the second utterance “Who is teaching artificial intelligence?” is entered, the current goal take AI is derived and all higher level goals derived (see Figure 3) from that using the body-action rule. The lowest level node where the two paths intersect becomes the relevant plan: concentrate-on-AI in this case. If the second utterance had been “When is operating systems offered?” the higher level goal fulfill electives would have been inferred since this is the only relation between the goals take operating systems and take natural language.

This method is essentially a search for the lowest common ancestor of the current goals of two consecutive utterances. When the third or any subsequent utterances are encountered the relevant goal is determined by performing the search for common ancestor using the previous relevant goal and the current goal of the new utterance.

Carberry (83) does present a method for tracking user goals over a sequence of discourse, building in the process a hierarchical model of user plans for the discourse. She uses this hierarchy and a set of focus heuristics to determine for the next incoming utterance which of several plausible plans the user could be focusing on. She does not specify which plan in the hierarchy best represents the overall discourse purpose and therefore should be addressed in succeeding explanations. Our model thus augments hers by providing this information.

§ When to Respond

The goal inference techniques just described allow the system to infer what a user's goal might be but this inference may be so tentative that explanations which always address such goals will be undesirable as those that never take a goal into account. Allen and Perrault themselves term their rules plausible inference rules since the goals they attribute to the user are only possibilities and not definite. However, goals derived from some discourse sequences seem intuitively more definite than those derived from other sequences.

If the user directly asserts his/her goal (as in 7)
then it can be definitely inferred. The plausible inference rules, however, will infer the same goal for an utterance like that shown in (8). Unless we have further indication that the user actually has the goal take natural language, then on receiving a follow-up question such as (9a), a neutral explanation as shown in (9b) is preferable to the tailored explanation in (9c). One problem for a system that generates tailored explanations is that it is being able to determine when to respond to a derived goal.

7 S I'm planning on taking NLP in the future. What are the prerequisites?
   plausible goal = take natural language

8 S Is natural language offered this semester?
   Plausible goal = take natural language

9a S I'm thinking of taking computability this semester. Would that be a good idea?
   plausible goal = take computability
b A Yes, it's your last requirement and it's a good idea to get it out of the way before going on to electives
c A Yes, computability is particularly important for NLP since it covers grammars so it's a good idea to take it first

To handle this problem we use three levels of likelihood of derived user goals. If we can distinguish between derived user goals that can definitely be attributed to the user, derived goals that are likely, and derived goals that are only plausible, then we have a basis for determining when to generate tailored explanations. Tailored explanations can be generated for definite and likely goals and a neutral explanation generated for plausible goals.

A goal is definite if a user states that s/he has that goal, as in "I want to concentrate in AI" or "I'd like to concentrate in AI," or "I'm interested in taking as much AI as possible." If not stated, it is difficult to infer without doubt that a user has a given goal, but there are cases where it is more likely than others. Space prohibits providing details but we note that a goal is more likely if it has been repeatedly derived from consecutive utterances as well as in cases where it is one step in a plan that the user has partially completed. The system is currently capable of deriving current and relevant goals for a discourse segment and classifying them as plausible or definite. Classification of goals as likely has been designed, but must still be implemented.

We have ignored, in this paper, the possibility of responding in other ways than providing explanations in some cases, in fact, it may be preferable for the system to ask the user to clarify his/her goal or to take the initiative in some other way. Determining when and how to take the initiative as an alternative to providing explanations is a topic addressed elsewhere (see Matthews 82).

8 How to Respond

Finally the system must be able to make use of the derived goal in constructing an explanation when a 'should' question follows a dialogue sequence. The underlying mini production system consists of working memory, rule base, and inference engine is invoked in this process.

To construct the explanation, the hierarchy representing the possible perspective is determined directly from the relevant goal, and information retrieved about the questioned object from that hierarchy is placed in working memory. The production system uses this information to derive the response, that is whether the user should or should not pursue the queried action. The trace of the reasoning is then available to provide the basis for the explanation as in the case in traditional expert systems. Note that the information extracted from one hierarchy will allow a different set of rules to fire than will information extracted from another, thus producing different explanation content.

As an example, consider again the question "Should I take both data structures and discrete this semester?" Assume that the system has determined that the user's goal is take required and that the goal should be taken into account in the explanation. After deducting that the student can take these courses, the production system will attempt to prove that the queried action helps the user achieve his/her goals. The information shown in Figure 4, extracted from the requirements hierarchy (refer back to Figure 1), enables rules 1 and 2 to fire with the instantiated fact as data structures by as discrete math and course as required. The extracted fact that discrete math is a co-requisite for data structures enables rule 2 to fire. Its consequence and the extracted fact that data structures is a prerequisite to required enables rule 1 to fire, which concludes that required can be taken. Thus, the advice is yes since take required is the user's goal and these two instantiated rules can then be used as the basis for the hypothetical explanation given earlier and reproduced in Figure 4. Other rules in the rule base (such as A course should be taken if the student is at the right year to take it) do not fire since information necessary to fire that rule does not exist in working memory. This processing is partially implemented, but much work is needed before the full explanation can be produced.

3 The questioned object is the course the user is inquiring about (e.g., data structures in "Should I take data structures?").

Regardless of whether the user's queried action helps him/her achieve the relevant goal, if it is not permissible or will prevent the student from completing the major, the advice is always negative. Recognizing such an absolute condition is called a barrier. Here, we assume the student has already taken the prerequisites to data structures and discrete math and is ready enough in his/her program that s/he will be able to finish on time, and thus the absolute condition is satisfied.
7 Future Directions

More research is needed on explanation, plan recognition, and user modelling for our approach to be effective for a broad range of human-computer dialogue. As for explanation, in the current implementation, the production system needs to be developed further and its reasoning trace interfaced to the operational surface generator for English output. On the theoretical side, we are investigating the use of discourse strategies to control the organization of the explanation. The plans in the current implementation were selected by examining transcripts of actual student-consultant sessions, but it would be desirable to have a much larger set of plans, knowledge about their base rates and importance and additional criteria for tracking their relevance and likelihood during the interaction. It seems likely, also, that better explanations will require a more complete user model incorporating static, global characteristics of the user as well as those dynamic, local characteristics available from the ongoing dialogue itself. Additionally, while we have touched on one way of representing and using point of view, others will doubtless be necessary. Such a comprehensive attack on the topics of explanation, plan recognition, and user modelling offers promise from both a theoretical and practical perspective.

8 Conclusion

We have demonstrated the need for tailoring explanations to users in consultative or problem-solving dialogues with a computer, and have addressed this problem with a new approach integrating research in plan recognition, user modelling, and explanation generation. Derivation of goals or plans is based on an extension of Perrault and Allen's [80] work which handles discourse segments rather than isolated utterances. Our model and implementation provide mechanisms for assessing which goal is relevant to the user at any moment during the discourse as well as when that point of view should be addressed in an explanation. It also makes progress toward the determination of how to tailor the explanation to the user's goal. In addition to enhancing previous work on goal inference, this report shows how research in natural language processing on goal derivation can be applied to generate explanations sensitive to the user's current perspective in expert system interactions.

Acknowledgments

We would like to thank Michael Letowski for suggestions on an earlier draft of this paper.

References


(Derr and McKeown 84) Derr, M A and K R McKeown, Using focus to generate complex and simple sentences, Proceedings of COLING-84, Tenth International Conference on Computational Linguistics, Stanford, July 1984, pp 319-26


(Mathews 83) Mathews, K. Initiation and reactive system roles in human-computer discourse, unpublished manuscript, AT&T Bell Laboratories, 1985


