Practical Issues in Mixed-Initiative Natural Language Dialog: An Experimental Perspective

Ronnie W. Smith
Department of Mathematics
East Carolina University
Greenville, NC 27858, USA
email: rws@cs.ecu.edu

Abstract

There are several theoretical and technical challenges to be resolved in developing effective computational models for mixed-initiative interaction. This paper describes some of the unresolved challenges encountered in the development of a computational model of dialog processing. This model was implemented in the Circuit Fix-It Shop, a spoken natural language dialog system that was experimentally evaluated with eight different users in 141 problem-solving dialogs. From our experiences in developing a system capable of experimental evaluation we have noted at least three difficulties in modeling mixed-initiative interaction: (1) maintaining coherence during an initiative change; (2) choosing an appropriate linguistic form for a response as a function of initiative; and (3) processing silence as a legitimate input and output. Our experiences in dealing with these issues are the focus of discussion in this paper.

Introduction

There are several theoretical and technical challenges to be resolved in developing effective computational models for mixed-initiative interaction. This paper describes some of the unresolved challenges encountered in the development of a computational model of dialog processing presented in (Smith & Hipp 1994). This model was implemented in the Circuit Fix-It Shop, a spoken natural language dialog system that assists users in repair of an electronic circuit. The original intent had been to make the system capable of mixed-initiative interaction and to test this capability experimentally. However, for reasons outlined later, the experimental evaluation focused on the *variable* initiative capabilities of the system. Thus, while the system and underlying computational model permit the system to interact with users at varying levels of initiative, during the experimental evaluation the initiative level was fixed for the duration of an interaction. After a brief review of our model for handling initiative, the paper examines three problems encountered during the development and evaluation of the Circuit Fix-It Shop: (1) maintaining coherence during an initiative change; (2) choosing an appropriate linguistic form for a response as a function of initiative; and (3) processing silence as a legitimate input and output.

Perspective on Initiative

For mixed-initiative collaborative problem-solving, we take the view that initiative lies with the participant whose goals currently have priority. Thus, while computer initiative is characterized by giving priority to the specific goals of the computer, user initiative is characterized by giving priority to the user's goals. If natural language dialog is the medium of interaction, then dialog must be used to indicate who has current control of the task. We characterize the relationship between dialog initiative and task initiative as follows: *the level of initiative in the dialog should mirror the level of initiative in the task.* Consequently, the level of initiative that the computer has in the dialog is primarily reflected in the degree to which the computer allows the user to interrupt the current subdialog in order to discuss a topic more relevant to the user's immediate goals.

Based on this view and the view that initiative is not an all or nothing prospect, four dialog modes were identified that characterize the level of initiative that the computer can have. These are described below.

1. *Directive*—The computer has complete dialog control. It recommends a task goal for completion and will use whatever dialog is necessary to complete this goal. No interruptions to other subdialogs are allowed.

2. *Suggestive*—The computer still has dialog control, but not as strongly. The computer will recommend a task goal for completion, but will allow minor interruptions to closely related subdialogs.

3. *Declarative*—The user has dialog control and can interrupt to any desired subdialog at any time. How-

ever, the computer is free to mention relevant facts as a response to the user's statements.

4. *Passive*—The user has complete dialog control. Consequently, the computer will passively acknowledge user statements. It will provide information only as a direct response to a user question.

These four modes were implemented in the Circuit Fix-It Shop. To provide for the experimental control necessary to demonstrate that there were fundamental differences in the interaction between computer and user as a function of initiative, the dialog mode was fixed throughout a given experimental session with a subject. The system was tested using directive mode and declarative mode during the experiment, and we did observe some fundamental differences in the interaction (see (Smith & Gordon 1997)). However, during development of the model for mixed-initiative dialog several problems were encountered. The remainder of this paper discusses some of these problems as they manifested themselves during the evaluation of the Circuit Fix-It Shop.

Coherent Transition during Initiative Changes

In the model of initiative just described, initiative change is characterized by switching the system to a new dialog mode. Given that mode-switching occurs, a key issue is how to ensure the coherence of the dialog when mode changes. For mode changes where the same conversational participant retains the initiative (i.e. between suggestive and directive or between declarative and passive), maintenance of coherence should be easy because the same participant's goals have priority. But for more extreme mode changes where the other participant gains control, there is the risk that coherency may be lost. For example, the following dialog segment occurred in an early version of the Circuit Fix-It Shop. Initially, the mode was declarative, with the user in control.

```
C: The LED is supposed to be displaying
     alternately flashing 1 and 7.
U: The LED is displaying nothing.
C: The power is on when the switch is up.
U: The switch is in the up position.
C: The switch is connecting to the battery
     when there is a wire between
     connector 111 and connector 120.

<User observes there's no wire, but says
     nothing about this>

U: Now what should I do?

<mode switch to suggestive>
```

Based on the theory outlined in the previous section, the change in initiative to a computer-controlled mode should cause the computer to direct the dialog toward a discussion of its goals. Unfortunately, the system produced the following response:

```
C: Turn the knob to 10.
```

a very confusing response with respect to the previous utterances.[1] In selecting a response topic, the computer ignored the context previously established and simply used the topic recommended by the domain reasoning subsystem. The problem is in the failure of the computer when it regained the initiative to adequately consider the established context in choosing the next topic of the dialog. A smooth transition when changing dialog mode is necessary in mixed-initiative dialog.

Preliminary efforts at correcting this problem focused on algorithms for determining the relationship between the computer's goal and the previous user goal, but no reasonable solution was found for the situation when there was no close relationship between the different goals. Negotiation subdialogs (Lambert & Carberry 1992) are a possible method of dealing with this problem. In general, we noted this was only a problem for mode changes where control moved from the user to the computer. For mode changes in the opposite direction, the user could initiate the topic change, and the computer would automatically shift the focus of its responses to the user goal. However, given the lack of an automatic mode switching algorithm, further work in developing a general linguistic-based model for maintaining coherence has not yet been completed. In general, there needs to be some form of metadialog processing that can effect a smooth transition between the current dialog topic and the new topic that the newly controlling dialog participant wishes to discuss.

Choice of Linguistic Forms

While the current dialog mode influences the selection of the topic to be discussed, another variable in the response formulation process is the level of assertiveness to be used in producing a response. For example, the dialog topic, "set the switch in the up position" can be introduced in a variety of ways as shown in figure 1. The topic of conversation for any of the first set of utterances may be getting the switch turned up,

[1] The reason for this topic selection is that the computer is interested in knowing what the LED is displaying when simultaneously the switch is up *and the knob is at position 10.*

```
Turn the switch up.                      (command)
Would you [please] turn the switch up? (request)
Can you turn the switch up?              (indirect request)
The switch can be turned up.             (statement of fact)
Turning the switch up is necessary.     (statement of fact)


What is the switch position?             (request)
I need to know the switch position.     (indirect request)
```

Figure 1: Sample introductions to a dialog topic

but the directness of the request varies greatly. Similarly, in the last two utterances, the topic of conversation could be learning what the switch position is, although the form of the utterance differs in each. If the dialog is to reflect the underlying task initiative, the more assertive utterances would be appropriate when the computer has the initiative, while the less assertive would be appropriate when the user has the initiative.

For the Circuit Fix-It Shop, assertiveness as a function of mode was fixed. In computer-controlled dialogs most computer responses were phrased as requests or commands while in user-controlled dialogs the computer utterances were primarily statements of fact and indirect requests. While this seems sufficient for the circuit repair domain, it may not be adequate in other domains.

Consider the dialog excerpt of figure 2. This excerpt from the murder mystery domain is a modification of an annotated computer-computer interaction presented in (Guinn 1996) where the actual messages passed between participants are logical forms rather than natural language utterances. The initial question has Holmes yielding control to Watson where Watson's task goal is to prove Suspect10 is the murderer. The initiative remains with Watson until utterance 12, where a conflict begins about who should have the initiative. In this revised excerpt, both speakers use requests and assertions when initiative is still with Watson and also when initiative is being negotiated. It is an open question as to how to incorporate choice about the surface form of the utterance for a fixed level of initiative within a mixed-initiative framework.

Silence Processing

In a collaborative interaction, silence may sometimes simply be a sign that a participant is working quietly while at other times it may indicate a collaborator is having difficulty completing a particular task. How the computer handles silence demonstrates its level of patience and will be a function of its current dialog mode. Silence can be considered a user input just like a verbal utterance. Intuitively, the level of patience demonstrated will be in inverse proportion to the current level of computer control in the dialog. Consequently, the system will be most patient in passive mode, least patient in directive mode, and the level of patience will gradually increase as mode moves from directive to suggestive to declarative to passive. Patience can be computationally characterized in two ways: (1) the length of time that the system will accept silence input before acting; and (2) the type of action taken once the computer decides that silence has gone on too long.

Determining the correct amount of time the system should wait for a response from the user should be a function of the current dialog mode as well as the current task step. For instance, measuring a voltage normally takes more time than identifying an object. Consequently, the computer should be willing to wait longer for a response after requesting a voltage measurement. The use of such information requires that such knowledge be available from the domain reasoning subsystem. The *time out parameter* will be the specification of the maximum amount of time the system should wait for a response to its most recent utterance.

If the elapsed time after the most recent output exceeds the allowed time out parameter, a *silence interrupt* is generated. We propose the following rules for generating a response after a silence interrupt. The rules are a function of the current dialog mode and the number of consecutive silence interrupts generated without an intervening user response.

1. Current Mode = directive or suggestive — The 1st interrupt invokes a query of, "Did you hear me?". The 2nd interrupt causes the computer to initiate attempts at getting the user to perform each of the substeps for the current task step. For example, if the current task step is to set the switch to some position, the computer may now respond by asking if the user sees the switch. Thus, the only differ-

```
 1: Holmes  Who is the murderer of Lord Dunsmore?
 2: Watson  Is it the case that Suspect10 is the murderer of Lord Dunsmore?
 3: Holmes  Do you know if Suspect10 had a motive to murder Lord Dunsmore?
 4: Watson  Yes.
 5: Holmes  Is it the case that Suspect10 had access to poison?
 6: Watson  Yes.
 7: Holmes  Is it the case that Suspect10 had an opportunity to administer the poison?
 8: Watson  I do not know.
 9: Holmes  I am unable to determine whether Suspect10 is the murderer of Lord Dunsmore.
10: Watson  I do not know either.
11: Watson  Is it the case that Suspect9 is the murderer of Lord Dunsmore?
12: Holmes  What about Suspect7?  Is Suspect7 the murderer of Lord Dunsmore?
13: Watson  I have proven that Suspect9 has a motive to murder Lord Dunsmore and Suspect9 had access to
            the poison.
14: Holmes: I have proven that Suspect7 had access to the poison, Suspect 7 had an opportunity to
            administer the poison, and Suspect7 has a criminal disposition.
15: Watson  Suspect 7 is the murderer of Lord Dunsmore.
```

Figure 2: Sample Annotated Computer-Computer Mixed-Initiative Dialog

ence between directive and suggestive mode, the two modes indicating computer control, is the value of the time out parameter (greater in suggestive mode).

2. Current Mode = declarative — As above, the 1st interrupt invokes a query of "Did you hear me?". However, the 2nd interrupt invokes a query of "Can I help you?", because in declarative mode the user is considered to be in control. Thus, the system exhibits more patience. Finally, the 3rd interrupt will cause the system to repeat the information previously stated before the original silence interrupt. The system demonstrates much greater restraint in its responses in this mode than in the modes indicating computer control.

3. Current Mode = passive — This mode will have the longest delay before a silence interrupt occurs. Once it does occur, the system will transition to declarative mode, and subsequently follow the steps presented above.

These rules reflect the idea about dialog mode that the level of patience should be inversely proportional to the level of initiative the computer has. For passive mode, the longest time interval is required for a silence interrupt to trigger. In declarative mode, the computer will prompt the user, but only about the current action in the dialog without pushing the user by giving more details. In the computer aggressive modes, directive and suggestive, the computer will try to push the user by mentioning substeps about the current action and will not wait as long for silence interrupts as in declarative and passive modes.

A consequence of this algorithm is that if silence is a valid input, then silence can also be a valid output as in the two situations described below.

1. After a silence interrupt when the computer has invoked the goal of "Did you hear me?", and the user has responded with "Yes." This indicates there is no need to repeat the previous computer utterance. A segment of dialog illustrating this is as follows:

```
C: Is anything else on the LED on?
   <silence interrupt>
C: Did you hear me?
U: Yes.
   <No response from the computer>
U: The LED is off.
```

2. After a silence interrupt when the computer has invoked the goal of "Can I help you?", and the user has responded with "No." This indicates the computer should continue patiently waiting as shown in the following dialog situation:

```
C: What is the voltage between connector
      121 and connector 120?
   <silence interrupt>
C: Can I help you?
U: No.
   <No response from the computer>
U: 9 volts.
```

There remain some unexplored possibilities where the computer should also not respond. These possibilities involve consideration about dialog mode and the level of initiative that the computer has. For example, when the user has more initiative, and has just acknowledged comprehension of some piece of factual information provided by the computer, it may be more in keeping with the concept of reduced initiative for the computer to wait patiently for further input from the user before providing another piece of factual information. This currently remains unexplored.

A crucial technical problem with silence processing is that it violates the turn-taking protocol. That is, sometimes the computer can give multiple responses without an intervening user response. In testing our system, this caused users to believe they could give multiple responses at any time, which was technically unfeasible. It was concluded that the implementation of the silence processing algorithm caused more confusion than assistance because the system did not permit arbitrary violations of the turn-taking protocol, and so silence processing was disabled for the formal experiment.

Issues in Evaluation

The most extensive study of the effectiveness of mechanisms for mixed-initiative dialog has been conducted by Guinn (see (Guinn 1994) and (Guinn 1996)). To achieve this, he used computer-computer simulations to obtain the large number of data points needed to make quantitative conclusions about the effectiveness of strategies for selecting the participant who should have the initiative at the current point in the dialog. Evaluating the effectiveness of the same strategies in actual human-computer interaction is a significantly more daunting challenge. From our experiences in evaluating the Circuit Fix-It Shop we can summarize some of the problems as follows.

- The time required to obtain an equivalent number of data points is prohibitive.
- In computer-computer interactions the participants can exchange messages using a logical form of representation while human-computer dialog requires construction of a fully functional natural language interface.
- Learning effects are difficult to control. If a longitudinal study is attempted, two different subjects with the same amount of experience with the system will possess different amounts of knowledge about how to successfully complete the task and interact with the system.
- Human subjects are likely to bring pre-existing biases about system capabilities that will influence their behavior when interacting with the system.

The traditional way to deal with variability in human subjects is to use such a large number of subjects that these variabilities cancel out over the large number of data points. This seems to be a problematic solution for the near future for evaluating human-computer dialog systems. Development of alternative strategies for measuring the effectiveness of computational models of mixed-initiative must coincide with the development of the models themselves.

Acknowledgments

The author wishes to express his appreciation to Steven A. Gordon who reviewed a draft of this paper. Other researchers who contributed to the development of the Circuit Fix-It Shop include Alan W. Biermann, Robert D. Rodman, Ruth S. Day, D. Richard Hipp, Dania Egedi, and Robin Gambill. The writing of this paper has been supported by National Science Foundation Grant NSF-IRI-9501571.

References

Guinn, C. 1994. *Meta-Dialogue Behaviors: Improving the Efficiency of Human–Machine Dialogue — A Computational Model of Variable Initiative and Negotiation in Collaborative Problem-Solving.* Ph.D. Dissertation, Duke University.

Guinn, C. I. 1996. Mechanisms for mixed-initiative human-computer collaborative discourse. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, 278–285.

Lambert, L., and Carberry, S. 1992. Modeling negotiation subdialogues. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, 193–200.

Smith, R., and Gordon, S. 1997. Effects of variable initiative on linguistic behavior in human-computer spoken natural language dialog. To appear in the March 1997 Issue of *Computational Linguistics.*

Smith, R., and Hipp, D. 1994. *Spoken Natural Language Dialog Systems: A Practical Approach.* New York: Oxford University Press.