The Role of Intention in Maintaining Coherent Human-Computer Dialog: Two Case Studies

Ronnie W. Smith
Department of Computer Science
East Carolina University
Greenville, NC 27858, USA
rws@cs.ecu.edu

Overview

One of the major challenges for natural language dialog systems has always been the need to respond coherently. This paper examines the relationship between the modeling of intention and the maintenance of dialog coherency for two implemented systems. The main lesson is that developing a proper computational model for intention is absolutely crucial in being able to engage in a flexible and coherent dialog. Furthermore, at some level this model must take into account the expectations of the human conversational partner.

Case Study 1: The Circuit Fix-It Shop

The Circuit Fix-It Shop (Smith, Hipp, & Biermann 1995) was developed in order to validate a theory for the design of a dialog system that would integrate a number of desired behaviors that enable intelligent completion of and transition between subdialogs. The behaviors include task-related problem solving, exploitation of user model and context dependent expectations, and varying dialog initiative.

The purpose of the Circuit Fix-It Shop is to assist a human user in the repair of an electronic circuit that uses a power switch and multiple transistor circuits in order to cause a Light Emitting Diode (LED) to alternately display a one and seven. Circuit failures are caused by one or more missing wires. The human user interacts with the system via spoken natural language dialog in order to receive assistance in detecting the source of the circuit failure, making the repair, and confirming correct behavior. Aspects of the interaction that require sophisticated dialog management include the following.

1. The possibility of user initiative that leads to helpful statements not necessarily directly relevant to the computer's request.
2. The need to engage in subdialogs for providing assistance with subtasks where some steps may already be completed or else knowledge of certain steps may have been communicated or inferred from other communication.

As the project progressed, a difficult to answer question kept recurring, "Why should the person and computer communicate?" A crucial problem was to provide a general, but computationally effective solution to the problem of "Why should the computer speak?" Many largely theoretical discussions of communicative intention are presented in (Cohen, Morgan, & Pollack 1990). Our answer to that question is the Missing Axiom Theory for language use. By considering the system's task goals (e.g. setting the position of the switch, reporting the content of the LED display, determining the presence or absence of a wire, etc.), as theorems to be proven, then language may be used as a means for inquiring about axioms that are missing from the theorem.

Consider the proof tree of figure 1. a is proven by proving b, c, and d; and b is proven by proving $b1$ and $b2$. Suppose that the process of proving a yields subgoal b which leads to subgoals $b1$ and $b2$. The system resorts to natural language and interacts with the user to try to verify the subgoals as shown below (C denotes the computer and U the user).

```
C: The knob is at the top of the green
   region.
U: I see it.
C: What is its setting?
U: Zero.
```

Completion of this dialog enables the proof of b and possibly, at a later time a.

This example also illustrates how user modeling is done. If an earlier interaction involves successfully referencing the knob, then $b1$ will be included as an achieved fact in the user model. The proof of b will trivially achieve $b1$ and omit the first comment about locating the knob.

Furthermore, a close correlation between subdialogs and domain actions is obtained. A subdialog consists of all the language interaction pertaining to a domain action. Movement between and within subdialogs is based on determining the domain action to which a user input is relevant.

Consequently, this theory not only provides a computational framework that connects relevant task goals for the human participant to the dialog but it also provides a constructive model for dynamically producing a representation of the tripartite discourse structure of (Grosz & Sidner 1986). A subdialog's intention is reflected in the theorem goal to be proven. The linguistic structure is based on the relationship of user utterances to the active missing axioms. This relationship is determined through the use of expectations for user responses when attempting to acquire missing axioms. These expectations model the attentional state.

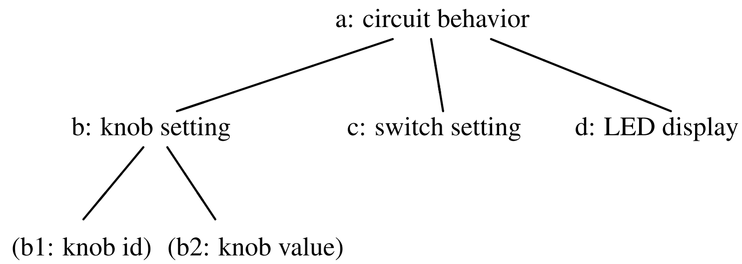

Figure 1: Sample theorem description

The development of the Circuit Fix-It Shop illustrates a case where the choice of the intentional model—acquiring missing axioms for completion proofs of task goals—led naturally to a coherent dialog interaction. The next case study illustrates a situation where there was one failed attempt because expectations of the conversational partner were not properly taken into account initially.

Case Study 2: The Command Post of the Future Story Capture System

The Command Post of the Future (CPOF) story capture system (Smith *et al.* 2002)[1] is a multimodal spoken natural language dialog system that was developed in order to provide a means for capturing a commander's expectations about planned military actions. If we think about these expectations as a story (Gershon & Page 2001), a possible machine representation for these expectations is a *story graph*, where the nodes of the graph represent expectations at particular moments in time and the arcs represent change of expectations between nodes. An example set of expectations in a node might be the following.

```
1. Delta company assumes an attack by
   fire position in building B.
2. Alpha company assumes an attack by
   fire position on hill C.
3. Echo company assumes an observation
   position on hilltop D.
```

Dialog context is used for handling issues concerning military echelons, multiple databases, dynamic name creation, and relative time references. A first attempt at developing the dialog model for this system focused on a communicative intention of determining the information for a story node. This was unsuccessful because the conversational participant, the commander, has no conception of a story node. What a commander does have as a working model as they plan is the notion of an *activity*. Various types of activities in which a military force can be engaged include movement, position establishment, and reconnaissance. Associated with an activity type are parameters, some of which are mandatory and some are optional. The dialog system will attempt to continue interaction about the current activity until values for all mandatory parameters are supplied. This approach is another instantiation of the Missing Axiom Theory of dialog. Modeling of the interaction of the dialog based on activities enables coherent interactions to occur. Ongoing work is focusing on extending the dialog model to allow for interaction about higher level strategic commander intent (Rogers 2003) (Adams 2006).

[1] Support for this system was originally provided by DARPA. Work has continued on this system as it serves as an interesting testbed for ideas.

Conclusion

This paper has briefly examined the role of modeling intention for enabling an intelligent system to engage in coherent human-computer dialog. What we have observed is that it is essential to have a correct model of intention that takes the expectations of the human conversational partner into account.

References

Adams, B. 2006. Activity based multimodal plan recognition of commander intent. Master's thesis, East Carolina University.

Cohen, P. R.; Morgan, J.; and Pollack, M., eds. 1990. *Intentions in Communication*. Cambridge, Mass.: MIT Press.

Gershon, N., and Page, W. 2001. What storytelling can do for information visualization. *Communications of the ACM* 31–37.

Grosz, B., and Sidner, C. 1986. Attentions, intentions, and the structure of discourse. *Computational Linguistics* 12(3):175–204.

Rogers, J. C. 2003. Developing bayesian networks for recognition of commander intent. Master's thesis, East Carolina University.

Smith, R.; Manning, B.; Rogers, J.; Adams, B.; Abdul, M.; and Alvarez, A. 2002. A dialog architecture for military story capture. In *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*, 184–187.

Smith, R.; Hipp, D.; and Biermann, A. 1995. An architecture for voice dialog systems based on Prolog-style theorem-proving. *Computational Linguistics* 21:281–320.