A Collaborative Assistant for Email

Dan Gruen, Candy Sidner, Carolyn Boettner

Lotus Development Corporation
55 Cambridge Parkway
Cambridge, MA 02142 USA
dan_gruen, csidner, carolyn_boettner@lotus.com

Charles Rich

MERL -- Mitsubishi Electric Research Lab
201 Broadway
Cambridge, MA 02139 USA
rich@merl.com

ABSTRACT

Software agents which communicate and collaborate with users to perform complex tasks constitute a new paradigm for human-computer interaction complementing existing graphical interfaces. We have recently completed a prototype agent of this kind for helping people with their email, based on our studies of people working with human assistants and Wizard-of-Oz studies. The prototype was constructed using application-independent software for modeling collaborative discourse (Collagen, see [4]) jointly developed by Lotus and Mitsubishi Electric and speech understanding technology from IBM Research. Users perform typical email tasks via a flexible combination of spoken language conversation with the agent and graphical interface actions (which are observed by the agent). The agent maintains a model of the user's goals and activities, and can act on its own initiative to assist the user. Having a high-level model of actions and goals allows speech to be used in a more natural, conversational, and effective manner than otherwise possible.

Keywords

Agent, collaboration, speech, conversation, email, calendar

INTRODUCTION

Collaborative interface agents constitute a new paradigm for human-computer interaction. Such agents require an explicit model of the user's task, represented in terms of the typical high-level goals and how these may be decomposed into lower-level goals and primitive actions. The agent uses this model to track progress by recognizing how the user's utterances and actions contribute to those goals, as well as to choose actions to perform on its own initiative which will advance the current goals. Our implementation software (Collagen) also allows us to easily vary our agent's level of initiative from very passive to very active, using the same task model.

UNDERSTANDING HUMAN ASSISTANTS

A crucial component of our research approach has been to study how human assistants work with their managers in their existing workplaces and in artificial settings we created. We first describe the results of these studies, and then we summarize the capabilities of our prototype agent.

Field Studies

We conducted interviews and observed work sessions with managers and assistants in several divisions of a large U.S. high-technology company. In the interviews, we asked about their procedures for working together and sharing information, and how these procedures had evolved. We observed episodes in which managers and assistants worked on email together. A common practice among high-level executives was to have their assistants print out their email and then go through the messages, page by page, at a daily meeting. During these meetings they would ask questions, dictate responses, review, modify and approve outgoing correspondence prepared by the assistant, delegate tasks, and handle outstanding matters brought to their attention by the assistant. The assistant would frequently jot notes on the email sheets as reminders of tasks to do later.

Wizard-of-OZ (WOZ) Studies

We also conducted a series of WOZ studies in which subjects performed tasks involving their email while speaking with a researcher who shared control of the email-reading software interface. Our goals in these studies were both to simulate a computer-based assistant and to elicit the email task model. Interestingly, as our studies progressed, we discovered that these two goals conflicted. The more we asked users to pretend they were dealing with a computer assistant, the more they seemed to restrict themselves to low-level commands rather than expressing their high-level goals. We were much more successful in eliciting high-level task structure when we told users to treat our experimenter as a human helper who understood general email and calendar issues.

RESULTS

Types of Assistance

We observed five categories of assistance provided by the human assistants:

Pre-Processing: Assistants frequently gathered additional information and prepared items before passing them to the manager.

Filtering/Prioritizing: Assistants selected the messages that needed the manager's attention and often ordered them by priority. They also determined if any messages required immediately interrupting the manager.

Adding Relevant Information: Assistants provided additional information, such as a reminder of a sender's affiliation, both when first presenting a message and while the manager read it.

Delegating Complex Tasks: Assistants frequently performed a number of complex steps in response to a single, often brief, request.

Peripheral Awareness/Drawing Attention to Items of Interest: Assistants frequently pointed out information they thought their managers would find important, such as the mention of a colleague in a message.

The Role of Speech

In both the observational and the WOZ studies, speech played a powerful role by providing an additional modality for communicating about items that were being attended to visually. Speech also allowed managers to express their higher level goals, which was simply not possible by only pointing and clicking on the graphical interface.

IMPLEMENTED SYSTEM

Based on our studies of human assistants we implemented a task model for email using the facilities provided in Collagen for representing objects, actions, goals, and goal decompositions (recipes). The model consists of about 70 goals and actions organized into a hierarchy by the associated recipes. Collagen provides our agent with predefined algorithms for using this recipe library to track the ongoing conversation, to participate in it, and to retain a history of the interaction hierarchically organized (segmented) according the goals that have been undertaken.

In particular the segmented interaction history is a valuable resource for users to remind them of the state of their current activity and what was done before. The history can either be viewed by the user as text or be accessed by voice. For example, upon returning from lunch or a meeting, the user may ask the agent *"What were we doing?"* and hear a summary of the high-level tasks that were performed.

As part of building our email agent, we also integrated a major extension (described in [3]) to Collagen, namely the capability of plan recognition. The addition of plan recognition dramatically reduced the amount the user needed to say to get things done, since the agent can now infer the intent of many actions.

Furthermore, earlier agents constructed with Collagen did not include any natural language or speech understanding. (Instead, the user was provided with a dynamically-changing menu of expected utterances from which to choose.) Our current agent, however, incorporates a complete speech understanding system from IBM Research, allowing users either to do their email work entirely in speech or with a mixture of speech and graphical actions, just as was done by our human subjects in the WOZ studies. For example, a user can ask *"Did I ever reply to that message from Robin"* while in the midst of reading a different message. The agent can say *"That's the week you're in Denver"* if the user opens an invitation to a meeting on a conflicting date, just as we observed human assistants doing.

In summary, we believe our email agent represents a new level of collaborative spoken-language capabilities well beyond previous state-of-the-art non-collaborating speech systems, such as Speech Acts [5], Jupiter at MIT [2], or the Trains system at the University of Rochester [1].

NEXT STEPS

We are expanding our email task model to incorporate more of the kinds of assistance we observed human assistants providing. We are beginning a new series of user studies in real work settings to determine how useful the agent is and the appropriate balance between agent initiative and intrusiveness. We are extending the agent to other areas, including calendar and interruption management. We also plan to use the segmented history as a substrate upon which to build support for reminders and task interleaving.

REFERENCES

1. Allen et al, "A Robust System for Natural Spoken Dialogue." *Proceedings of 34th Annual Meeting of the ACL*, 1996.

2. Chung. G. And Seneff, S. "Improvements in Speech Understanding Accuracy through the Integration of Hierarchical Linguistic, Prosodic, and Phonological Constraints in the Jupiter Domain" *ICSLP 98, Sydney, Australia,* November 1998.

3. Lesh, N. Rich, C. & Sidner, C. Using Plan Recognition in Human-Computer Collaboration, Lotus TR 98-14, 1998

4. Rich, C. & Sidner, C. "Collagen: A Collaboration Manager for a Collaborative Interface Agent." *UMUAI Journal*, Fall, 1998

5. Yankelovich, N., Levow, G., & Marx, M. "Designing Speech Acts: Issues in Speech User Interfaces" *CHI '95 Proceedings,* 369-376