Lecture Minutes: "Towards Conversational Human-Computer Interaction" by James Allen, Donna Byron, Myroslava Dzikovska, George Ferguson, Lucian Galescu, and Amanda Stent

Presented by: Cory Lueninghoener
Scribe: Robert Glaubius

March 13, 2002

1 Topic

The lecture for Wednesday, March 13 was given by Cory Lueninghoener, regarding the paper "Towards Conversational Human-computer Interaction", by Allen *et al*. Topics discussed were background information from Chapter 22 of [2], in addition to [1]. We began by discussing the background information with the idea of a speech act.

2 Background

A speech act is a communication interaction between a speaker, and a hearer, involving utterances. Speech acts can also include other events that convey information, such as body language. These acts can have various intents, such as informing, querying, answering, commanding, promising, acknowledging, and sharing.

A large portion of the difficulty of NLP stems from recognizing the intention of a speech act. This task involves parsing speech acts to determine the relevance of an utterance.

Approaches to conversational NLP is based on several communication steps involving both the speaker and the hearer. The speaker must first form an intention, then generate a statement that conveys this intent, which is then synthesized to form the actual speech act. The hearer, after perceiving this, must analyze and

disambiguate to determine the speakers intent, and then must take some action with respect to that act.

In his presentation, Cory distinguished between two communication models. The first of these, the encoded message model, involves complete statements that are able to stand on their own. In other words, a single speech act or statement conveys all necessary information. The second model is the situated language model, in which the context of the statement is essential for determining meaning.

3 Conversational HCI

The discussion then moved on to cover human-computer interfaces. One of the motives of conversational interfaces is to reduce the role of graphical user interfaces (GUIs). This is based on the observations that speech is easier for a human user, and that feature-rich GUIs tend to be cluttered. However, it was noted that we do not wish to remove the GUI component completely, as a completely dialogue-driven system requires the user to memorize a potentially large set of arbitrary commands.

The paper under discussion provides an overview of some approaches used in the development of the TRIPS system. TRIPS stands for the The Rochester Interactive Planning System, and is a system that the authors' have been working on for ten years. The system attempts to use dialogue-based interaction to accomplish concrete tasks.

A concrete task is one that satisfies some of the following. It attempts to solve a specific problem, or it intends to extract certain information. It may deliver advice and tutoring, or serve a command and control role. This focus on concrete tasks is derived from the *practical dialogue hypothesis*.

- *Practical Dialogue Hypothesis*: Conversational competence required for practical dialogues is simpler to achieve than general conversational competence.

Simply put, it is easier for an automated system to focus on a concrete task than on conducting a general-topic conversation. General-topic conversations require some proficiency in many different areas, while a concrete task requires high proficiency in one area.

To introduce modularity into the dialogue system, the authors propose and take advantage of the *Domain-Independence Hypothesis*.

- *Domain-Independence Hypothesis*: The bulk of complexity in a language interpretation and dialogue management is independent of the task being performed.

Basically, this hypothesis states that we can separate a conversation agent from the domain-dependent reasoning system, allowing the conversation agent to serve as a broker between the user and the reasoning system.

4 Challenges

The authors in [1] distinguish four challenges for a dialogue system that they've confronted in their work. These are as follows.

- *Parsing languages into practical dialogues*
- *Integrating dialogue and task performance*
- *Intention recognition*
- *Mixed initiative dialogue*

4.1 Parsing languages into practical dialogues

An interesting aspect of the work of Allen *et al*. is the focus on statement meaning rather than structure. Other work in this area has focused on parsing statements in order to determine syntactic form, whereas this work focuses on determining semantic form.

With the use of semantics to form a context for statements, it is then possible to attempt disambiguation. With a structure-based approach this would likely be more difficult. The context places constraints on meanings of ambiguous words, such as distinguishing between corn chips and silicon chips in the context of eating.

One tool in a meaning-based approach is the use of utterances over sentences. A sentence is not likely to be sufficient, since dialogue's don't tend to contain all the necessary information in every sentence. Utterances, on the other hand, are likely to preserve more of the relevant context.

4.2 Integrating dialogue and task performance

A goal of the TRIPS system was the use of a flexible dialogue system. This would allow the agent-based back-end to be separate from the dialogue system, which would handle transactions between the user and the back-end. Dr. Choueiry noted that the underlying dialogue structure resembles a planning problem. This structure involves objectives (goals), solutions, resources, and situations that must be organized. These objects are manipulated through utterances.

4.3 Intention recognition

Intention recognition is the problem of determining what the speaker meant by an utterance. Cory gave three examples drawn from the paper that displayed this difficulty, when we're dealing with a current objective: moving people from Abyss to Delta using truck one.

- "Can we use a helicopter to get the people from Abyss?"
- "Can we use a helicopter to get the people at Barnacle?"
- "Can we use a helicopter to get the people from Delta?"

The first statement involves the addition of a second transition that may help accomplish our goal. The second involves the addition of another subgoal, while the third involves extending our plan.

TRIPS approaches the intention recognition problem with a two step process. First, it derives the possible intentions of an utterance; this set is then pruned to remove irrelevant intentions with respect to the context.

4.4 Mixed Initiative dialogue

Previous systems have focused on dialogue simplifications involving conversational initiative. A member of a conversation has initiative if he, she, or it drives the conversation. While human-human dialogue typically involves exchanges of initiative between conversants, implemented systems often restrict initiative. These fixed-initiative systems restrict initiative to either the user or the system. Allen *et al*. reject this approach as overly restrictive, and focus on the more complex problem of mixed initiative.

Mixed initiative dialogue is complex due to the fact that it is necessary to determine which conversant has the initiative at a given moment (scribe's note: for those who've read *Lord of the Flies*, think "conch"), as well as handling interruptions which may cause an arbitrary transfer of initiative, as well as disrupt context.

5 TRIPS

TRIPS has been constructed incrementally from a simple dialogue system, based on the result of studying human-human interaction. It is divided into three components: Interpretation, Generation, and Behavior. Interpretation focuses on determining the meaning and intent of utterances, while the behavior agent determines appropriate action. The behavior agent interfaces with the back-end application.

The Generation manager synthesizes a reply to the user. The three components are fairly independent, and interact at need.

6 Conclusion

We were rendered bereft of Cory's multimedia extravaganza by unfortunate circumstances that prevented us to use realplayer. Due to this, we jumped directly to the conclusion. We covered some of the practical benefits of dialogue-based interfaces. For instance, it seems that this may be a favorable method for introducing objects into a planning system. Another that was not discussed is the application to those who do not have eyesight.

However, there are a large number of complex issues, as were discussed earlier. Simplifications, such as the focus on practical dialogue help to make the system more effective. Additionally, the use of context, a focus on utterances over sentences, and modular architecture help to strengthen the approach in TRIPS.

7 Discussion

- Tibor
  - Mentioned that this is an exciting research area, especially in terms of combining the use of dialogue and GUI.
  - Stated his preference for fixed user initiative as a deterrent to HAL taking over the shuttle
  - Asked if anyone has successfully used speech-driven cell phone features (e.g. automated dialing); several people indicated that they have successfully used or observed this, as well as OS speech commands and voice driven web page navigation.
- Dan
  - Disagreed with the authors' comments that statistical methods for disambiguation are not as useful as other approaches.
  - Pointed out that the statement proclaiming that dialogue systems spark a new revolution may be a bit shaky.
- Rob

  - Raised a question about the domain-independence hypothesis and its applicability. It was pointed out that domain-independence refers to the use of generic plan objects in specific domains.

The following questions were not raised during class due to time constraints. These are taken from their written comments.

- Lin Xu
  - Since there are several significant technical problems remaining after 10 years of work, what have they actually been doing?
  - Mentions that he enjoyed the paper, and that even with a lot of work left to do, they still think it's important work.
  - Asks how bad the delay between question and response in the system is.
- Amy
  - Points out the use of the limited-domain heuristic from NLP, e.g. the idea that NLP can be successful when the domain is limited. This is an interesting application in that the limitation is in terms of dialogue types rather than the specific domain.
  - Mentions that the focus on speech acts rather than sentences is interesting, and feels that it is likely to be a more successful approach.
  - Wonders if the rejection of statistical training methods implies that systems are hand-coded, or what training method is used. How is the knowledge encoded?

References

[1] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, and L. Galescu. Towards conversational human-computer interaction, 2001.

[2] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Englewood Cliffs, New Jersey, 1995.