Managing Uncertainty in Dialogue Information State for Real Time Understanding of Multi-Human Meeting Dialogue

Alexander Gruenstein, Lawrence Cavedon, John Niekrasz, Dominic Widdows, and Stanley Peters
Center for the Study of Language and Information
Stanford University, Stanford, CA 94305
{alexgru,lcavedon,niekrasz,dwiddows,peters}@csli.stanford.edu

1 Introduction

Human speech processing is riddled with ambiguity and uncertainty on a number of levels: *e.g.* uncertainty of speech-processing; lexical and structural ambiguity in parsing; dialogue–act classification; intention recognition and interpretation. Information-state approaches to dialogue management typically only maintain a single current state and utilize strategies for resolving ambiguities and uncertainty immediately they arise.[1]

We are concerned with tracking and understanding dialogue between multiple human participants—specifically, in meetings—in such a way that the dialogue system does not intervene. In this scenario, the system is not able to provide feedback on whether or not it has understood, and is unable to ask for clarification or ambiguity resolution. Our ultimate aim is to model human-human dialogue (to the extent that it is feasible) in real-time, providing useful services (*e.g.* relevant document retrieval) and answering queries about the dialogue state and history (*e.g.* "what action items do we have so far?"). Our approach has been to extend our existing dialogue system, based on the information-state update approach—which supports a rich semantic interpretation of multi-utterance constructions—to cope with the added uncertainty inherent in two-person meetings in which the participants speak, point, and draw on a whiteboard.

[1] Some previous work has considered the issue of dialogue management under uncertainty (e.g. (Levin et al., 2000; Roy et al., 2000)) but has not generally involved rich semantic dialogue states, linking speech directly to action.

1.1 Meeting artifacts and information state

We focus exclusively on meetings about *artifacts*: *i.e.* meetings that produce some constructed object as its end, such as a project plan with tasks and deadlines (*i.e.* a Gantt chart), or budget in some sort of spreadsheet format. This focus provides a concrete frame for interpretation of drawing and of spoken language.

Artifacts are represented in an ontology designed using Protégé,[2] including classes for the objects themselves (*e.g.* a plan and its components), relations among these entities, and the events which affect state-change in the entities or relations. The current artifact state, as represented by the ontology, is part of the information state of the dialogue and contributes to the interpretation of utterances. Indeed, most utterance sequences in our scenario can be viewed to have semantics defined in *operations* over the artifact under discussion.

A meeting history viewer graphically displays the relationships between changes to the artifacts in the information state and the utterances and actions which caused those changes. This provides a useful visual into the internals of the system, and comprises a tool by which a meeting can be indexed, allowing a user to skip to the dialogue segment associated with additions or changes to the artifact (*e.g.* revisiting the negotiation associated with the choice of a milestone date). Unlike standard meeting summarization systems, the history viewer is cross-indexed by both artifact and dialogue.

[2] protege.stanford.edu

2 Uncertainty management

There has been much work on dialogue management systems to detect and resolve ambiguity, such as by combining multiple sources of evidence—*e.g.* multimodal systems that combine speech and drawing/gesture (Oviatt, 2000), or systems that use prosodic features to help classify speech-acts (Venkataraman et al., 2003)—or by using corpus-based statistical techniques to identify most likely interpretation. However, little work has been done on *maintaining* the uncertainty that arises from such ambiguity over extended periods of time, rather than resolving it soon after its detection.

Previous applications of our dialogue management system—e.g. (Lemon et al., 2002))—have ignored uncertainty in interpretation and have resolved ambiguity immediately as it arose: *e.g.* only the top item of the speech-recognizer's n-best list was considered (regardless of probability), and clarification questions were used to resolve ambiguity. However, in the meeting-understanding application, uncertainty management becomes necessary as the system has only limited mechanisms for resolving detected ambiguities without intruding on the normal flow of the meeting.

2.1 Incorporating ASR uncertainty into dialogue state

An initial implementation of uncertainty in our dialogue state framework is to incorporate multiple results from an n-best list into the Dialogue Move Tree (DMT). As in previous work, each incoming utterance is classified as a type of dialogue move, and a corresponding node is attached to the DMT using an attachment algorithm (see (Lemon et al., 2002)). Here, however, all speech-rec results which can be interpreted in context are simultaneously attached to the dialogue move tree—these assignments are weighted depending on recognizer and dialogue-move classification confidences. As more evidence becomes available, either through subsequent utterances or through multimodal evidence,[3] nodes which represent unlikely interpretations are pruned from the DMT. The idea is that the tree may contain arbitrarily long threads representing competing interpretations of conversations which will be pruned as new evidence rules out unlikely threads.

[3]Multimodal integration is performed in collaboration with the Center for Human-Computer Communication at Oregon Graduate Institute; see (Kaiser et al., 2003)

2.2 Current and future directions

Many meetings have at least an outline of structure, such as a formal or pre-agreed agenda. Some agenda items may be directly related to the meeting artifact or component thereof (*e.g.* deciding the delivery date of a task). A direction we are currently exploring, one which does not seem to have been pursued in previous meeting-understanding projects, is to include some representation of meeting-state—as measured by progress against an agenda—to the dialogue information state. We are also investigating techniques for automatically detecting topic shifts. Such information can be added to dialogue-state and used to prime ASR language-models and disambiguate spoken utterances. Links from utterance to agenda-item or topic are themselves highly uncertain of course, and will require more sophisticated probabilistic models to be incorporated into the dialogue management process.

References

E. Kaiser, A. Olwal, D. McGee, H. Benko, A. Corradini, X. Li, P. Cohen, and S. Feiner. 2003. Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In *ICMI 2003*.

O. Lemon, A. Gruenstein, and S. Peters. 2002. Collaborative activities and multi-tasking in dialogue systems. *Traitment automatique des langues*, 43(2).

E. Levin, R. Pieraccini, and W. Eckert. 2000. A stochastic model of human-machine interaction for learning dialogue strategies. *IEEE Transactions on Speech and Audio Processing*, 8(1).

S. L. Oviatt. 2000. Taming speech recognition errors within a multimodal interface. *CACM*, 43(9).

N. Roy, J. Pineau, and S. Thrun. 2000. Spoken dialog management for robots. In *ACL 2000*, Hong Kong.

A. Venkataraman, L. Ferrer, A. Stolcke, and E. Shriberg. 2003. Training a prosody-based dialog act tagger from unlabeled data. In *IEEE-ICASSP*, Hong Kong.