Modeling Situated Conversational Agents as Partially Observable Markov Decision Processes

Will Thompson
Dept. of Linguistics
Northwestern University
william-thompson@northwestern.edu

Darren Gergle
Dept. of Communication Studies
Northwestern University
dgergle@northwestern.edu

ABSTRACT

A *Situated Conversational Agent* (SCA) is an agent that engages in dialog about the context within which it is embedded. An SCA is distinguished from non-situated conversational agents by an intimate connection of the agent's dialog to its embedding context, and by intricate dependencies between its linguistic and physical actions. Constructing an SCA that can interact naturally with users while engaged in collaborative physical tasks requires the agent to interleave decision making under uncertainty, action execution, and observation while maximizing expected utility over a sequence of interactions. These requirements can be fulfilled by modeling an SCA as a *partially observable Markov decision process* (POMDP). We show how POMDPs can be used to formalize and implement psycholinguistic proposals on how situated dialog participants collaborate in order to make and ground dialog contributions.

Author Keywords

situated conversational agents, decision-theoretic planning

ACM Classification Keywords

natural language interfaces, theory and models

INTRODUCTION

A *Situated Conversational Agent* (SCA) is an agent that engages in dialog about the context within which it is embedded. An SCA is distinguished from non-situated conversational agents by an intimate connection of the agent's dialog to its embedding context, and the interleaving of linguistic with physical actions. The research described in this paper focuses on situated dialog while engaged in *collaborative physical tasks* [5],[6],[9], in which multiple agents interact with one another in order to collaboratively achieve some goal with respect to the physical context within which they are embedded. Collaborative physical tasks, such as fixing a bicycle [9], building a Lego[TM] model [2], or solving an on-line jigsaw puzzle [5],[6],[7], can potentially occur in a wide variety of contexts, either real or virtual.

For an SCA engaged in a collaborative physical task, the decision about whether to select a linguistic or physical action is based on a variety of interacting factors. These include the degree of uncertainty the agent has about the dialog context, the immediate cost incurred by an action versus its potential future payoff, as well as the likelihood of success for the action – factors that are determined in turn by properties of the available communication media and the nature of the task itself. All of this information can potentially influence an SCA's decision making about which actions to take, and representing and using this information appropriately is necessary for a situated agent to engage in efficient and natural dialogs in support of collaborative physical tasks.

Within the field of psycholinguistics, these phenomena have been most directly addressed by Herb Clark's *grounding theory* of dialog [1]. Grounding theory emphasizes the idea that dialog is a joint activity, in which participants collaborate in order to add information to their common ground of shared beliefs. Adding information to the common ground is not an automatic effect of making an utterance. Rather, when a speaker initiates a dialog contribution, it is only accepted as part of the common ground when the addressee provides evidence of understanding sufficient for current purposes, in a process known as *grounding*. According to grounding theory, interlocuters attempt to ground utterances to a degree sufficient for current purposes, while attempting to minimize collaborative effort in the process.

Applying these ideas to the modeling and construction of a conversational agent naturally suggests the use of *decision theory*, which marries probability theory with the concept of utility. The goal of a decision-theoretic agent is to select an action that maximizes expected utility. Maximizing expected utility over a sequence of actions is the domain of *partially observable Markov decision processes* (POMDPs) [8]. POMDPs provides a principled basis for modeling an SCA, where the agent must decide whether or not to execute a linguistic or physical action. Within a POMDP approach, both information gathering and physical actions are subjected to a type of value of information calculation, and can be weighed against each other when choosing an action that maximizes expected utility over a particular time horizon.

Our research is closely related to recent work on formulating spoken dialog systems as POMDPs [12]. We extend this work by applying POMDPs to situated dialog. Our re-
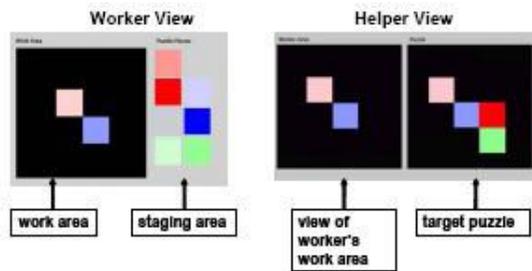
Figure 1. The Puzzle Task

Figure 2. Example Puzzle Task transcripts

search is also closely related to that of [10], who describe a decision-theoretic multimodal dialog architecture. The main difference between their work and ours is our adoption of a POMDP approach, which models sequential decision making rather than one-shot decisions. The work of [4] is also closely related. They describe an implemented dialog system where sentential ambiguities are translated into uncertainty about dialog context, and they model dialog agents that maintain a running estimate of uncertainty. We differ from this primarily in the adoption of a decision-theoretic model of planning. Finally, our research is inspired by David Traum's original work on computational models of grounding [11], where the idea of using a decision-theoretic approach to conversational grounding is proposed.

THE PUZZLE TASK

Our data on dialog about collaborative physical tasks come from a series of experiments performed by Gergle, Kraut and Fussell [5],[6],[7]. These experiments involved pairs of subjects working together online to solve a virtual jigsaw puzzle, such as the one shown in Figure 1. In this task, the subjects were given the roles of *helper* and *worker*. During each trial the helper had private access to a solved version of the puzzle, and her task was to provide verbal instructions to the worker on how to build the puzzle. The worker's task was to follow these instructions by selecting from a set of puzzle pieces and moving them into a workspace arranged in the goal configuration. In a subset of trials the helper shared a view of the worker's workspace. In other trials, the helper had no view of the worker's workspace, and their interaction was restricted to audio communication only.

An analysis of the data obtained from these experiments reveals that both communication efficiency and communicative processes show substantial differences across the shared vs. not-shared condition. Pairs were about a third quicker at solving puzzles, with fewer words per unit of time, when a high fidelity view of the workspace was shared, and this difference was greater when task complexity was higher. The worker's efficiency in this respect was affected to a greater degree than the helper's. Figure 2 (from [7]) illustrates how language changes when a shared visual workspace is available. When the shared visual workspace is available, responsibility for checking the state of the puzzle shifts from the worker to the helper, there are fewer verbal grounding acts, fewer words, and fewer turns.

Based on these results, and results from similar studies, we highlight two properties of situated dialog about collaborative physical tasks:

P1 *Interchangeability of language and action.* In a collaborative physical task, worker actions have the dual role of communicating information as well as accomplishing task goals. A shared visual workspace provides both agents with high fidelity observations about the state of the task, and the state of the ongoing actions of the participants. Agents who are aware their actions are being observed produce fewer explicit reports of their current state of understanding. Agents that can observe others' actions produce fewer explicit requests for grounding feedback [6].

P2 *Interleaving of action and language in fine increments.* Situated dialog often consist of short speech increments, finely timed with others' actions. The pauses between the speech increments are precisely timed to correspond with the Worker's actions, and these actions can in turn determine the subsequent course of the dialog [2]. The participants work together in order to *ground* dialog contributions at a fine time-scale, with dialog often consisting of repeated productions of short sentential fragments, followed by verbal or non-verbal responses.

Properties of P1 and P2 can be explained under the grounding theory of dialog, according to which dialog participants use the most efficient means at their disposal to communicate and ground information, and achieve the task goal of solving the puzzle. Property P1 indicates that physical actions affect not only the task state, but also dialog state. When a dialog participant executes a physical action that she knows is observable by the other participants, it becomes unnecessary to verbally describe either the action or the results of the action. When actions are cheap, reliable, and reversible, as in the Puzzle Task, it can be more efficient to indicate understanding by executing a physical act than to verbally ground a helper's instruction. Property P2 can be explained under an analysis in which the helper is minimizing collaborative effort over a *sequence* of interactions. The helper can produce shorter (and less effortful) instructions,
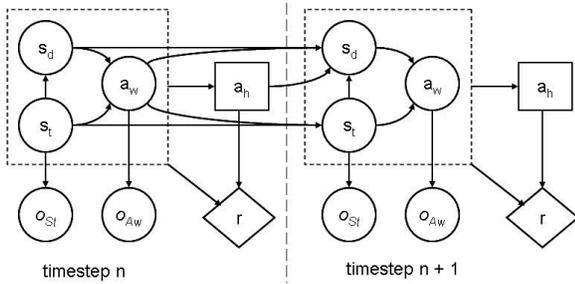
Figure 3. Dynamic influence diagram of a helper POMDP

because the helper is aware that the worker can provide immediate, cheap, and highly reliable feedback in the form of a physical action. In the next section we outline how these intuitive explanations of P1 and P2 can be formalized and implemented using a POMDP approach to situated dialog.

A POMDP MODEL OF THE PUZZLE TASK

A POMDP is formally defined as a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \Omega, \mathcal{O} \rangle$ where $\mathcal{S}$ is a finite set of states of the world, $\mathcal{A}$ is a finite set of actions, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\mathcal{S})$ is the state-transition function, giving for each world state and agent action a probability distribution over world states, $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function giving the expected immediate reward gained by the agent for taking each action in each state, Ω is a finite set of observations of the world, and $\mathcal{O} : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\Omega)$ is the *observation function*, giving for each action and resulting state, a probability distribution over possible observations [8].

After each action, an agent receives an observation about the world. This observation is only probabilistically related to the actual state of the world and to the action executed by the agent. Since a POMDP agent does not have direct access to the world, it must maintain a running estimate of what state it is in. This running estimate, the *belief state*, is updated after every action and observation, and forms the basis for what actions will be taken next. A solution to a POMDP is a *policy*, which is a function $\pi : b(s) \rightarrow A$, that maps from belief states to actions. An *optimal policy* (usually denoted π^*) is a policy that maximizes expected return over some time horizon. A variety of techniques exist for finding exact or approximate optimal policies for POMDPs [8].

A particularly perspicuous way to represent a POMDP is with a dynamic influence diagram [3]. A dynamic influence diagram is structured like a dynamic Bayesian network, with oval chance nodes representing random variables. In addition to chance nodes however, there are square nodes representing agent decisions, and diamond nodes representing agent rewards. A dynamic influence diagram representing the helper's role in the Puzzle Task is shown in Figure 3 (cf. Figure 1 in [12]).

The diagram in Figure 3 shows the world state (in the dashed box) factored into three components, $\langle S_t, S_d, A_w \rangle$. These represent, respectively, the puzzle task state, the dialog state, and the last action taken by the worker, which may be ei-

ther a task action or a dialog action. The arrows in the diagram represent influence: taking just these three variables, the state of the task influences the state of the dialog, while the task state and dialog state jointly influence the worker's action. In addition to these state variables, there are two observation variables, $\langle O_{S_t}, O_{A_w} \rangle$, representing observations of the task state and worker action, respectively.

At each time step, the helper receives these observations, and uses this evidence (along with knowledge of its own previous action) to update the belief state. An action is chosen based on this belief state, and a reward r is received. This influence diagram therefore models the case where there is a shared visual workspace. For the helper POMDP, the only actions available are linguistic actions. This is reflected in the influence diagram, where an arc leads from the helper's action node to the dialog state in the next time step. The worker, on the other hand, can influence both the dialog state and the task state, since the worker can manipulate the contents of the workspace.

Given this model of the helper, we can now say more on how property P2 can be derived, with respect to the identification of the next puzzle piece to be added to the workspace. We start by providing the helper POMDP with a set of referring expressions for identifying puzzle pieces, of varying incremental lengths. Longer referring expressions contain more information, but cost more to execute than shorter referring expressions. Because the helper can directly observe the worker's physical actions, and the resulting changes to the task state, the resultant belief state mass from these observations leaves little uncertainty with respect to the dialog and task states. This reduction in uncertainty leads to higher expected utilities. The helper POMDP will therefore be led into making shorter, incremental referring expressions, on the expectation that visual feedback from the worker will result in a highly certain estimate of dialog and task state.

The diagram in Figure 4 shows a POMDP for the worker role in the Puzzle Task. The world state representation is similar to that of the helper POMDP, but with the action chance variable and corresponding observation replaced by the action of the helper. Likewise, the decision node now consists of the worker's action, and the reward node represents the worker's reward. The structure of the influence arcs has also changed, with the helper's action influencing only the subsequent dialog state (representing the fact that the helper cannot directly modify the task state), and the worker's actions influencing both task state and dialog state.

Given this model of the worker, we can now say more about how property P1 can be derived, again with respect to the identification of the next puzzle piece to be added to the workspace. The worker POMDP can select from a set of task actions (here, moving a particular puzzle piece into the workspace) and dialog actions (proposing a description of a puzzle piece that matches the one the helper specified). As before, task actions result in a high degree of certainty with respect to the world state – in particular for this case, the dialog state. This high degree of certainty leads to higher
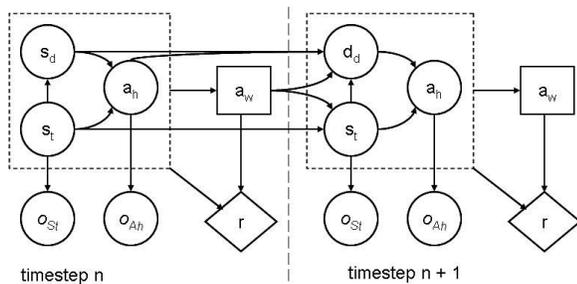
3

Figure 4. Dynamic influence diagram of a worker POMDP

expected utility, and leads the worker agent to prefer (in this case) to take a task action rather than a dialog action. In order to make this even more likely, it is possible to apply a higher cost to executing dialog actions than to task actions. This is reasonable if we make the assumption that both types of action require the worker to do the work of identifying a puzzle piece, but the task action is simpler to execute because it does not require any further process of planning and generating an appropriate referring expression.

CONCLUSION AND FUTURE WORK
Situated dialog is distinguished from non-situated dialog by an intimate connection of the dialog to the embedding context, and by intricate dependencies between linguistic and physical actions. An analysis of data from the Puzzle task corpus [5],[6],[7] reveals substantial differences in both communication efficiency and communicative processes when dialog participants collaborating on a physical task share a visual workspace. Worker actions substitute for language, the amount of explicit verbal grounding is reduced, and dialog turns often consist of incremental instructions from the helper interleaved with task actions by the worker. A Situated Conversational Agent (SCA) that engages in natural and efficient dialog with a user should be able to duplicate these properties in an appropriate manner. Partially observable Markov decision processes (POMDPs) provide a natural means for modeling the properties of situated dialog highlighted in this paper.

We are currently working on an implementation of the models described above. One of the main tasks here is to estimate the probabilities in the model. Some of these we can obtain from the Puzzle Task corpus itself, while others (such as the dialog model) can be handcrafted using dialog rules (see [12]). The reward functions also need to be constructed, in such a manner that the desired behavior manifests itself in the solved POMDPs. We plan to evaluate the POMDP policies we obtain against a set of handcrafted policies (again see [12]) in order to compare their relative performance.

In another planned evaluation, we will construct minimally different versions of the worker and helper POMDPs that correspond to the condition in the Puzzle Task experiment in which the participants do not share a visual workspace. We are interested in examining how the lack of a shared visual workspace impacts turn-taking and grounding behaviors (properties P1 and P2). For example, we would predict that

the helper POMPD will produce longer and more detailed referring expressions, since cheap and reliable feedback from the worker is no longer available in the form of visual evidence. If specified properly, this behavior will fall out from the POMDP models without the need to explicitly program them to perform this way.

REFERENCES
1. H. Clark. *Using Language*. Cambridge University Press, 1996.

2. H. Clark and M. Krych. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1):62–81, 2004.

3. T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence Journal*, 5(3):142–150, 1989.

4. D. DeVault and M. Stone. Managing ambiguities across utterances in dialogue. In *The 2007 Workshop on the Semantics and Pragmatics of Dialogue (DECALOG 2007)*, 2007.

5. D. Gergle, R. Kraut, and S. Fussell. Action as language in a shared visual space. In *Proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW 04)*, pages 487–496, 2004.

6. D. Gergle, R. Kraut, and S. Fussell. Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, 23:491–517, 2004.

7. D. Gergle, D. Millen, R. Kraut, and S. Fussell. Persistence matters: Making the most of chat in tightly-coupled work. In *Proceedings of CHI 2004*, 2004.

8. L. Kaelbling, M. Littman, and A. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.

9. R. Kraut, S. Fussell, and J. Siegel. Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18:13–49, 2003.

10. T. Paek and E. Horvitz. Conversation as action under uncertainty. In *Proceedings of the 16th Annual Conference on Uncertainty in Artificial Intelligence (UAI-00)*, San Francisco, CA, 2000. Morgan Kaufmann.

11. D. Traum. Computational models of grounding in collaborative systems. In *Working Notes of the AAAI Fall Symposium on Psychological Models of Communication*, pages 124 – 131. AAAI, November 1999.

12. J. D. Williams, P. Poupart, and S. Young. Factored partially observable Markov decision processes for dialogue management. In *4th Workshop on Knowledge and Reasoning in Practical Dialog Systems, IJCAI, Edinburgh*, 2005.