

Demonstration of a POMDP Voice Dialer

Jason Williams

AT&T Labs – Research, Shannon Laboratory
180 Park Ave., Florham Park, NJ 07932, USA
jdw@research.att.com

Abstract

This is a demonstration of a voice dialer, implemented as a partially observable Markov decision process (POMDP). A real-time graphical display shows the POMDP's probability distribution over different possible dialog states, and shows how system output is generated and selected. The system demonstrated here includes several recent advances, including an action selection mechanism which unifies a hand-crafted controller and reinforcement learning. The voice dialer itself is in use today in AT&T Labs and receives daily calls.

1 Introduction

Partially observable Markov decision processes (POMDPs) provide a principled formalism for planning under uncertainty, and past work has argued that POMDPs are an attractive framework for building spoken dialog systems (Williams and Young, 2007a). POMDPs differ from conventional dialog systems in two respects. First, rather than maintaining a single hypotheses for the dialog state, POMDPs maintain a probability distribution called a *belief state* over many possible dialog states. A distribution over a multiple dialog state hypotheses adds inherent robustness, because even if an error is introduced into one dialog hypothesis, it can later be discarded in favor of other, uncontaminated dialog hypotheses. Second, POMDPs choose actions using an optimization process, in which a developer specifies high-level goals and the optimization works out the detailed dialog plan. Because

of these innovations, POMDP-based dialog systems have, in research settings, shown more resilience to speech recognition errors, yielding shorter dialogs with higher task completion rates (Williams and Young, 2007a; Williams and Young, 2007b).

Because POMDPs differ significantly from conventional techniques, their operation can be difficult to conceptualize. This demonstration provides an accessible illustration of the operation of a state-of-the-art POMDP-based dialog system. The system itself is a voice dialer, which has been operational for several months in AT&T Labs. The system incorporates several recent advances, including efficient large-scale belief monitoring (akin to Young et al., 2006), policy compression (Williams and Young, 2007b), and a hybrid hand-crafted/optimized dialog manager (Williams, 2008). All of these elements are depicted in a graphical display, which is updated in real time, as a call is progressing. Whereas previous demonstrations of POMDP-based dialog systems have focused on showing the probability distribution over dialog states (Young et al., 2007), this demonstration adds new detail to convey how actions are chosen by the dialog manager.

In the remainder of this paper, Section 2 presents the dialog system and explains how the POMDP approach has been applied. Then, section 3 explains the graphical display which illustrates the operation of the POMDP.

2 System description

This application demonstrated here is a voice dialer application, which is accessible within the AT&T research lab and receives daily calls. The dialer's vo-

cabulary consists of 50,000 AT&T employees.

The dialog manager in the dialer is implemented as a POMDP. In the POMDP approach, a distribution called a belief state is maintained over many possible dialog states, and actions are chosen using reinforcement learning (Williams and Young, 2007a). In this application, a distribution is maintained over all of the employees' phone listings in the dialer's vocabulary, such as Jason Williams' office phone or Srinivas Bangalore's cell phone. As speech recognition results are received, this distribution is updated using probability models of how users are likely to respond to questions and how the speech recognition process is likely to corrupt user speech. The benefit of tracking this belief state is that it synthesizes all of the ASR N-Best lists over the whole dialog – i.e., it makes the most possible use of the information from the speech recognizer.

POMDPs then choose actions based on this belief state using reinforcement learning (Sutton and Barto, 1998). A developer writes a reward function which assigns a real number to each state/action pair, and an optimization algorithm determines how to choose actions in order to maximize the expected *sum* of rewards. In other words, the optimization performs planning and this allows a developer to specify the trade-off to use between task completion and dialog length. In this system, a simple reward function assigns -1 per system action plus +/- 20 for correctly/incorrectly transferring the caller at the end of the call. Optimization was performed roughly following (Williams and Young, 2007b), by running dialogs in simulation.

Despite their theoretical elegance, applying a POMDP to this spoken dialog system has presented several interesting research challenges. First, scaling the number of listings quickly prevents the belief state from being updated in real-time, and here we track a distribution over *partitions*, which is akin to a beam search in ASR (Young et al., 2006). At first, all listings are undifferentiated in a single master partition. If a listing appears on the N-Best list, it is separated into its own partition and tracked separately. If the number of partitions grows too large, then low-probability partitions are folded back into the master undifferentiated partition. This technique allows a well-formed distribution to be maintained over an arbitrary number of concepts in real-time.

Second, the optimization process which chooses actions is also difficult to scale. To tackle this, the so-called “summary POMDP” has been adopted, which performs optimization in a compressed space (Williams and Young, 2007b). Actions are mapped into clusters called *mnemonics*, and states are compressed into state feature vectors. During optimization, a set of template state feature vectors are sampled, and values are computed for each action mnemonic at each template state feature vector.

Finally, in the classical POMDP approach there is no straightforward way to impose rules on system behavior because the optimization algorithm considers taking any action at any point. This makes it impossible to impose design constraints or business rules, and also needlessly re-discovers obvious domain properties during optimization. In this system, a hybrid POMDP/hand-crafted dialog manager is used (Williams, 2008). The POMDP and conventional dialog manager run in parallel; the conventional dialog manager nominates a *set* of one *or more* allowed actions, and the POMDP chooses the optimal action from this set. This approach enables rules to be imposed and allows prompts to easily be made context-specific.

The POMDP dialer has been compared to a convention version in dialog simulation, and improved task completion from 92% to 97% while keeping dialog length relatively stable. The system has been deployed in the lab and we are currently collecting data to assess performance with real callers.

3 Demonstration

A browser-based graphical display has been created which shows the operation of the POMDP dialer in real time, shown in Figure 1. The page is updated after the user speech has been processed, and before the next system action has been played to the user. The left-most column shows the system prompt which was just played to the user, and the N-Best list of recognized text strings, each with its confidence score.

The center column shows the POMDP belief state. Initially, all of the belief is held by the master, undifferentiated partition, which is shown as a green bar and always shown first. As names are recognized, they are tracked separately, and the top 10

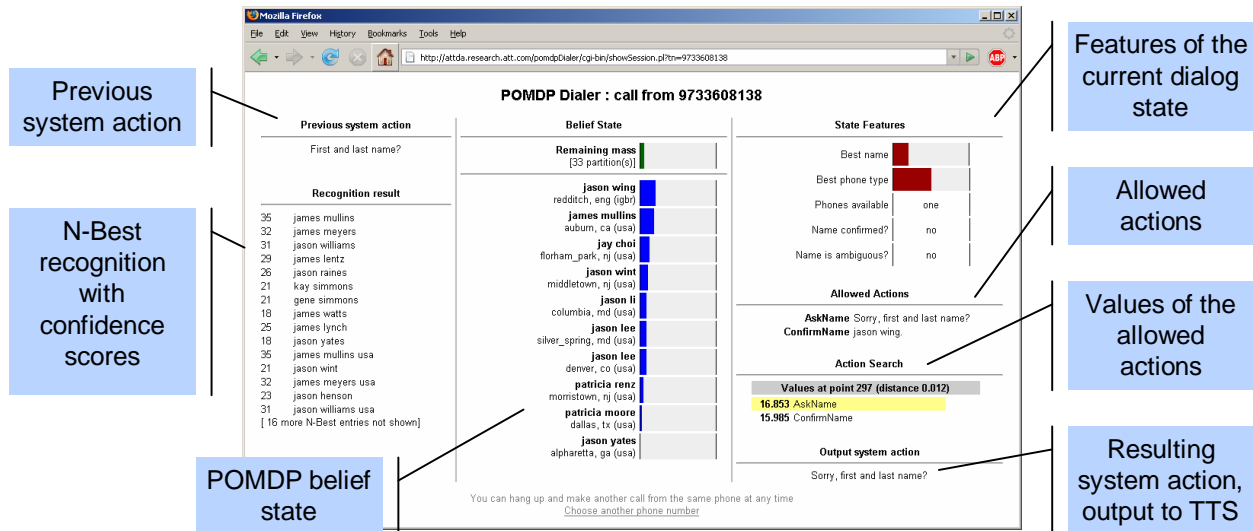


Figure 1: Overview of the graphical display. Contents are described in the text.

names are shown as blue bars, sorted by their belief. If the system asks for the phone type (office or mobile), then the bars sub-divide into a light blue (for office) and dark blue (for mobile).

The right column shows how actions are selected. The top area shows the features of the current state used to choose actions. Red bars show the two continuous features: the belief in the most likely name and most likely type of phone. Below that, three discrete features are shown: how many phones are available (none, one, or both); whether the most likely name has been confirmed (yes or no); and whether the most likely name is ambiguous (yes or no). Below this, the allowed actions (i.e., those which are nominated by the hand-crafted dialog manager) are shown. Each action is preceded by the action mnemonic, shown in bold. Below the allowed actions, the action selection process is shown. The values of the action mnemonic at the closest template point are shown next to each action mnemonic. Finally the text of this action, which is output to the caller, is shown at the bottom of the right-hand column. Figure 2 shows the audio and video transcription of an interaction with the demonstration.

4 Conclusion

This demonstration has shown the operation of a POMDP-based dialog system, which incorporates recent advances including efficient large-scale belief

monitoring, policy compression, and a unified hand-crafted/optimized dialog manager. A graphical display shows the operation of the system in real-time, as a call progresses, which helps make the POMDP approach accessible to a non-specialist.

Acknowledgments

Thanks to Iker Arizmendi and Vincent Goffin for help with the implementation.

References

- R Sutton and A Barto. 1998. *Reinforcement Learning: an Introduction*. MIT Press.
- JD Williams and SJ Young. 2007a. Partially observable Markov decision processes for spoken dialog systems. *Computer Speech and Language*, 21(2):393–422.
- JD Williams and SJ Young. 2007b. Scaling POMDPs for spoken dialog management. *IEEE Trans. on Audio, Speech, and Language Processing*, 15(7):2116–2129.
- JD Williams. 2008. The best of both worlds: Unifying conventional dialog systems and POMDPs. In *(In submission)*.
- SJ Young, JD Williams, J Schatzmann, MN Stuttle, and K Weilhammer. 2006. The hidden information state approach to dialogue management. Technical Report CUED/F-INFENG/TR.544, Cambridge University Engineering Department.
- SJ Young, J Schatzmann, B R M Thomson, KWeilhammer, and H Ye. 2007. The hidden information state dialogue manager: A real-world POMDP-based system. In *Proc NAACL-HLT, Rochester, New York, USA*.

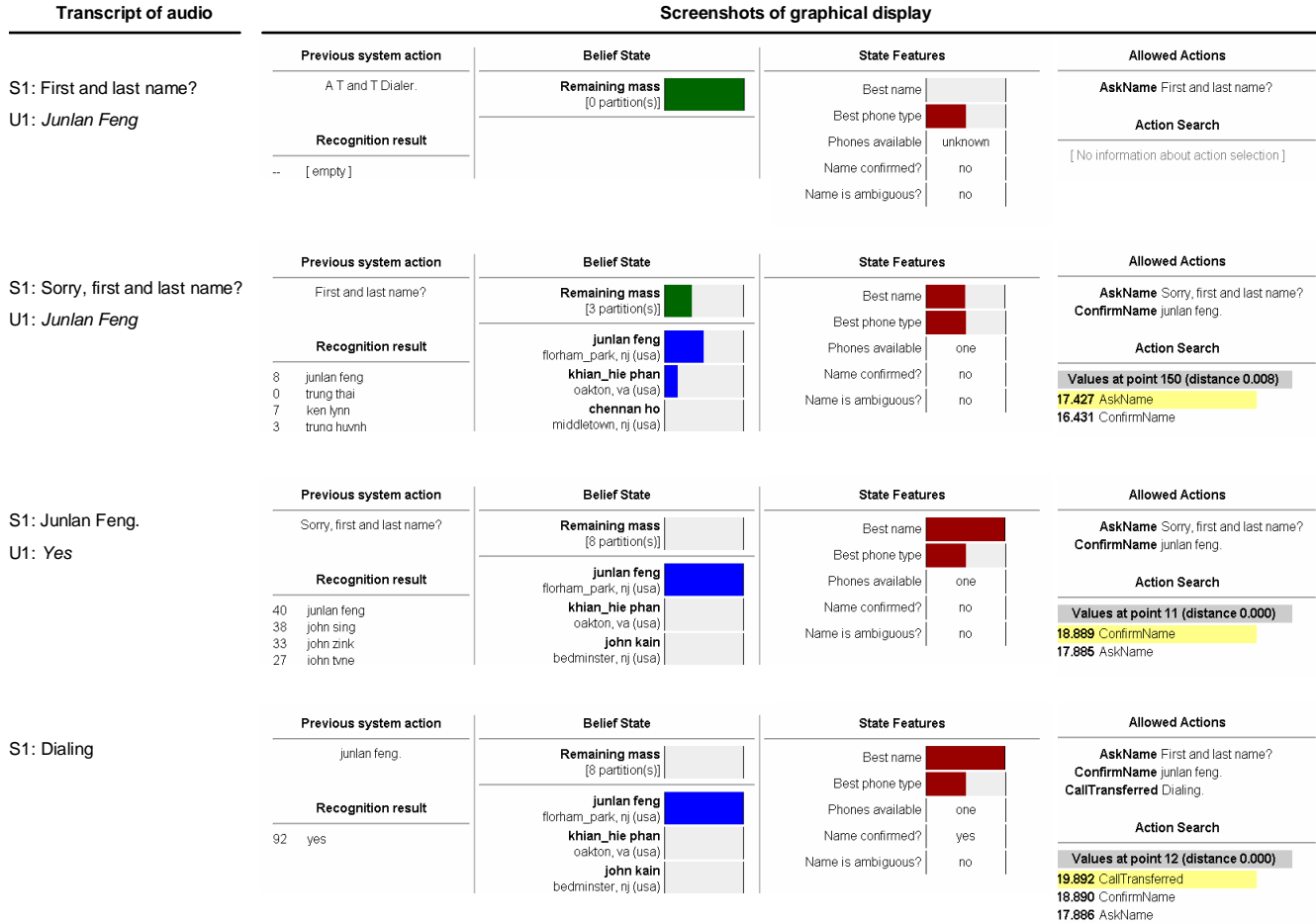


Figure 2: The demonstration’s graphical display during a call. The graphical display has been cropped and re-arranged for readability. The caller says “Junlan Feng” twice, and although each name recognition alone carries a low confidence score, the belief state aggregates this information. This novel behavior enables the call to progress faster than in the conventional system and illustrates one benefit of the POMDP approach. We have observed several other novel strategies not in a baseline conventional dialer: for example, the POMDP-based system will confirm a callee’s name at different confidence levels depending on whether the callee has a phone number listed or not; and uses yes/no confirmation questions to disambiguate when there are two ambiguous callees.