Investigating Clarification Strategies in a Hybrid POMDP Dialog Manager

Sebastian Varges and **Silvia Quarteroni** and **Giuseppe Riccardi** and **Alexei V. Ivanov**

Department of Information Engineering and Computer Science

University of Trento, 38050 Povo di Trento, Italy

{varges|silviaq|riccardi|ivanov}@disi.unitn.it

Abstract

We investigate the clarification strategies exhibited by a hybrid POMDP dialog manager based on data obtained from a phone-based user study. The dialog manager combines task structures with a number of POMDP policies each optimized for obtaining an individual concept. We investigate the relationship between dialog length and task completion. In order to measure the effectiveness of the clarification strategies, we compute concept precisions for two different mentions of the concept in the dialog: first mentions and final values after clarifications and similar strategies, and compare this to a rule-based system on the same task. We observe an improvement in concept precision of 12.1% for the hybrid POMDP compared to 5.2% for the rule-based system.

1 Introduction

In recent years, probabilistic models of dialog have been introduced into dialog management, the part of the spoken dialog system that takes the action decision. A major motivation is to improve robustness in the face of uncertainty, in particular due to speech recognition errors. The interaction is characterized as a dynamic system that manipulates its environment by performing dialog actions and perceives feedback from the environment through its sensors. The original sensory information is obtained from the speech recognition (ASR) results which are typically processed by a spoken language understanding module (SLU) before being passed on to the dialog manager (DM).

The seminal work of (Levin et al., 2000) modeled dialog management as a Markov Decision Process (MDP). Using reinforcement learning as the general learning paradigm, an MDP-based dialog manager incrementally acquires a policy by obtaining rewards about actions it performed in specific dialog states. As we found in earlier experiments, an MDP can learn to gradually drop the use of clarification questions if there is no noise. This is due to the fact that clarifications do not improve the outcome of the dialog, i.e. the reward. However, with extremely high levels of noise, the learner prefers to end the dialog immediately (Varges et al., 2009). In contrast to deliberate decision making in the pragmatist tradition of dialog processing, reinforcement learning can be regarded as low-level decision making.

MDPs do not account for the observational uncertainty of the speech recognition results, a key challenge in spoken dialog systems. Partially Observable Markov Decision Process (POMDPs) address this issue by explicitly modeling how the distribution of observations is governed by states and actions.

In this work, we describe the evaluation of a divide-and-conquer approach to dialog management with POMDPs that optimizes policies for acquiring individual concepts separately. This makes optimization much easier and allows us to model the confusability of concrete concept values explicitly. This also means that different clarification strategies are learned for individual concepts and even individual concept values. The use of the POMDP policies is orchestrated by an explicit task structure, resulting in a hybrid approach to dialog management. The evaluation involved a user study of 20 subjects in a tourist information domain. The system is compared against a rule-based baseline system in the same domain that was also evaluated with 20 subjects.

2 Hybrid POMDP dialog management

In this section we introduce the hybrid POMDP dialog manager that was used in the data collection.

2.1 Concept-level POMDPs

The domain is a tourist information system that uses 5 different policies that can be used in 8 different task roles (see below). For each concept we optimized an individual policy. The number of states of the POMDP can be limited to the concept values, for example a location name such as `trento`. The set of actions consists of a question to obtain the concept (e.g. `question-location`), a set of clarification actions (e.g. `verify-trento`) and a set of submit actions (e.g. `submit-trento`). POMDP modeling including a heuristically set reward structure follows the (simpler) 'tiger problem' that is well-known in the AI community (Kaelbling et al., 1998): the system has a number of actions to obtain further information which it can try and repeat in any order until it is ready to commit to a concept value. For optimization we used the `APPL` solver (Kurniawati et al., 2008).

2.2 Task structure and dialog management

The use of individual policies is orchestrated by an explicit task structure that activates and deactivates them. The task structure is essentially a directed `AND-OR` graph with a common root node. The dialog manager maintains a separate belief distribution for each concept. Figure 1 shows the general system architecture with a schematic view of the task structure, and additionally a more detailed view of an active location node. In the example, the root node has already finished and the system is currently obtaining the location for a lodging task. The term 'role' refers to a concept's part in the task, for example a month may be the check-in or check-out month for accommodation booking.

At the beginning of a dialog, the task structure is initialized by activating the root node. A top level function activates nodes of the task structure and passes control to that node. Each node maintains a belief b_c for a concept c, which is used to rank the available actions by computing the inner product of policy vectors and belief. The top-ranked action a_m is selected by the system, i.e. it is exploiting the policy, and passed to the natural language generator (NLG). Next, the top-ranked SLU results for the active node and concept are used as observation $z_{u,c}$ to update the belief to b'_c, which
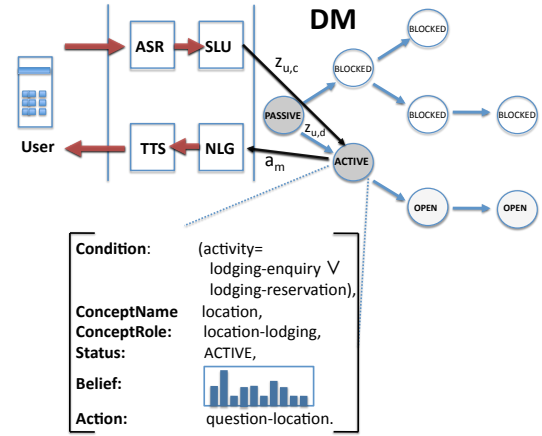

Figure 1: System architecture with Task Structure (task node example in detailed view)

follows the standard method for POMDPs:

$$b'_c(s') = \sum_{s \in S} b_c(s)\, T(s, a_m, s')\, O(a, s', z_{u,c})/p_{z_{u,c}}$$

(1)

where probability $b'_c(s')$ is the updated belief of being in state s', which is computed as the sum of the probabilities of transitioning from all previous belief points s to s' by taking machine action a_m with probability $T(s, a_m, s')$ and observing $z_{u,c}$ with (smoothed) probability $O(a_m, s', z_{u,c})$. Normalization to obtain a valid probability distribution is performed by dividing by the probability of the observation $p_{z_{u,c}}$.

A concept remains active until a submit action is selected. At that point, the next active node is retrieved from the task structure and immediately used for action selection with an initially uniform belief. Submit actions are not communicated to the user but collected and used for the database query at the end of the dialog.

Overanswering, i.e. the user providing more information than directly asked for, is handled by *delayed belief updating*: the SLU results are stored until the first concept of a matching type becomes active. This is a heuristic rule designed to ensure that a concept is interpreted in its correct role. Operationally, unused SLU results $z_{u,d}$ (where concept $d \neq c$) are passed on to the next activated task node (see also figure 1).

3 Experiments and data analysis

We conducted user studies with two systems involving 20 subjects and 8 tasks in each study. The systems use a Voice XML platform to drive ASR and TTS components. Speech recognition is

	Lodging Task		Event Enquiry	
	TCR	#turns	TCR	#turns
Rule-based DM	75.5%	13.7	66.7%	8.7
	(40/53)	(σ=4.8)	(28/42)	(σ=3.3)
POMDP-DM	78.1%	23.0	84.3%	14.4
	(50/64)	(σ=8.8)	(27/32)	(σ=4.5)

Table 1: Task completion and length metrics

	Rule-based DM			POMDP-DM		
	first	final	Δ%	first	final	Δ%
a) activity	0.78	0.74	-4.1	0.83	0.88	5.0
b) location	0.64	0.74	15.8	0.69	0.73	6.3
c) starrating	0.67	0.70	3.4	0.90	0.97	7.7
d) month	0.85	0.89	4.3	0.76	0.86	12.7
e) day	0.70	0.76	8.3	0.61	0.76	25.3
ALL (a-e)	0.74	0.78	**5.2**	0.74	0.83	**12.1**
Clarifications	0.84	0.85	1.5	0.96	0.87	-8.8

Table 2: Concept precision of first vs final value

based on statistical language models for the opening prompt, and is grammar-based otherwise. One system used the hybrid POMDP-DM, the other is a rule-based dialog manager that uses explicit, heuristically set confidence thresholds to trigger the use of clarification questions (Varges et al., 2008).

Dialog length and task completion Table 1 shows task completion rates ('TCR') and durations ('#turns') for the POMDP and rule-based systems. Task completion in this metric is defined as the number of tasks of a certain type that were successfully concluded. Duration is measured in the number of turn pairs consisting of a system action followed by a user action. We combine the counts for two closely related lodging tasks. The number of tasks is shown in brackets. Table 1 shows that the POMDP-DM successfully concludes more and longer lodging tasks and almost as many event tasks. In general, the POMDP policies can be described as more cautious although obviously the dialog length of the rule system depends on the chosen thresholds.

Concept precision at the value level In order to measure the effect of the clarification strategies in both systems, we computed concept precisions for two different mentions of a concept in a dialog (table 2): first mentions and final values after clarifications and similar strategies. The rationale for this metric is that the last mentioned concept value is the value that the system ultimately obtains from the user, which is used in the database query:

- if the system decides not to use clarifications, the only mentioned value is the accepted one,

- if the system verifies and obtains a positive answer, the last mentioned value is the accepted one,

- if the system verifies and obtains a negative answer, the user will mention a new value (which may or may not be accepted).

Thus, this metric is a uniform way of capturing the obtained values from systems that internally use very different dialog managers and representations. The actual precision of a concept C is calculated by comparing SLU results to annotations and counting true positives (matches M) and false positives (separated into mismatches N and entirely un-annotated concepts U): $Prec(C) = \frac{M}{M+N+U}$. Unrecognized concepts, on the other hand, are recall related and not counted since they cannot be part of any system belief.

As table 2 clearly shows, the use of clarification strategies has a positive effect on concept precision in both systems. The exception is the precision of concept activity in the rule-based system for which the system reprompted rather than verified.[1] In table 2, row 'All' refers to the average weighted precision of the five concepts. Both systems start from a similar level of overall precision. The relative improvement of the POMDP-DM for all concepts is 12.1%, compared to 5.2% of the rule-based DM.

We conducted a statistical significance test by computing the delta in the form of three values for individual data points, i.e. dialogs, and assigned `+1` for all changes from non-match to match, `-1` for a change in the opposite direction and `0` for everything else (e.g. from mismatch to mismatch). We found that, although there is a tendency for the POMDP-DM to perform better, the difference is not statistically significant at p=0.05 (a possible explanation is the data size since we are using human subjects).

We furthermore measured the precision of recognizing 'yes/no' answers to clarification questions. In contrast to actual concepts, there is no belief distribution for these in the DM since clarification actions are part of the concept POMDP models. We are thus dealing with individual one-off recognition results that should be entirely independent of each other. However, as table 2 (bottom)

---

[1]The second value obtained may be incorrect but above the confidence threshold; note that the rule system does not maintain a belief distribution over values.

shows, the precision of verifications decreases for the hybrid POMDP system. A plausible explanation for this is the increasing impatience of the users due to the longer dialog duration.

Characterization of dialog strategies For some concepts, the best policy is to ask the concept question once and then verify once before committing to the value (assuming the answer is positive). Other policies verify the same value twice. Another learned strategy is to ask the original concept question twice and then only verify the value once (assuming that the understood value was the same in both concept questions). In other words, the individual concept policies show different types of strategies regarding uncertainty handling. This is in marked contrast to the manually programmed DM that always asks the concept question once and verifies it if needed (concept activity being the exception).

HCI and language generation The domain is sufficiently simple to use template-based generation techniques to produce the surface forms of the responses. However, the experiments with the POMDP-DM highlight some new challenges regarding HCI aspects of spoken dialog systems: the choice of actions may not be 'natural' from the user's perspective, for example if the system asks for a concept twice. However, it should be possible to better communicate the (change in the) belief to the user.

4   Related work

The pragmatist tradition of dialog processing uses explicit representations of dialog structure to take decisions about clarification actions. These models are more fine-grained and often deal with written text, e.g. (Purver, 2006), whereas in spoken dialog systems a major challenge is managing the uncertainty of the recognition. Reinforcement learning approaches to dialog management learn decisions from (often simulated) dialog data in a less deliberative way. For example, the Hidden Information State model (Young et al., 2010) uses a reduced summary space that abstracts away many of the details of observations and dialog state, and mainly looks at the confidence scores of the hypotheses. This seems to imply that clarification strategies are not tailored toward individual concepts and their values. (Bui et al., 2009) uses factored POMDP representations that seem closest to

our approach. However, the effect of clarifications does not seem to have been investigated.

5   Conclusions

We presented evaluation results for a hybrid POMDP system and compared it to a rule-based one. The POMDP system achieves higher concept precision albeit at the cost of longer dialogs, i.e. there is an empirically measurable trade-off between concept precision and dialog length.

Acknowledgments

References

T.H. Bui, M. Poel, A. Nijholt, and J. Zwiers. 2009. A tractable hybrid DDN-POMDP approach to affective dialogue modeling for probabilistic frame-based dialogue systems. *Natural Language Engineering*, 15(2):273–307.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134.

H. Kurniawati, D. Hsu, and W.S. Lee. 2008. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In *Proc. Robotics: Science and Systems*.

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

Matthew Purver. 2006. CLARIE: Handling clarification requests in a dialogue system. *Research on Language and Computation*, 4(2-3):259–288, October.

Sebastian Varges, Giuseppe Riccardi, and Silvia Quarteroni. 2008. Persistent information state in a data-centric architecture. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, Columbus, Ohio.

Sebastian Varges, Giuseppe Riccardi, Silvia Quarteroni, and Alexei V. Ivanov. 2009. The exploration/exploitation trade-off in reinforcement learning for dialogue management. In *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*.

S. Young, M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu. 2010. The Hidden Information State Model: a practical framework for POMDP-based spoken dialogue management. *Computer Speech and Language*, 24:150–174.