

A POMDP approach to Affective Dialogue Modeling

Trung H. BUI¹, Mannes POEL, Anton NIJHOLT, and Job ZWIERS
Human Media Interaction, University of Twente, the Netherlands

Abstract. We propose a novel approach to developing a dialogue model that is able to take into account some aspects of the user's affective state and to act appropriately. Our dialogue model uses a Partially Observable Markov Decision Process approach with observations composed of the observed user's affective state and action. A simple example of route navigation is explained to clarify our approach. The preliminary results showed that: (1) the expected return of the optimal dialogue strategy depends on the correlation between the user's affective state & the user's action and (2) the POMDP dialogue strategy outperforms five other dialogue strategies (the random, three handcrafted and greedy action selection strategies).

Keywords. Dialogue modeling, dialogue management, affective computing, Partially Observable Markov Decision Processes, Dynamic Bayesian Networks

Introduction

We aim to develop dialogue management models which are able to act appropriately by taking into account some aspects of the user's affective state. These models are called *affective dialogue models*. Concretely, our affective dialogue manager processes two main inputs, namely the user's action (e.g., dialogue act) and the user's affective state, and selects the most appropriate system's action based on these inputs and the context. In human-computer dialogue, this work is difficult because the recognition results of the user's action and affective state are ambiguous and uncertain. Furthermore, the user's affective state can change over time. Therefore, an affective dialogue model should take into account both the basic dialogue principles (such as turn-taking and grounding) and the dynamic aspects of the user's affective state.

We found that Partially Observable Markov Decision Processes (POMDPs) are suitable for use in designing these affective dialogue models because of three main reasons. First, the POMDP allows for realistic modeling the user's affective state, user's intention, and other user's hidden states by incorporating them into the state space. Second, the recent dialogue management research for spoken dialogue systems [1,2,3,4] has shown that the POMDP-based dialogue model is able to cope well with the uncertainty that can occur at many levels inside a dialogue system from the speech recognition, natural language understanding to the dialogue management. Third, the transition model and observation model of a POMDP are usually represented by a set

¹ Corresponding Author: Trung H. Bui, Human Media Interaction, Department of Computer Science, University of Twente, Drienerlolaan 5, 7522 NB, Enschede, the Netherlands; Email: t.h.bui@ewi.utwente.nl

of Dynamic Bayesian Networks. These networks are suitable for modeling the user affect and for simulating the behavior of the user.

In this paper, we first introduce a short overview of POMDP and its application to the dialogue management problem. Second, a general affective dialogue model using POMDP is described. Then, we present a simple example to illustrate our ideas and discuss future work.

1. POMDP and Dialogue Management

A POMDP is defined by the tuple $\langle S, A, Z, T, O, R \rangle$, where S is the set of states (of the environment), A is the set of the agent's actions, Z is the set of observations the agent can experience of its environment, T is the transition model, O is the observation model, and R is the reward model (Figure 2a).

In a dialogue management context (Figure 1), the agent is the system (i.e., the dialogue manager) and a part of the POMDP environment represents the user's state. The system uses a state estimator (SE) to compute its internal belief about the user's current state and a policy π to select actions. SE takes as its input the previous belief state, the most recent action and the most recent observation, and returns an updated belief state. The policy π selects actions based on the system's current belief state [5].

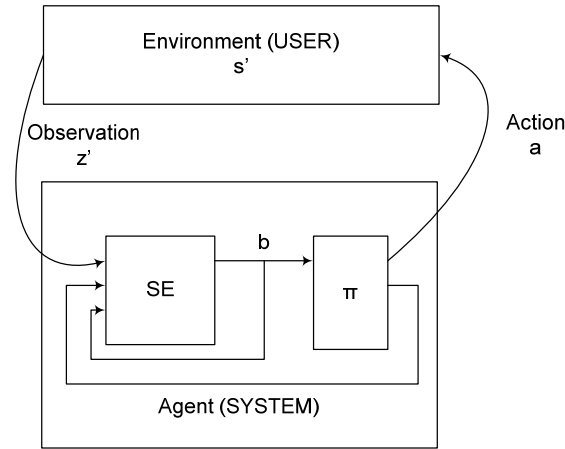


Figure 1. The interaction between the agent (the system) and its environment (the user) in a dialogue management context

Concretely, the system starts with an initial belief state b_0 . At time t , the system belief is b , it selects action a and sends to the user. The user's state changes to s' . State s' is unobservable and the system only gets observation z' . At this moment the system needs to update its belief state b' given knowing b, a, z' :

$$b'(s') = SE(b, a, z') = \alpha O(s', a, z') \sum_{s_i \in S} T(s_i, a, s') b(s_i),$$

where $\alpha = 1/P(z' | a, b)$ is the normalizing constant.

The system task is usually involved in finding the optimal policy (i.e. optimal dialogue strategy) $\pi^* = \text{argmax}_{\pi} E[V^{\pi}(b)]$, where $E[V^{\pi}(b)]$ is the expected value function and

$$V^{\pi}(b) = E_{\pi} \left[\sum_{k=0}^{\infty} \gamma^k \sum_{s_i \in \mathcal{S}} b_{t+k}(s_i) R(s_i, a_{t+k}) \mid b_t = b \right],$$

γ is a discount factor which ensures the sum is finite ($0 \leq \gamma < 1$) and the closer γ to 1 the more effect future rewards have on current system action selection.

The first work that applies POMDP for the dialogue management problem is proposed by Roy and his colleagues for the nursing home robot application [1]. In this application, a flat POMDP is used where the states represent the user's intentions; the observations are the speech utterances from the user; and the actions are the system responses. They show that the POMDP dialogue manager handles well with noisy speech utterances, for example their POMDP-based dialogue manager makes fewer mistakes than an MDP dialogue manager and it automatically adjusts the dialogue policy when the quality of the speech recognition degrades. Zhang's model [2] extends Roy's model in several dimensions: (1) a factored POMDP [6] is deployed for the state and observation sets, (2) the states are composed of the user's intentions and "hidden system states", (3) the observations are the user's utterances and other observation being inferred from lower-level information of speech recognizer, robust parser, and other input modalities. Williams's model [3,4] further extends Zhang's model by adding the state of the dialogue from the perspective of the user to the state set. All these approaches focus on spoken dialogue systems.

Our POMDP dialogue model extends the previous work by integrating the user's affective states and observed user's affective states into state and observation spaces. Furthermore, we propose to verify the performance of the POMDP-based dialogue strategy as well as to simulate the user behavior using a set of parameters (to generate the transition and observation models). The detail of the model is described in the next section.

2. A POMDP Approach to Affective Dialogue Modeling

We select the factored POMDP [6] for representing our affective dialogue model. The state set and observation set are composed of six features. The state set is composed of the user's goal (G_u), the user's affective state (E_u), the user's action (A_u), and the user's grounding state (D_u) (similar to the user's dialogue state described in [3,4]). The observation set is composed of the observed user's action (OA_u) and the observed user's affective state (OE_u). Depending on the complexity of the application's domain, these features can be represented by more specific features. For example, the user's affective state can be encoded by continuous variables such as valence and arousal, and can be represented using a continuous-state POMDP [7]. The observed affective state might be represented by a set of observable effects such as response speech, speech pitch, speech volume, posture, and gesture [8].

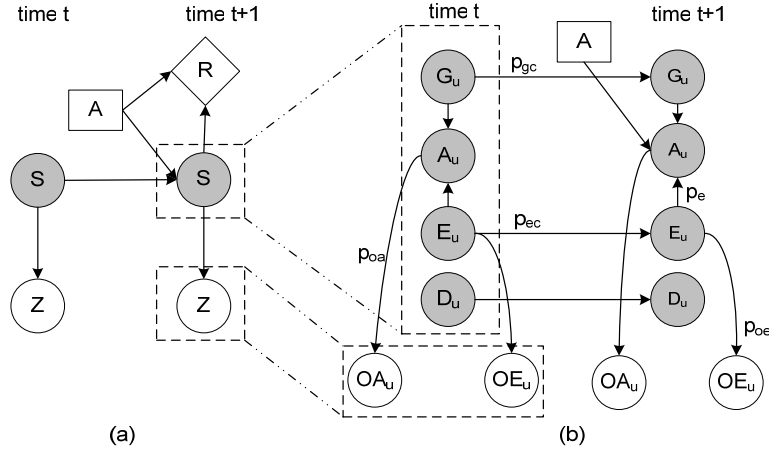


Figure 2. (a) Standard POMDP, (b) Two time-slice of factored POMDP for ADM, where state set S is factored into four features G_u , E_u , A_u , and D_u , the observation set Z is factored into two features OA_u and OE_u

At the moment we are focusing on finite-state discrete-time POMDPs. Figure 2b shows our affective dialogue model (ADM). The features of the state set, action set, observation set, and their correlations form a two time-slice Dynamic Bayesian Network (2TBN). The 2TBN in Figure 2b is built for our route navigation example that will be presented in Section 4. We can easily modify this 2TBN for representing other correlations, for example the correlation between the user's goal and affective state. Parameters p_{gc} , p_{ec} , p_e , p_{oa} , and p_{oe} are used to produce the transition and observation models in case no real data is available, where p_{gc} and p_{ec} are the probabilities that the user's goal and emotion change; p_e is the probability of the user's action error induced by emotion; p_{oa} and p_{oe} are the probabilities of the observed action and observed affective state errors.

The reward model depends on each specific application. Therefore, it is not specified in our general affective dialogue model.

3. Example: Route Navigation in an Unsafe Tunnel

We illustrate our affective dialogue model described in Section 2 by a simulated toy route navigation example. An accident happened in a tunnel. A rescue member (denoted by "the user") is sent to the unsafe part of the tunnel to evacuate some injured victims. Suppose the user is in one of three locations (v_1, v_2, v_3). The user interacts with the system which is located at the operation center. The system is able to produce the route description when knowing the user's current location. Furthermore, the system can detect the user's stressful state (*nostress* or *stress*) and uses this information to act appropriately. In this simple example, the system can *ask* the user about his current location, *confirm* a location provided by the user, *show* route description (*ok*) of a given location, and *stop* the dialogue by connecting the user with the operator.

The POMDP for this problem is represented by $S = \langle G_u \times A_u \times E_u \times D_u \rangle = \langle \{v_1, v_2, v_3\} \times \{\text{answer}(v_1), \text{answer}(v_2), \text{answer}(v_3), \text{yes}, \text{no}\} \times \{\text{stress}, \text{nostress}\} \times \{\text{notstated}, \text{stated}\} \rangle$, $A = \{\text{ask}, \text{confirm}(v_1), \text{confirm}$

$(v_2), confirm(v_3), ok(v_1), ok(v_2), ok(v_3), stop\}, O = \langle OA_u \times OE_u \rangle = \langle \{answer(v_1), answer(v_2), answer(v_3), yes, no\} \times \{stress, nostress\} \rangle$.
The full flat-POMDP model is composed of 61 states (including a special end state), eight actions, and ten observations.

The transition and observation models are generated from the 2TBN (Figure 2b). We assume that the observed user's action only depends on the true user's action (i.e. $P(oa_u | a_u) = (1 - p_{oa})$ if $oa_u = a_u$, otherwise $P(oa_u | a_u) = p_{oa}/4$). The observed user's affective state is computed in a similar way. We use two criteria to specify the reward model, helping the user obtain the correct route description as soon as possible and maintaining the dialogue appropriateness [3]. Concretely, if the system confirms when the user's dialogue state is not stated, the reward is -2, the reward is -5 for action stop, the reward is 10 for action $ok(v_i)$ if $g_u = v_i$, otherwise the reward is -10. The reward for any action taken in the absorbing end state is 0. The reward for any other action is -1. Whenever the system selects actions ok or stop, the current state changes to the end state and the dialogue episode (or dialogue session) is ended.

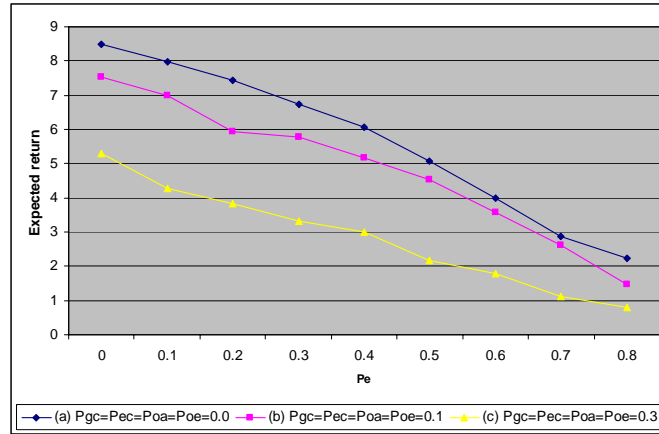


Figure 3. Expected return vs. the user's action error induced by stress p_e

The expected return (i.e. the expected amount of future discounted reward the agent can gather in a large number of steps) of the optimal policy (Figure 3) is computed using the Perseus [9] which is an approximate POMDP algorithm that requires two inputs, a number of belief points and a maximum runtime value. We found 1000 belief points and a runtime of 60 seconds be a good choice for testing our problem. The probability of the user's action error being induced by stress p_e changes from 0 (stress has no influence to the user's action selection) to 0.8 (the user is highly stressed and acts almost randomly). Three lines in Figure 3 are: no observation error ($p_{oa}=p_{oe}=0.0$); low observation error ($p_{oa}=p_{oe}=0.1$); and high observation error ($p_{oa}=p_{oe}=0.3$). All these lines show that the expected return of the optimal policy depends on p_e .

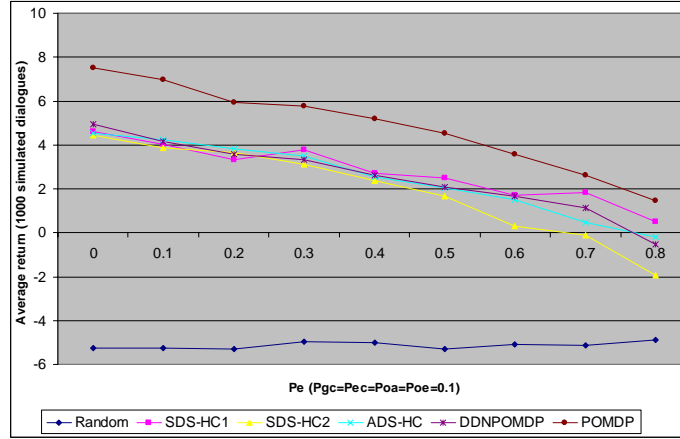


Figure 4. Comparison between POMDP and other dialogue strategies

Figure 4 shows a quantitative comparison between the POMDP dialogue strategy and a set of five other system dialogue strategies (1. select action randomly, 2. first ask and give route description (SDS-HC1), 3. first ask, then confirm, and give route description (SDS-HC2), 4. first ask, then confirm if stress, and give route description (ADS-HC), 5. select the greedy action using a set of 2 time-slice Dynamic Decision Networks (DDNPOMDP)). Strategy 1 aims to show the difference between the random dialogue strategy and other dialogue strategies. Strategies 2 and 3 are considered as the non-affective dialogue strategies since they ignore the user's stress state. Strategy 4 uses commonsense rules to generate the system behavior. Strategy 5 is a special case of the POMDP-based dialogue with the discount factor $\gamma=0$ (this strategy is used in [10,11]). The result is obtained by letting each strategy interact with the simulated user (the simulated user model is constructed as a 2TBN described in Figure 2b). The average return is the average dialogue episode reward the agent receives (1000 dialogue episodes are carried out for each strategy). As expected, the POMDP dialogue strategy outperforms all other strategies, see Figure 4.

4. Conclusions and Future Work

We have presented a POMDP approach to affective dialogue modeling and illustrated our affective dialogue model by a simple example. The 2TBN representation allows integrating the features of states, actions, and observations in a flexible way. We have also shown that even if the observation is perfect, the expected return of the optimal dialogue strategy depends on the correlation between the user's affective state and the user's action. The POMDP dialogue strategy outperforms five other strategies (the random, three handcrafted and greedy action selection strategy). Furthermore, the POMDP dialogue strategy copes well with different types of errors such as speech recognition error [1,2,3,4] and the user's action error being induced by stress as showed in Section 3.

However, solving the POMDP problem (i.e. finding the optimal policy) is computationally expensive. Therefore, all currently developed POMDP dialogue management work is limited to toy frame-based dialogue problems with the size of

several slots [4]. We are currently working with the scaling up issue; especially we focus on the online belief update for real-world dialogue systems.

Acknowledgements

This work is part of the ICIS program (<http://www.icis.decis.nl>). ICIS is sponsored by the Dutch government under contract BSIK 03024.

References

- [1] N. Roy, J. Pineau, and S. Thrun. Spoken Dialogue Management using Probabilistic Reasoning. In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, 2000.
- [2] B. Zhang, Q. Cai, J. Mao, and B. Guo. Spoken Dialog Management as Planning and Acting under Uncertainty. In Proceedings of Eurospeech, 2001.
- [3] 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, 2005.
- [4] J.D. Williams and S. Young. Scaling POMDPs for Dialog Management with Composite Summary Point-based Value Iteration (CSPBVI). In AAAI Workshop on Statistical and Empirical Approaches for Spoken Dialogue Systems, July 2006.
- [5] A.R. Cassandra, L.P. Kaelbling, and M.L. Littman. Acting Optimally in Partially Observable Stochastic Domains. In Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94), volume 2, pages 1023–1028, Seattle, Washington, USA, 1994. AAAI Press/MIT Press.
- [6] C. Boutilier and D. Poole. Computing Optimal Policies for Partially Observable Decision Processes using Compact Representations. In AAAI/IAAI, Vol.2, pages 1168–1175, 1996.
- [7] A. Brooks, A. Makarenko, S. Williams, and H. Durrant-Whyte. Planning in Continuous State Spaces with Parametric POMDPs. In IJCAI Workshop Reasoning with Uncertainty in Robotics, July 2005.
- [8] E. Ball. A Bayesian Heart: Computer Recognition and Simulation of Emotion. In Paolo Petta, Robert Trappl, and Sabine Payr, editors, Emotions in Humans and Artifacts, chapter 11, pages 303–332. The MIT Press, 2003.
- [9] M.T.J. Spaan and N. Vlassis. Perseus: Randomized Point-based Value Iteration for POMDPs. Journal of Artificial Intelligence Research, 24:195–220, 2005.
- [10] T. Paek and E. Horvitz. Conversation as action under uncertainty. In UAI '00: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, pages 455–464, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.
- [11] S. Young, J.D. Williams, J. Schatzmann, M. Stuttle, and K. Weilhammer. The hidden information state approach to dialogue management. Technical Report CUED/F-INFENG/TR.544, University of Cambridge, 2005.