Learning Observation Models for Dialogue POMDPs

Hamid R. Chinaei, Brahim Chaib-draa, Luc Lamontagne

Computer Science and Software Engineering Department, Laval University
Quebec, PQ, Canada
hrchinaei@damas.ift.ulaval.ca {Brahim.Chaib-Draa, Luc.Lamontagne}@ift.ulaval.ca

Abstract. The SmartWheeler project aims at developing an intelligent wheelchair for handicapped people. In this paper, we model the dialogue manager of SmartWheeler in MDP and POMDP frameworks using its collected dialogues. First, we learn the model components of the dialogue MDP based on our previous works. Then, we extend the dialogue MDP to a dialogue POMDP, by proposing two observation models learned from dialogues: one based on learned keywords and the other based on learned intentions. The subsequent keyword POMDP and intention POMDP are compared based on accumulated mean reward in simulation runs. Our experimental results show that the quality of the intention model is significantly higher than the keyword one.

1 Introduction

The SmartWheeler project aims at developing an intelligent wheelchair that minimizes the physical and cognitive load required in steering it [6]. A sample of SmartWheeler dialogues is shown in the left part of Table 1. The first line noted as u_1 shows the true user utterance, that is the one which has been extracted manually from user audio recordings. The following line noted as $\tilde{u}_1$ is the transcribed user utterances by Automatic Speech Recognition (ASR). Finally, the line noted as a_1 shows the performed action in response to the ASR output at the time of collecting dialogues. For each dialogue utterance in ASR output, the machine goal is to estimate the user intention and to perform the best action which satisfies the user intention. The task of dialogue manager is difficult mainly because of ASR output which is not reliable for taking an action (see $\tilde{u}_2$ in Table 1).

The recent literature of spoken dialogue management has used Markov Decision Processes (MDPs) and Partially Observable Markov Decision Processes (POMDPs) for dialogue policy learning [8]. In a previous paper [3], we demonstrated that unsupervised learning methods based on topic modelling can be used for learning user intentions inside dialogues. Then in a follow up work [2], we showed that by using such user intentions as states of dialogue MDPs/POMDPs we can properly model the environment dynamics.

Built on our previous works [3,2], in this paper we first learn a dialogue MDP using SmartWheeler unannotated and noisy dialogues. For background on MDP/POMDP frameworks, the reader is referred to [8]. Then, we introduce two observation models: one is based on learned keywords and the other based on learned user intentions. Our experiments show that the intention model outperforms the keyword model.

Table 1: Left is SmartWheeler sample dialogues. Right is sample results of applying the two observation models on SmartWheeler dialogues (Refer to the end of Section 3 for explanation).

u_1 : turn right	u_1 : turn right
$\tilde{u}_1$: [turn right]	$\tilde{u}_1$: [turn right]
a_1 : (TURN RIGHT)	o_1 : *right (i3o)*
u_2 : turn right a little	u_2 : turn right a little
$\tilde{u}_2$: [10 writer little]	$\tilde{u}_2$: [10 writer little]
a_2 : (PLEASE REPEAT YOUR COMMAND)	o_2 : *confusedObservation (i3o)*
u_3 : turn right a little	u_3 : turn right a little
$\tilde{u}_3$: [turn right to lead a]	$\tilde{u}_3$: [turn right to lead a]
a_3 : (TURN RIGHT A LITTLE)	o_3 : *right (i3o)*

2 Capturing a dialogue MDP Model for SmartWheeler

This section is the application of our previous works [3,2] on the SmartWheeler dialogues. That is, we learn a dialogue MDP model from SmartWheeler unannotated and noisy dialogues [6], using an unsupervised learning method called Hidden Topic Markov Model (HTMM) [4]. In fact, HTMM is an unsupervised hidden topic learning model based on probabilistic latent Dirichlet analysis [1]. The specific property of HTMM consists of adding the Markovian property to topic learned for each sentence of a document. In this way, successive user utterances are more likely to have the same topic.

In a previous work, we showed that HTMM can be adapted for learning user intentions in dialogues [3]. Then, we used the learned user intentions to learn a dialogue MDP [2]. Our experiments in [2] have been performed on SACTI-1 dialogues [7], publicly available at: http://mi.eng.cam.ac.uk/projects/sacti/corpora/.[1]

Based on [2], we learned a dialogue MDP for SmartWheeler. First, we did a quick preprocessing of dialogues to remove stop words such as determiners and auxiliary verbs. Then, we learned the user intentions for SmartWheeler dialogues. Table 2 shows the learned intentions with their four top words. Most of the learned intentions show a specific user *command* desired by the users including: *I1: move forward little*, *I2: move backward little*, *I3: turn right little*, *I4: turn left little*, *I5: follow left wall*, *I6: follow right wall*, *I8: go door*, and *I11: stop*. However, there are some learned intentions which either represent two commands such as *I7: turn degree right/left*, or they loosely represent a command such as: *I9: set speed* and *I10: follow person*.

For learning the model component of the dialogue MDP, we used all the above intentions as states of the MDP. Note that for the intention $I7$, we used it as the state for the command *turn degree right* as the word *right* occurs slightly with higher probability in the intention $I7$. Then, we learned a transition model using a maximum likelihood estimator (c.f. [2]).

The assumed reward model is as follows: reward of +1 for the SmartWheeler performing the right action at each state, and 0 otherwise. Moreover, for the action PLEASE REPEAT YOUR COMMAND occurring in every state the reward is considered as +0.4.

[1] SACTI stands for Simulated ASR Channel Tourist Information.

Table 2: Learned intentions for SmartWheeler.

I1		I2		I3		I4		I5		I6	
forward	0.18029	**backward**	0.38028	**right**	0.20998	**left**	0.18937	**left**	0.24280	**right**	0.27995
move	0.16151	drive	0.33318	turn	0.17108	turn	0.17110	wall	0.22988	wall	0.21239
little	0.11464	little	0.10944	little	0.13348	little	0.13801	follow	0.18831	follow	0.19720
drive	0.08187	top	0.01752	bit	0.07403	right	0.09082	fall	0.03202	left	0.06483
I7		I8		I9		I10		I11			
turn	0.37373	**go**	0.35880	**for**	0.08802	**top**	0.14378	**stop**	0.94263		
degree	0.18612	door	0.28955	word	0.08004	stop	0.13194	stopp	0.02238		
right	0.16560	forward	0.07144	speed	0.05824	follow	0.09850	scott	0.00754		
left	0.16239	backward	0.06523	set	0.05409	person	0.09602	but	0.00275		

3 Learning the Observation Model for the Dialogue POMDP

The model components of a dialogue POMDP includes an observation model. In this section, we learn an observation model for SmartWheeler dialogue POMDP from its dialogues. In order to be able to apply POMDPs in real domains, we need to reduce the number of observations significantly. This is because the time complexity for learning the optimal strategy of a POMDP is double exponential to the number of observations [5]. In SmartWheeler dialogues, there exist around 400 words which can potentially be used as observations. In this case, approximating the optimal policy of such a POMDP is intractable. Thus, we reduce the number of observations in two ways and learn two observation models: one is keyword model and the other is intention.

Keyword observation model: For each state, this model uses a keyword which best represents the state. In fact, we use the 1-top word of each state learned in Table 2 as observations (highlited words). That is, observations are O = *forward, backward, right, left, turn, go, for, top, stop*. Note that states 3 and 6 share the same observation keyword. This is also the case for states 4 and 5. This problem can be avoided by using for instance the two top words of each state as observations, however this increase the size of observations which is not appealing to us. Thus, an auxiliary observation noted as *confusedObservation* is used for the case where none of the learned keyword observations occur in a recognized user utterance. Moreover, if an utterance includes more than one keyword as observation, then *confusedObservation* is also used as the observation. For the keyword observation model, we define a maximum likelihood observation model:

$$\Omega(o',a,s') = Pr(o'|a,s') = \frac{Count(a,s',o')}{Count(a,s')}$$

To make a more robust observation model, we apply smoothing to the maximum likelihood observation model, for instance δ smoothing where $0 \leq \delta \leq 1$. We set δ to 1:

$$\Omega(o',a,s') = Pr(o'|a,s') = \frac{Count(a,s',o') + 1}{Count(a,s') + K'}$$

where $K' = |A||S||O|$.

Intention observation model: Given recognized user utterance $\tilde{u} = [w_1, \ldots, w_n]$, the highest probable underlying intention is used as POMDP observation. That is:

$$o = \arg\max_z \prod_i \beta_{w_i z} \tag{1}$$

where $\beta_{w_i z}$ is the vector for probability of each word w given intention z. Notice that for the intention model, each state itself is the observation. Then, the set of observations is equivalent to the set of intentions. For instance, for SmartWheeler the observations are O = *i1o, i2o, i3o, i4o, i5o, i6o, i7o, i8o, i9o, i10o, i11o* respectively for states *i1, i2, i3, i4, i5, i6, i7, i8, i9, i10, i11*. Similar to the keyword model, the intention observation model can be defined as:

$$\Omega(o', a, s') = Pr(o'|a, s') = \frac{Count(a, s', o')}{Count(a, s')}$$

Note that in the intention observation model, we essentially end up with a MDP model. This is because we use the highest maximum likelihood intention as state and we use the highest maximum likelihood intention as observation as well. So, we end up with a deterministic observation model, which makes the framework a MDP. However, we can use sort of smoothing to allow a small probability for other observations than the observation corresponding to the current state. In the experiments section, we use the intention model without smoothing as intention MDP.

Additionally, we can estimate the intention observation model using recognized utterances $\tilde{u}$ inside the training dialogue d using vector β_{wz} and θ_z where θ is a local vector for each dialogue d, and retains the probability of intentions in the dialogue d (c.f. [3]). Assume that we want to estimate $Pr(o')$, then we have:

$$\begin{aligned} Pr(o') &= \sum_w Pr(w, o') \\ &= \sum_w Pr(w|o') Pr(o') \\ &= \sum_w \beta_{wo'} \theta_{o'} \end{aligned} \tag{2}$$

where o' is drawn from Equation 1. To estimate $Pr(o'|a, s')$, the multiplication in Equation 2 is performed only after visiting the action state pair (a, s'). As such, we use this calculation to learn the intention observation model. In the experiment section, the dialogue POMDP with the intention observation model is noted as intention POMDP.

The right of Table 1 shows the sample dialogue from SmartWheeler after learning the two observation models using the dialogues. The line o_1 is the observation for the recognized utterance $\tilde{u}_1$. If keyword observation model is used the observation will be *right*, but if intention observation model is used then the observation inside parenthesis is used, i.e., *i3o*. In fact, *i3o* is an observation with high probability for *i3* state, and with low probability for the rest of states. Notice however that in o_2 for the case of keyword observation, *confusedObservation* is learned. This is because for the keyword model, none of the keyword observations occurs in the recognized utterance $\tilde{u}_2$. But, the intention observation is *i3o* which is interestingly the same as the intention observation in o_1.

Table 3: Performance of intention POMDP vs. keyword POMDP.

Used Framework	Mean Reward	Conf95Min	Conf95Max
SmartWheeler intention MDP	10.000	10.000	10.000
SmartWheeler intention POMDP	8.914	8.904	8.922
SmartWheeler keyword POMDP	4.784	4.767	4.802
SACTI-1 intention MDP	186.413	169.074	201.498
SACTI-1 intention POMDP	197.769	180.977	215.02
SACTI-1 keyword POMDP	135.843	116.608	153.464

4 Experiments

In SmartWheeler, there are eight dialogues with healthy users and nine dialogues with target users of SmartWheeler. The dialogues with target users, who are the elderly, are somehow more noisy than the ones with healthy users. However, we used all the data for healthy and target users in order to perform the experiments on a larger amount of data. The average word error rate (WER) is 13.88% for dialogues of healthy users and 19.43% for dialogues of target users. In order to perform our experiments on a larger amount of data, we used all the dialogues for healthy and target users. In total, there are 2853 user utterances and 422 distinct words in the used dialogues. We performed the methods proposed in this paper to learn dialogue POMDP models for SmartWheeler. The SmartWheeler dialogue POMDP models consist of 11 states, 24 actions and 10 observations if keyword observation model is used, and 11 observations in the case of intention observation model.

Moreover, we performed the same set of experiments on SACTI-1 dialogues [7]. We learned intention MDP, intention POMDP, and keyword POMDP for SACTI-1 dialogues. For details on SACTI-1 learned dialogue MDP, the reader is referred to [2]. The SACTI-1 dialogue POMDP models consist of 5 states, 14 actions and 5 observations for both keyword observation and intention observation models.

We solved our POMDP models, using ZMDP software available online at: http:// www.cs.cmu.edu/ trey/zmdp/. We set a uniform distribution on states, and set the discount factor to 90%. Table 3 shows the comparison of intention POMDP, keyword POMDP, and intention MDP for the two domains based on accumulated mean reward in 1000 simulation runs by ZMDP software. In Table 3, *Conf95Min* and *Conf95Max* are respectively the minimum 95% confidence and the maximum 95% confidence of accumulated mean reward. The table shows that intention POMDP accumulates strongly higher mean reward than the keyword POMDP.

Moreover, the results show that the intention MDP performance is comparable with intention POMDP. In particular, for SmartWheeler intention MDP the mean reward is slightly higher than that of the intention POMDP. The reason is that the intention MDP is actually an intention POMDP with a "deterministic" observation model. So, for example, when the MDP receives intention $i1o$ as observation it assumes $i1$ as its state with 100% probability. However, in the intention POMDP $i1o$ can bring uniform distribution for the observation model, i.e., no information. This is particularly the case when $i1o$ has not been observed after performing some actions in training dialogues.

5 Conclusion and Future Work

In this paper, first we learned the model components of dialogue MDP for SmartWheeler, using its noisy dialogues. Moreover, we introduced two observation models to form a keyword dialogue POMDP and an intention dialogue POMDP. Furthermore, we developed the intention dialogue MDP, the keyword dialogue POMDP, and the intention dialogue POMDP using SACTI-1 dialogues. Finally, we evaluated the models of the two domains and we saw that the intention POMDP performed strongly higher than the keyword POMDP. Moreover, the intention MDP performance was comparable to the the intention POMDP.

In the future, we would like to further improve the performance of intention POMDP by increasing the number of observations, for instance by considering both the learned keywords and learned intention observations. Moreover, the intention MDP can be transformed to a POMDP by making a non-deterministic observation model, for instance by making the observation model of intention MDP smoothed. Lastly, we would like to verify the performance of the enhanced models based on other criteria for example to compare policy of the learned models with respect to the human expert actions.

6 Acknowledgement

The authors would like to thank Professor Joelle Pineau from Computer Science at McGill University for kindly providing us with SmartWheeler dataset. The dataset has been collected with contributions from researchers at McGill University, Ecole Polytechnique de Montreal, Universite de Montreal, and the Centres de readaptation Lucie-Bruneau and Constance-Lethbridge.

References

1. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
2. Hamid R. Chinaei and Brahim Chaib-draa. Learning Dialogue POMDP Models from Data. *Advances in Artificial Intelligence*, pages 86–91, 2011.
3. Hamid R. Chinaei, Brahim Chaib-draa, and Luc Lamontagne. Application of Hidden Topic Markov Models on Spoken Dialogue Systems. *Agents and Artificial Intelligence*, pages 151–163, 2010.
4. Amit Gruber, Michal Rosen-Zvi, and Yair Weiss. Hidden Topic Markov Models. In *Artificial Intelligence and Statistics (AISTATS)*, San Juan, Puerto Rico, 2007.
5. Joelle Pineau, Geoffrey Gordon, and Sebastian Thrun. Point-based Value Iteration: An Anytime Algorithm for POMDPs. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1025 – 1032, Acapulco, Mexico, August 2003.
6. Joelle Pineau, Robert West, Amin Atrash, Julien Villemure, and Francois Routhier. On the Feasibility of Using a Standardized Test for Evaluating a Speech-Controlled Smart Wheelchair. *International Journal of Intelligent Control and Systems*, 16(2):124–131, 2011.
7. Jason D. Williams and Steve Young. The SACTI-1 Corpus: Guide for Research Users. Cambridge University Department of Engineering. Technical report, 2005.
8. Jason D. Williams and Steve Young. Partially Observable Markov Decision Processes for Spoken Dialog Systems. *Computer Speech and Language*, 21:393–422, 2007.