A NONPARAMETRIC BAYESIAN APPROACH TO LEARNING MULTIMODAL INTERACTION MANAGEMENT

Zhuoran Wang and Oliver Lemon

MACS, Heriot-Watt University, Edinburgh, UK

ABSTRACT

Managing multimodal interactions between humans and computer systems requires a combination of state estimation based on multiple observation streams, and optimisation of time-dependent action selection. Previous work using partially observable Markov decision processes (POMDPs) for multimodal interaction has focused on simple turn-based systems. However, state persistence and implicit state transitions are frequent in real-world multimodal interactions. These phenomena cannot be fully modelled using turn-based systems, where the timing of system actions is a non-trivial issue. In addition, in prior work the POMDP parameterisation has been either hand-coded or learned from labelled data, which requires significant domain-specific knowledge and is labor-consuming. We therefore propose a nonparametric Bayesian method to automatically infer the (distributional) representations of POMDP states for multimodal interactive systems, without using any domain knowledge. We develop an extended version of the infinite POMDP method, to better address state persistence, implicit transition, and timing issues observed in real data. The main contribution is a "sticky" infinite POMDP model that is biased towards self-transitions. The performance of the proposed unsupervised approach is evaluated based on both artificially synthesised data and a manually transcribed and annotated human-human interaction corpus. We show statistically significant improvements (e.g. in ability of the planner to recall human bartender actions) over a supervised POMDP method.

Index Terms— Multimodal Interaction, HDP, POMDP

1. INTRODUCTION

In order to address planning problems under uncertainty, partially observable Markov decision process (POMDP) models have recently been demonstrated in several successful applications in spoken dialogue systems (SDS) [1, 2] and for human-robot interaction (HRI) [3]. However, existing POMDP-style interactive systems are usually turn-based, where belief state updates will only be considered following explicit system actions, and without considering implicit user state transitions. This simplification underestimates the complexity of multimodal communication, where humans can generate important state transitions without intervening system actions. Figure 1 shows the kind of multiple input streams that we observe in human multimodal communication, where we take a robot bartender as our example application.

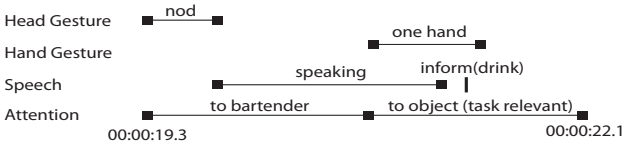

Fig. 1. Example observations from the human-human interaction corpus.

A human agent could start from a miscellaneous state without attempting to be involved in an interaction with the robot bartender. Then at some time, it may decide to request the bartender's attention to start an interaction round (e.g. to order a drink). In this case, there must be a mechanism in the system to allow (belief) state updates even though no explicit system action is performed. Traditional approaches could rely on predefined trigger events to handle this situation, but in HRI the observations are multimodal, i.e. as well as speech inputs, a user's intention could also be realised by various nonverbal behaviours, such as hand gestures, body postures, facial expressions, gaze, etc, which the robot controller would receive from a vision system frame-by-frame as an event stream. Unlike in traditional speech-based dialogue systems, where the boundary of a user state can be identified by observing silence durations above a certain length in the speech input (i.e. the"end of speech" signal), the trigger events to segment user states in such multimodal observations are more difficult to recognise. For example, the user state "request for attention" can be realised by "looking at the bartender" until a system response is received, which means that the state duration varies and there may not be an obvious boundary for such a state to trigger the system's action planner. Hence, the timing of system actions is also a non-trivial issue in such a real-world HRI task.

In addition, previous POMDP parameterisations have been either hand-coded [1] or learned from labelled data [2], both of which methods rely on complicated predefined semantic correlations (in terms of probability dependencies)

among the states, observations and system actions. The construction of such a system requires significant domain-specific knowledge and is labor-consuming. Moreover, when more complex interactions are involved (e.g. in a real-world HRI system as described above), it might not be easy to develop reasonable assignments of such semantic correlations.

Therefore, in this paper we propose a nonparametric Bayesian method to automatically infer the (distributional) representations of POMDP states for complex interactive systems, without using any domain knowledge (see Section 3). Our work essentially follows the *infinite POMDP* (iPOMDP) model [4]. Firstly, we discuss its application in modelling multimodal observations. Then we propose an extended version, in order to better address the state persistence, implicit state transition, and timing issues. Our main contribution is a *"sticky" iPOMDP* that is biased towards self-transitions for implicit *null* system actions. The proposed approach works on frame-based observations and offers a unified framework to jointly solve the state persistence, implicit transition, and time-dependent action selection problems.

We evaluate the state inference performance of the sticky iPOMDP in two ways. First we use artificially synthesised data. After this, we evaluate its planning effects based on a manually transcribed and annotated human-human interaction corpus, by comparing our system's action selection outcomes against the true human actions. Promising results are obtained in the both experimental settings. In the second experiment, the proposed method selects system actions agreeing with the true human actions 74% of the time, and its correct actions also tend to be produced at the timing close to but reasonably faster than the original human decisions.

The reminder of this paper is organised as follows. Section 2 briefly reviews some fundamental knowledge about POMDPs. Section 3 discusses the iPOMDP and how it can be extended to handle several multimodal interaction problems, which leads to the "sticky" iPOMDP. We explain the inference and planning algorithms for the proposed method in Section 4 and Section 5, respectively. Our experimental results are reported in Section 6. Previous work related to the problems of interest in this paper is discussed in Section 7. Finally, we conclude in Section 8.

2. POMDP BASICS

A POMDP is a tuple $\{S, A, O, T, \Omega, R, \eta\}$, where the components are defined as follows. S, A and O are the sets of states, actions and observations respectively. The transition function $T(s'|s,a)$ defines the conditional probability of transiting from state $s \in S$ to state $s' \in S$ after taking action $a \in A$. The observation function $\Omega(y|s,a)$ gives the probability of the occurrence of observation $y \in O$ in state s after taking action a. $R(s,a)$ is the reward function specifying the immediate reward of a state-action pair. Whilst, $0 \leq \eta \leq 1$ is a discount factor. In this paper, we will focus on POMDPs with discrete state and action spaces, but the observations can take either discrete or continuous values.

A standard POMDP operates as follows. At each time step, the system is in an unobservable state s, for which only an observation y can be received. A distribution over all possible states is therefore maintained, called a belief state, denoted by b, where the probability of the system being in state s is $b(s)$. Based on the current belief state, the system selects an action a, receives a reward $R(s,a)$ and transits to a new (unobservable) state s' where it receives an observation y'. Then the belief state is updated to b' based on y' and a as:

$$b'(s') = \frac{1}{Z(a,y')}\Omega(y'|s',a)\sum_{s} T(s'|a,s)b(s) \tag{1}$$

where $Z(a,y') = \sum_{s'}\Omega(y'|s',a)\sum_{s}T(s'|a,s)b(s)$ is a normalisation factor. Let π be a policy that maps each belief state b to an action $a = \pi(b)$. The value function of π is then defined to be the expected sum of discounted rewards, as:

$$V^{\pi}(b) = R(b,\pi(b)) + \eta\sum_{b'} T(b'|b,\pi(b))V^{\pi}(b') \tag{2}$$

Now there are two inevitable problems in constructing a POMDP: how to estimate the model parameters T, Ω and R; and how to seek an optimal policy π maximising $V_{\pi}(b)$ for every b, given T, Ω and R. Focusing on multimodal and time-dependent interactions, we address these two questions respectively in the following sections of this paper.

3. THE INFINITE POMDP AND ITS EXTENSIONS

The infinite hidden Markov model (iHMM) as an application of the hierarchical Dirichlet process (HDP) has been proven to be a powerful tool for inferring generative models from sequential data [5]. The iPOMDP derived in [4] directly extends the iHMM, of which we give a brief review for the convenience of future discussions.

3.1. Review of the iPOMDP

An iPOMDP utilises an HDP to define a prior over POMDPs as follows. To generate a model from the prior, we:

- Draw the state distribution prior $\beta \sim \mathrm{GEM}(\lambda)$
- For each state-action pair (s,a):
  - Draw a transition parameter $\mathrm{T}_{s,a} \sim \mathrm{DP}(\alpha,\beta)$
  - Draw a reward parameter $\Theta_{s,a} \sim H_R$
- For each state s:
  - Draw an observation parameter[1] $\Omega_s \sim H_{\Omega}$

[1]Note that here, we assume that the observation function $\Omega(y|s)$ is independent of the previous system action a [6]. This is because if the original definition $\Omega(y|s,a)$ is utilised, the HDP tends to cluster state-action pairs based on their observations, according to our experiments, which can confuse the planning process.

where H_Ω and H_R are the respective prior distributions for Ω_s and $\Theta_{s,a}$, $\mathrm{GEM}(\lambda)$ stands for the stick-breaking construction procedure with a concentration parameter λ (see e.g. [5]), and $\mathrm{DP}(\alpha, \beta)$ is a Dirichlet process (DP) with a concentration parameter α and the base probability measure β.

Then for an interaction sequence consisting of a trajectory of N observations and actions $\{(y_1, a_1), (y_2, a_2), \ldots, (y_N, a_N)\}$, the generative process is defined as:

- For $i = 1, \ldots, N$:
  - Draw a transition $s_i \sim P_{\mathrm{T}}(\cdot|s_{i-1}, a_i)$
  - Draw an emission $y_i \sim P_\Omega(\cdot|s_i)$
  - Draw a reward $r_i \sim P_\Theta(\cdot|s_i, a_{i+1})$

where the reward function $R(s, a)$ is rewritten as $P_\Theta(r|s, a)$, a conditional distribution describing the probability of observing reward r on state-action pair (s, a).

3.2. Multimodal Observations

To adapt the iPOMDP to the multimodal case, one essential challenge is to construct a joint distribution function for the multiple channels of observations. Such observations are usually presented using different representations. For example, a common representation for speech inputs is an n-best list of parsed dialogue acts with semantics, each with a normalised confidence score [1, 2]. However, gesture and facial expression recognisers tend to provide continuous (frame-based in practice) streams of events with discrete values. On the other hand, the gaze and position (3D coordinates) information of a human agent can be in the form of streams with continuous values. Therefore, we have to define a distribution for every observation channel and let the joint observation distribution be their tensor products. Concretely, assume that there are K channels of observations. For each frame of the input $z = y^{1:K}$, the observation probability will be computed by $P_\Omega(z|s) = \prod_{k=1}^{K} P_{\Omega^k}(y^k|s)$. Hence, in the iPOMDP, we will have the observation generation process as:

- For each state s:
  - For each channel k:
    - Draw an observation parameter $\Omega_s^k \sim H_\Omega^k$

Different distributions $P_{\Omega^k}(\cdot|s)$ and priors H_Ω^k can be used for different types of channels. We start from the simplest observation type, the binary indicators. The Bernoulli distribution that has a conjugate Beta prior is a natural choice to model such observations. When generalised to the multivariate case, it also models the occurrences of events in n-best lists. Furthermore, one can assume that the normalised confidence score associated with each event is drawn from a separate Beta distribution. Although Beta likelihood does not have a conjugate prior, one can either employ Metropolis-Hastings algorithms to seek a target posterior [7], or perform a Bernoulli trial to choose one of its two parameters to be 1 and apply a conjugate Gamma prior for the other one [8]. Finally, to model streams of events, multinomial or multivariate Gaussians can be used to draw the respective discrete or continuous observation in each frame, for which conjugate priors are the well-known Dirichlet distribution and Normal-Inverse-Wishart distribution, respectively.

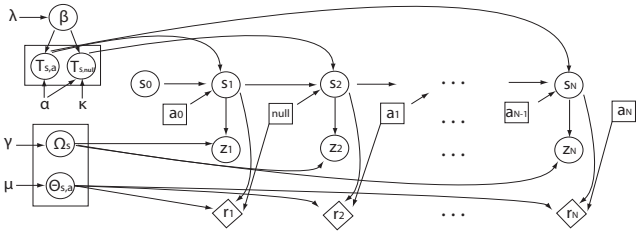

Fig. 2. Graphical representation of the sticky iPOMDP.

3.3. The "Sticky" iPOMDP

Regarding the previous discussions, state persistence and implicit state transitions commonly exist in real-world multimodal interactions. A natural strategy addressing the timing of system actions is to model a POMDP that allows the system to select an action (including a *null* action, or a *wait* action in other words) at every unit timestamp. This requires the iPOMDP to infer the hidden user states frame-by-frame.

However, as an HDP, the iPOMDP tends to cluster observations into states, which suggests that slight changes among the observations over time might result in them being clustered to many different states. Therefore, if directly applied here, the standard iPOMDP may experience unexpected fast state switches (see Section 6). To better model state persistence, we give a bias to self state transitions when the system performs a *null* action, by drawing for each state s:

$$\mathrm{T}_{s,null} \sim \mathrm{DP}(\alpha + \kappa, \frac{\alpha\beta + \delta_s\kappa}{\alpha + \kappa}) \tag{3}$$

where $\kappa > 0$ is a hyperparameter to weight the bias, and δ_s is a Kronecker indicator. $(\alpha\beta + \delta_s\kappa)$ means that an amount κ is added to the s-indexed element in $\alpha\beta$. The idea directly follows the sticky HDP-HMM [9], but in the iPOMDP context, self-transitions are only biased for *null* actions and should be eliminated for explicit system actions. Note here, the self-transition bias assumes that the user tends stay in the same state if no system action is explicitly performed, however the probabilities for implicit state transitions are still preserved. On the other hand, when a system performs an action, the user normally would not remain in the same state as the previous one. So self-transitions should be eliminated, which can be done by setting $P_{\mathrm{T}}(s|s, a) = 0$ and renormalise $P_{\mathrm{T}}(\cdot|s, a)$ every time a transition distribution is drawn. Figure 2 illustrates a graphical representation of the sticky iPOMDP.

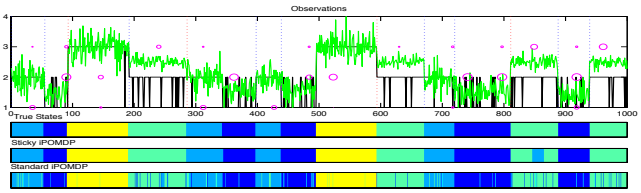

Fig. 3. State inference on synthetic data: Black solid lines are generated randomly from underlying multinomials to represent stream events with discrete values; Green solid lines are generated from Gaussians representing a stream of continuous observations; Magenta circles are generated from multivariate Bernoulli distributions and simulate n-best lists of recognised events (e.g. dialogue acts from parsed speech recogniser hypotheses) with their radii generated from corresponding Beta distributions representing confidence scores.

4. INFERENCE ALGORITHM

The weak-limit sampler used in [9] is adapted here to solve inference problems for our sticky iPOMDP. Firstly, we approximate the HDP transition prior by a finite L-dimensional Dirichlet distribution, which is called a degree L weak-limit approximation. Then the HMM forward-backward procedure can be employed to jointly sample the state sequence $s_{1:N}$ given the observations sequence $z_{1:N}$ and action sequence $a_{1:N}$, which will significantly improve the inference speed. Moreover, it provides a tractable finite state space for the POMDP planning process.

Concretely, assume that we are at position $1 < i \leq N$ within an action and observation sequence $\{(z_1, a_1), (z_2, a_2), \ldots, (z_N, a_N)\}$. For each $s \in S$, the backward message $m_{i,i-1}(s)$ passed from i to $i-1$ can be computed by:

$$m_{i,i-1}(s) = \sum_{s' \in S} m_{i+1,i}(s') P_{\mathrm{T}}(s'|s, a_i) P_{\Omega}(z_i|s') \qquad (4)$$

where we define $m_{N,N+1}(s) = 1$ for all $s \in S$. Based on the backward messages, we can work sequentially forward to sample the state assignments, as:

$$s_j \sim \sum_{s \in S} P_{\Omega}(z_j|s) P_{\mathrm{T}}(s|s_{j-1}, a_j) m_{j+1,j}(s) \delta(s, s_j) \qquad (5)$$

where $\delta(\cdot, \cdot)$ denotes the Kronecker indicator. After this, we can sample the auxiliary variables to update the global transition distribution, and resample new transition distributions for each state. Finally, conditioning on those sampled states, the posterior parameters for observations and rewards can be sampled. Note that, since self-transitions are ruled out for explicit system actions in the sticky iPOMDP, geometric auxiliary variables needs to be sampled for transitions conditioned on explicit actions to complete the data to allow conjugate inference, as suggested in [10], whereas binomial override auxiliary variables similar to [9] are required for transition parameters depending on the *null* actions.

5. PLANNING

Due to the possibly infinitely large (continuous) observation space together with the model uncertainty raised by HDP, seeking an optimal policy via value iteration techniques [11, 12] is difficult in our case. Hence, in this work we employ a classic forward search method to solve our POMDPs as proposed in [4] for the standard iPOMDPs, where we sample a set of models to compute a weighted-averaged Q-value, and only maintain a finite set of observations generated by Monte-Carlo sampling at each node of the search tree.

6. EXPERIMENTS

We evaluate the performance of state inference of the proposed sticky iPOMDP as well as its actual planning effects in comparison with the standard iPOMDP based on a synthetic data sequence as well as a transcribed and manually annotated human-human interaction corpus [13]. In addition, on the second data set, a supervised learning based POMDP model is also trained as a baseline system.

6.1. State Inference on Synthetic Data

Figure 3 illustrates the state inference performance of the sticky iPOMDP in comparison with the standard iPOMDP on an artificially synthesised data sequence. The sequence consists of 1000 data points generated based on 4 hidden states, 2 explicit actions (red and blue dash lines), and 3 multimodal observation channels.

Note that two implicit state transitions happen here, between point 300 and point 400 and around point 500. The initial results suggest that the sticky iPOMDP achieves a better alignment between the inferred and true states than the standard iPOMDP, whereas the latter suffers from frequent state switches, as can be seen in Figure 3.

6.2. Planning on Transcribed Corpus

The planning performance of the proposed model was also evaluated based on a human-human interaction corpus [13], which contains 50 interaction sequences between customers and a bartender, manually transcribed and annotated from 50 video clips recorded in a real German bar. There are 6 user states, 4 explicit system (bartender) actions, and 4 observation channels in the data. The observation channels consist of speech, hand gestures, head gestures, and attention information. The last three types of observations are all in the form of streams of discrete events. However, to simulate the situation one can normally expect in an HRI setting with vision systems and a standard speech recogniser, we split the speech channel into two sub-channels as follows. Firstly, when a customer starts talking, the system will keep observing a *speaking* event. After this, only in the last frame of the speaking

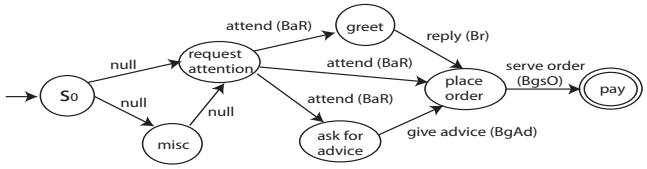

Fig. 4. State transitions in the transcribed corpus: Vertices are user/customer states and edges are bartender/system actions. We define a common start state s_0 for all interaction sequences, and force them to finish at *pay* states.

stream, a dialogue act will be received. Note here, that since the data is manually transcribed, there is no uncertainty in the observations. However, the uncertainty comes from the state inference. Without losing generality, noisy observations can be fed into our models in real HRI applications. The interactions are illustrated in Figure 4. Note here, the true user states are annotated in this corpus, but this information is reserved when training our models, and is only used for training a baseline system and designing the evaluation metric.

The evaluation metric is designed as follows. We conduct a leave-one-out test for each interaction sequence. In each state, we feed the observations frame-by-frame from the beginning of that state into a model trained on the remaining 49 examples, until an expected action is output by the planner or the state finishes. Then we move to the next state and repeat this procedure. Note here, due to the limited data (i.e. no data on user reactions to unusual bartender actions) we assume that if the system outputs an incorrect action, the user will just ignore that action and remain in the same state continuing what he/she is doing. This is by necessity a preliminary simulation of real users, since we only have an offline corpus available.

We take the transcription chunk corresponding to every 0.1s video clip as a frame to generate the training data, based on which the sticky and standard iPOMDPs are trained. Degree 50 weak-limit approximations are utilised as described in Section 4, and the sampling procedures are run for 200 iterations. After this, a forward search Monte-Carlo planner is employed for each of the two iPOMDPs, where 5 POMDP models are sampled from the posterior, and the search depth and number of (joint multimodal) observations sampled for each search node are set to 3 and 10 respectively.

In addition, the reward distributions in both cases are constructed as follows. Firstly, a three-dimensional Dirichlet distribution with the concentration parameter [1, 0.01, 0.01] is used as the prior for all (s, a) pairs, where the three corners of the simplex correspond to reward values -10, 0 and 10 respectively. Then after the state inference procedure, an observed (s, a) is assigned a reward 0 if $a = null$ and 10 otherwise. Hence, the distributions $P_\Theta(\cdot|s, a)$ drawn from the posterior will tend to reward the explicit state-action pairs that have been seen during the sampling, penalise those unseen state-action combinations, and stay neutral for *null* actions.

We also train a baseline POMDP model using the annotations in our corpus, where the transition probabilities and observation probabilities are estimated in a supervised manner (frequency-based probabilities with add-one smoothing), and the reward function is designed by simply assigning a positive reward 10 to the explicit state-action pairs observed in the corpus, 0 reward to state-*null*-action pairs, and a negative reward -10 to those unseen state-action combinations. Leave-one-out test is also performed for the baseline model, and in each round its policy is optimised offline using Perseus [12]. Note that, the supervised model will naturally achieve a bias on self state transitions, as it is trained on frame-based state sequences, where state persistences are frequently seen.

We measure four quantities: *Precision* – the percentage of the planned explicit actions agreeing with the human actions, *Recall* – the percentage of the human actions recovered by the planner, *F-score* – the harmonic mean of precision and recall, and *Relative Timing* - the average amount of time in seconds by which those correctly planned actions are ahead of or behind the human actions (note that human action timing may not be optimal).

The results for the first three quantities are shown in Table 1. It can be found that all the models can produce satisfactory plans highly agreeing with the human bartenders' decisions. However, interestingly and surprisingly, the two unsupervised methods achieve precisions comparable to the supervised baseline with optimised policies, and even slightly outperform the supervised baseline according to the F-score. (The results are statistically significant based on approximate randomisation tests [14], where the significance level $p < 0.01$.) This suggests that the states inferred by the iPOMDPs can capture more information than the rather general state annotations. In addition, the sticky iPOMDP works better than the standard iPOMDP, which is due to the bias on self-transitions allowing the probabilities to propagate more properly during belief updates, since that is the only difference between the two models. The fourth quantity tends to be action-dependent, hence we evaluate it for each action separately and show the results in Table 2, where the findings indicate that the timing decisions of our methods are also close to the human bartender's action timing, with some actions (especially *BaR*) selected reasonably faster than the human bartender.

7. RELATED WORK

Time-dependent POMDP planning problems have previously been discussed in [3], where the timing issue was solved by explicitly defining a time-indexed state representation in the POMDP. We argue that our sticky iPOMDP offers a more flexible solution in comparison with his work, due to its potential ability in modelling large state duration variance.

Bohus and Horvitz [15, 16] introduced a multimodal dialogue system that utilises supervised learning techniques to classify multiparty engagement states and make correspond-

Model	Precision	Recall	F-score
Sticky iPOMDP	0.74	0.96	**0.84**
Standard iPOMDP	0.71	**0.98**	0.82
Supervised POMDP	**0.78**	0.85	0.81

Table 1. Accuracy of planning evaluated based on transcribed real-world interaction sequences.

	BaR	*Br*	*BgAd*	*BgsO*
Sticky	-1.6±1.4	-0.7±1.4	0	-0.1±0.3
Standard	-1.3±1.5	-0.6±1.3	+0.1±0.0	-0.1±0.3
Supervised	-1.5±1.6	-0.6±1.3	0	-0.1±0.2

Table 2. Relative timing (s) of planning evaluated based on transcribed real-world interaction sequences.

ing decisions. In their work, the timing issue is handled by modelling state transitions based on a dynamic graphical model with explicitly defined variable dependencies among the features for engagement states and observations. A remarkable advantage of their approach is that the model can be trained based on automatically collected observations and state labels without explicit developer supervision. To address several real-world situations very similar to the discussions in [15, 16], this paper attempts an alternative that employs recent advances in unsupervised machine learning, where no state labels or domain-specific knowledge is required at all.

8. CONCLUSION

This paper introduces a nonparametric Bayesian POMDP model to jointly solve several issues that commonly exist in real-world multimodal HRI tasks, but have rarely been discussed in previous work. The main advantages of the proposed technique over previous approaches using POMDPs are its abilities in modelling state persistence and implicit transitions, in seeking proper action timing, and in employing unsupervised learning.

Satisfactory results are obtained in evaluations for both the state inference and the planning procedures, where the proposed method selects system actions agreeing with the true human actions 74% of the time at reasonable timing. Moreover, this unsupervised technique outperforms a supervised model at statistically significant levels, which demonstrates its feasibility and potential application in solving multimodal interaction problems. The issues of its application in physical robotic systems and different domains are left open at this stage, but will be investigated in our future research.

9. ACKNOWLEDGMENTS

The research leading to these results was supported by the EC FP7 projects JAMES (ref. 270435) and Spacebook (ref. 270019).

10. REFERENCES

[1] J. Williams and S. Young, "Partially observable Markov decision processes for spoken dialog systems," *Computer Speech and Language*, 2007.

[2] B. Thomson and S. Young, "Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems," *Computer Speech and Language*, 2010.

[3] F. Broz, *Planning for Human-Robot Interaction: Representing Time and Human Intention*, Ph.D. thesis, Carnegie Mellon University, 2008.

[4] F. Doshi-Velez, "The infinite partially observable Markov decision process," in *NIPS*, 2009.

[5] Y. W. Teh, M. I. Jordan, M. Beal, and D. Blei, "Hierarchical Dirichlet processes," *Journal of the American Statistical Association*, 2004.

[6] A. Atrash and J. Pineau, "Efficient planning and tracking in POMDPs with large observation spaces," in *AAAI Workshop on Statistical and Empirical Approaches for Spoken Dialogue Systems*, 2006.

[7] M. Hamada, C. S. Reese, A. G. Wilson, and H. F. Martz, *Bayesian Reliability*, Springer, 2008.

[8] T. Masada, D. Fukagawa, A. Takasu, Y. Shibata, and K. Oguri, "Modeling topical trends over continuous time with priors," in *IEEE ISNN*, 2010.

[9] E. Fox, E. Sudderth, M. I. Jordan, and A. Willsky, "An HDP-HMM for systems with state persistence," in *ICML*, 2008.

[10] M. Johnson and A. Willsky, "The hierarchical Dirichlet process hidden semi-Markov model," in *UAI*, 2010.

[11] A. R. Cassandra, *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes*, Ph.D. thesis, Brown University, Providence, RI, 1998.

[12] M. T. J. Spaan and N. Vlassis, "Perseus: Randomized point-based value iteration for POMDPs," *Journal of Artificial Intelligence Research*, 2005.

[13] K. Huth, "Wie man ein Bier bestellt," M.S. thesis, Universität Bielefeld, 2011.

[14] E. W. Noreen, *Computer-Intensive Methods for Testing Hypotheses: An Introduction*, Wiley-Interscience, 1989.

[15] D. Bohus and E. Horvitz, "Learning to predict engagement with a spoken dialog system in open-world settings," in *SIGDIAL*, 2009.

[16] D. Bohus and E. Horvitz, "Models for multiparty engagement in open-world dialog," in *SIGDIAL*, 2009.