Towards Learning Human-Robot Dialogue Policies Combining Speech and Visual Beliefs

Heriberto Cuayáhuitl, Ivana Kruijff-Korbayová

Abstract We describe an approach for multi-modal dialogue strategy learning combining two sources of uncertainty: speech and gestures. Our approach represents the state-action space of a reinforcement learning dialogue agent with relational representations for fast learning, and extends it with belief state variables for dialogue control under uncertainty. Our approach is evaluated, using simulation, on a robotic spoken dialogue system for an imitation game of arm movements. Preliminary experimental results show that the joint optimization of speech and visual beliefs results in better overall system performance than treating them in isolation.

1 Introduction

Reinforcement learning dialogue agents have a promising application for adaptive human-robot interaction [1]. One of the main problems that affect its practical application is that the learning agents have to operate under uncertainty. Dialogue control under uncertainty has been addressed by sequential decision-making models under uncertainty [2, 3, 4, 5]. In this paper we describe an approach for tractable dialogue strategy learning under uncertainty derived from speech and gesture recognition errors in human-robot dialogues. Our scenario is an imitation game of arm movements between a child and a robot . In this memory game the child makes an arm movement (e.g. right arm up) and the robot has to imitate that movement followed by adding another one (e.g. right arm up, left arm up), then the child imitates the robot and adds another movement (e.g. right arm up, left arm up, left arm down), and so on until a maximum number of movements is reached. In these interactions the robot receives verbal (spoken) and non-verbal (gestures) observations. The rest of the paper explains how to learn dialogue strategies combining such observations.

Heriberto Cuayáhuitl, Ivana Kruijff-Korbayová
German Research Center for Artificial Intelligence (DFKI), Stuhlsatzenhausweg 3, 66123, Saarbrücken, Germany, e-mail: Heriberto.Cuayahuitl@dfki.de

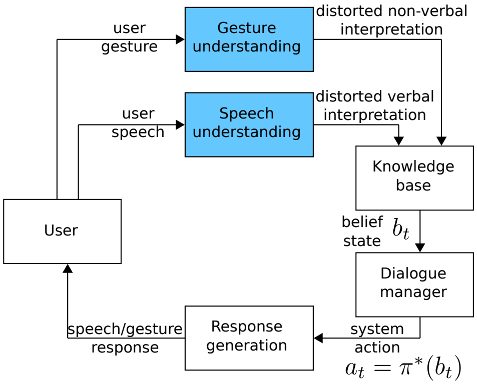

Fig. 1 A pipeline model of human-robot interaction under uncertainty, where belief dialogue state b_t is used by the dialogue manager to choose action a_t.

2 Learning Human-Robot Dialogues Under Uncertainty

A human-robot dialogue can be defined as a finite sequence of verbal and/or non-verbal information units conveyed between conversants, where the information can be described at different levels of communication such as speech signals, gestures, words, and dialogue acts. Figure 1 illustrates a model of human-robot interaction. An interaction between conversants can be briefly described as follows: the robot receives user verbal and non-verbal contributions from which it extracts interpreted user dialogue and/or gesture acts and enters them into its knowledge base. The robot then updates its belief dialogue state b_t (i.e. a probability distribution over observable dialogue states) with information extracted from its knowledge base. This dialogue state is received by the dialogue manager in order to choose a machine dialogue and/or gesture act a_t, which is received by the response generation module to generate the corresponding verbal and non-verbal response conveyed to the user.

A human-robot dialogue follows such a sequence of interactions in an iterative process between both conversants until one of them terminates it. Such sequences can be used by a reinforcement learning agent to optimize the robot's dialogue behaviour. In this paper we apply the learning approach proposed by [6], which extends the state representation of Markov Decision Process (MDP) reinforcement learning dialogue agents with relational representations and beliefs states. We apply this approach to child-robot dialogues in an imitation game of arm movements.

3 Using Bayesian-Relational State Representations for Optimizing Human-Robot Dialogues

Our employed approach unifies two concepts: relational representations and belief states. Whilst the former describes the dialogue state with expressive, compact, and high-level representations (rather than propositional ones with exponential growth), the latter provides the mechanism to handle uncertainty in the interaction [6].

A relational Markov Decision Process (MDP) can be characterized as a 5-tuple $<S, A, T, R, L>$, where the first four elements are well known in a standard MDP, and element L is a language that provides the mechanism to express logic-based representations. We describe L by a context-free grammar to represent formulas consisting of predicates, variables, constants and connectives similar to first order logic [7]. Whilst the state set S is generated from an enumeration of all possible logical forms in L, the actions A per state are constrained by the logical forms in L.

We approximate the belief states of a relational MDP with belief state variables, defined as $b(s) = \frac{1}{Z} \Pi p(X_i \in s)$, where $p(X_i \in s)$ is the probability distribution of predicate X_i in state s, and Z is a normalization constant. We maintain a Bayesian Network (BN) for each predicate $X_i \in s$. A BN models a joint probability distribution over a set of random variables and their dependencies based on a directed acyclic graph, where each node represents a variable Y_j with parents $pa(Y_j)$ [8]. The Markov condition implies that each variable is only dependent on its parents, resulting in a unique joint probability distribution expressed as $p(Y) = \Pi p(Y_j | pa(Y_j))$, where every variable is associated with a conditional probability distribution $p(Y_j | pa(Y_j))$. To that end, we use the variable elimination and junction tree algorithms [9].

4 Experiments and Results

We tested our approach in a reinforcement learning agent for the imitation game described in Section 1. The task consists in imitating arm movements by taking turns and by incrementing the movement sequence by one movement each turn until reaching a maximum length. This makes the task a memory game. The scenario represents two sources of uncertainty, coming from verbal and non-verbal user contributions, i.e. speech and gestures. Our hypothesis is that the joint optimization of speech and visual beliefs in human-robot dialogue strategy learning results in better overall system performance than treating them in isolation.

4.1 The Simulated Conversational Environment

Our simulated dialogues are based on the Dialogue Acts (DAs) and Gesture Acts (GAs) shown in Table 1 (10 user DAs and 22 system GAs). We employed the conditional probability distribution $p(u|a)$ for simulating user dialogue/gesture acts u given the last robot dialogue/gesture acts a. The user responses were coherent with probability 0.9 and random otherwise, a speech recognition error rate of 20% and a gesture recognition error rate of 30% was assumed. In addition, we modelled speech and gesture recognition events from *beta* continuous probability distributions, which have been applied to statistical dialogue simulation by [6]. The *beta* distribution is parameterized by two positive shape parameters referred to as α and β. Our simulations used (α=2,β=5; α=5,β=2) for bad and good recognition, respectively.

Table 1 Dialogue and gesture acts for the conversations in the arm imitation game.

Conversant	Dialogue/Gesture Acts	Sample Utterance
User	Provide(YourTurn)	Your turn
	Provide(IncorrectMovement)	That was incorrect
	Confirm(ContinuePlaying)	Yes
	Move(RightArmUp)	Right arm up [speech and gesture]
	Move(RightArmDown)	Right arm down [speech and gesture]
	Move(LeftArmUp)	Left arm up [speech and gesture]
	Move(LeftArmDown)	Left arm down [speech and gesture]
	Request(RepeatSequence)	Repeat that please
	Request(LastMovement)	What was the last movement?
	Silence()	[*remain in silence*]
Robot	Provide(GameInstructions)	In this game we take turns to imitate our arm movements...
	Provide(YourTurn)	Your turn
	Provide(MyTurn)	Okay, my turn
	Provide(SequenceAgain)	Okay, here is the sequence again.
	Provide(NewMovement)	The new movement is
	Provide(GamePerformance)	We scored together X out of Y points.
	Request(ContinuePlaying)	Are you ready?
	Request(Movement)	Show and tell me the first movement.
	Request(RepeatSequence)	Try again
	Request(LastMovement)	Sorry what was your last movement?
	Apology(ContinuePlaying)	I didn't hear that. Are you ready?
	Apology(GestureSequence)	I didn't get that. Show me the sequence again.
	Acknowledgement(Play)	Good
	Acknowledgement(NoPlay)	Too bad
	Move(RightArmUp)	Right arm up [speech and gesture]
	Move(RightArmDown)	Right arm down [speech and gesture]
	Move(LeftArmUp)	Left arm up [speech and gesture]
	Move(LeftArmDown)	Left arm down [speech and gesture]
	Move(NodYes)	Yes [speech and nod]
	Move(NodNo)	No [speech and nod]
	Express(Success)	Great [move body showing happiness]
	Express(Failure)	Upps [move body showing sadness]

4.2 Characterization of the Learning Agent

Figure 2 shows the context-free grammar specifying the language for the relational states in our learning agent, see sample dialogue in Table 2. The goal state is reached when the game is over. Notice that the enumeration of states using propositional representations results in 3.5 billion states, corresponding to the following vector of state variables and domain values: GameInstructions with 2 values, PlayGame with 2 values, Gestures with $5 \times 4^7 \times 10$ values (assuming a decomposed predicate), MatchedSequence with 2 values, LastGesture with 3 values, Turn with 3 values, MaxMovements with 2 values, NewGesture with 4 values, GameScore with 2 values and Timeout with 2 values.

In contrast, the relational state representations only require 666 thousand combinations (less than 0.1% of the propositional representation). We constrain the ac-

$L := l_{01}\ l_{02}\ l_{03}\ l_{04}\ l_{05}\ l_{06}\ l_{07}\ l_{08}\ l_{09}\ l_{10}\ l_{11}\ l_{12}\ l_{13}\ l_{14}\ l_{15}\ ...\ l_{27}$
l_{01}:= GameInstructions(unprovided)
l_{02}:= GameInstructions(provided) $\wedge$ PlayGame(unknown) $\wedge$ TimeOut(no)
l_{03}:= GameInstructions(provided) $\wedge$ PlayGame(unknown) $\wedge$ TimeOut(yes)
l_{04}:= GameInstructions(provided) $\wedge$ PlayGame(no)
l_{05}:= GameInstructions(provided) $\wedge$ PlayGame(yes)
l_{06}:= GameInstructions(provided) $\wedge$ PlayGame(ready)
l_{07}:= l_{06} $\wedge$ Gestures(unfilled) $\wedge$ Turn(none)
l_{08}:= l_{06} $\wedge$ Gestures(unfilled) $\wedge$ Turn(user) $\wedge$ TimeOut(no)
l_{09}:= l_{06} $\wedge$ Gestures(unfilled) $\wedge$ Turn(user) $\wedge$ TimeOut(yes)
l_{10}:= l_{06} $\wedge$ Gestures(filled, sequence, score) $\wedge$ Turn(user)
l_{11}:= l_{06} $\wedge$ Gestures(filled, sequence, score) $\wedge$ LastGesture(correct) $\wedge$ Turn(user)
l_{12}:= l_{06} $\wedge$ Gestures(filled, sequence, score) $\wedge$ LastGesture(incorrect) $\wedge$ Turn(user)
l_{13}:= l_{06} $\wedge$ Gestures(filled, sequence, score) $\wedge$ LastGesture(missing) $\wedge$ Turn(user)
l_{14}:= l_{06} $\wedge$ Gestures(recognized) $\wedge$ MatchedSequence(no) $\wedge$ Turn(user)
l_{15}:= l_{06} $\wedge$ Gestures(recognized) $\wedge$ MatchedSequence(yes) $\wedge$ Turn(unknown)
l_{16}:= l_{06} $\wedge$ Gestures(recognized) $\wedge$ MatchedSequence(yes) $\wedge$ Turn(robot)
l_{17}:= l_{06} $\wedge$ Gestures(provided) $\wedge$ NewGesture(unknown) $\wedge$ Turn(robot)
l_{18}:= l_{06} $\wedge$ Gestures(provided) $\wedge$ NewGesture(known) $\wedge$ Turn(robot)
l_{19}:= l_{06} $\wedge$ Gestures(provided) $\wedge$ NewGesture(provided) $\wedge$ Turn(robot)
l_{20}:= l_{06} $\wedge$ Gestures(corrected) $\wedge$ NewGesture(unknown) $\wedge$ Turn(robot)
l_{21}:= l_{06} $\wedge$ Gestures(corrected) $\wedge$ NewGesture(known) $\wedge$ Turn(robot)
l_{22}:= l_{06} $\wedge$ Gestures(corrected) $\wedge$ NewGesture(mentioned) $\wedge$ Turn(robot)
l_{23}:= l_{06} $\wedge$ Gestures(corrected) $\wedge$ NewGesture(provided) $\wedge$ Turn(robot)
l_{24}:= l_{06} $\wedge$ MaxMovements(yes) $\wedge$ GameScore(good, non-expressed) $\wedge$ Turn(robot)
l_{25}:= l_{06} $\wedge$ MaxMovements(yes) $\wedge$ GameScore(bad, non-expressed) $\wedge$ Turn(robot)
l_{26}:= l_{06} $\wedge$ MaxMovements(yes) $\wedge$ GameScore(performance, expressed) $\wedge$ Turn(robot)
l_{27}:= $l_{04} \vee (l_{06} \wedge (l_{24} \vee l_{25})$ $\wedge$ GameOver(yes))
sequence := [combinations of four arm movements of length seven (i.e. 4^7 sequences)]
score := $0.1 \vee 0.2 \vee 0.3 \vee 0.4 \vee 0.5 \vee 0.6 \vee 0.7 \vee 0.8 \vee 0.9 \vee 1$

Fig. 2 Context-free grammar defining the language L for the dialogue states in the imitation game of arm movements. The notation l_i denotes logical forms, i.e. groups of dialogue states of size ≥ 1.

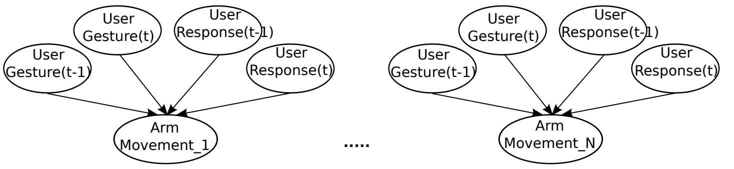

Fig. 3 Bayesian networks for modelling the beliefs of a sequence of arm movements. The random variables values are {Nothing, LeftArmUp, LeftArmDown, RightArmUp, RightArmDown}. The predicate 'Gestures()' uses multiple Bayesian nets in order to avoid large amounts of probabilities.

tions per state based on the logical forms in language L (we refer to this partially-specified dialogue policy as 'hand-coded policy'). In addition, the core Bayesian Networks (BNs) for modelling the beliefs of a sequence of arm movements are shown in Figure 3. The BNs for the remaining predicates assumed observed values (i.e. one value with probability 1; e.g., GameInstructions(provided)=1 and GameInstructions(unprovided)=0). Finally, we used the same learning settings as in [6].

Agent	State ID	Dialogue Act	Utterance (with mirrored arm movements)
Robot	l_{01}	Provide(GameInstructions)	In this game we take turns to imitate our arm movements. You do the first pose. Once you give me my turn I imitate it, and add another one. Then you imitate the two poses and add one more, then me again, and so on until a seven-long sequence.
	l_{02}	Request(ContinuePlaying)	Are you ready?
User		Confirm(ContinuePlaying)	*Yes I am*
Robot	l_{05}	Acknowledgement(Play)	Good
	l_{06}	Request(Movement)	Show and tell me the first movement.
User		Move(RightArmUp)	*Right arm up [speech and gesture]*
		Provide(YourTurn)	*Your turn*
Robot	l_{16}	Move(LeftArmUp)	Left arm up [speech and gesture]
	l_{18}	Move(RightArmUp)	Right arm up [speech and gesture]
	l_{19}	Provide(YourTurn)	Up to you.
User		Move(RightArmUp)	*Right arm up [speech and gesture]*
		Move(LeftArmUp)	*Left arm up[speech and gesture]*
		Provide(YourTurn)	*Your turn*
Robot	l_{13}	Request(NewMovement)	Sorry I didn't see the last movement?
	l_{14}	Request(RepeatSequence)	Try again.
User		Move(RightArmUp)	*Right arm up [speech and gesture]*
		Move(LeftArmUp)	*Left arm up [speech and gesture]*
		Move(LeftArmDown)	*Left arm down [speech and gesture]*
Robot	l_{16}	Provide(MyTurn)	Okay, my turn
...	...	...	...
Robot	l_{24}	Express(Success)	Great.
	l_{26}	Provide(GamePerformance)	We scored together 72 out of 100 points.

Table 2 Fragment of a human-robot dialogue playing the arm imitation game.

4.3 Experimental Results

We observed that a learnt policy with belief state variables achieves better performance (due to more accurate recognitions) than without tracking joint belief states. The latter considered either only speech recognition hypotheses (ignoring gestures) or gesture recognition hypotheses (ignoring spoken responses). We measured the average reward of the last 1000 (out of 10 thousand) training dialogues and observed that learning with joint beliefs outperforms its counterpart (without joint beliefs) by an absolute 10% (for only speech) and 13% (for only gestures) in terms of average reward. We also compared the average reward of the first 1000 training dialogues and the last 1000 training dialogues for the best policy (joint beliefs), and noticed that the latter phase outperformed the first one by 9.4%. This indicates that the dialogue policy with hand-coded constraints was improved by policy learning.

The learning dialogue agent described in this paper (sample dialogue in Table 2) has been incorporated into a complex integrated robotic system [10] (see Figure 4). This system is being used as a testbed for investigating adaptive child-robot interaction in the context of the ALIZ-E project (`www.aliz-e.org`).

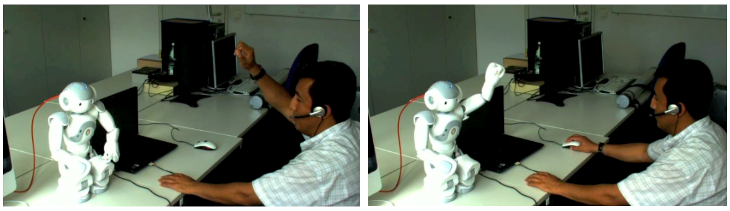

Fig. 4 *Illustrative interaction playing the arm imitation game using the NAO robot.*

5 Conclusion and Future Work

We have described an approach for learning human-robot dialogue policies, which aims for efficient and robust operation combined with straightforward design. For such a purpose we use logic-based representations in the state-action space, and extend them with belief states derived from multiple Bayesian networks. Our experimental results provide initial evidence to conclude that our method is promising because it combines more scalable learning (than propositional state representations) with robust operation. Future work consists in extending our proposed game with other games using a hierarchy of learning agents modelling belief states at different levels of granularity, and its corresponding evaluation in a realistic environment.

Acknowledgements Funding by the EU-FP7 project ALIZ-E (ICT-248116) is gratefully acknowledged.

References

[1] R. Stiefelhagen, H. Ekenel, C. Fugen, P. Gieselmann, H. Holzapfel, F. Kraft, K. Nickel, M. Voit, and A. Waibel, "Enabling multimodal human-robot interaction for the Karlsruhe humanoid robot," *IEEE Transactions on Robotics*, vol. 23, no. 85, pp. 840–851, 2007.

[2] J. Williams, "Partially observable Markov decision processes for spoken dialogue management," Ph.D. dissertation, Cambridge University, 2006.

[3] J. Henderson and O. Lemon, "Mixture model POMDPs for efficient handling of uncertainty in dialogue management," in *International Conference on Computational Linguistics (ACL)*, Columbus, Ohio, USA, Jun 2008, pp. 73–76.

[4] B. Thomson, "Statistical methods for spoken dialogue management," Ph.D. dissertation, University of Cambridge, 2009.

[5] Y. Young, M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, T. B., and K. Yu, "The hidden information state model: a practical framework for POMDP-based spoken dialogue management," *Computer Speech and Language*, vol. 24, no. 2, pp. 150–174, 2010.

[6] H. Cuayáhuitl, "Learning dialogue agents with Bayesian relational state representations," in *Workshop on Knowledge and Reasoning in Practical Dialogue Systems (IJCAI)*, Barcelona, Spain, Jul 2011.

[7] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Pearson Educ., 2003.

[8] F. Jensen, *An Introduction to Bayesian Networks*. Springer Verlag, New York, 1996.

[9] F. G. Cozman, "Generalizing variable elimination in Bayesian networks," in *Workshop on Probabilistic Reasoning in Artificial Intelligence (IBERAMIA/SBIA)*, Sao Paulo, Brazil, 2000.

[10] I. Kruijff-Korbayová, G. Athanasopoulos, A. Beck, P. Cosi, H. Cuayáhuitl, T. Dekens, V. Enescu, A. Hiolle, B. Kiefer, H. Sahli, M. Schroeder, G. Sommavilla, F. Tesser, and W. Verhelst, "An event-based conversational system for the Nao robot," in *Workshop on Paralinguistic Information & its Integration in Spoken Dialogue Systems (IWSDS)*, Granada, Spain, 2011.