Improving Human Behavior Using POMDPs with Gestures and Speech Recognition

João A. Garcia and Pedro U. Lima

Abstract This work proposes a decision-theoretic approach to problems involving interaction between robot systems and human users, with the goal of estimating the human state from observations of its behavior, and taking actions that encourage desired behaviors. The approach is based on the Partially Observable Markov Decision Process (POMDP) framework, which determines an optimal policy (mapping beliefs onto actions) in the presence of uncertainty on the effects of actions and state observations, extended with information rewards (POMDP-IR) to optimize the information-gathering capabilities of the system. The POMDP observations consist of human gestures and spoken sentences, while the actions are split into robot behaviors (such as speaking to the human) and information-reward actions to gain more information about the human state. Under the proposed framework, the robot system is able to actively gain information and react to its belief on the state of the human (expressed as a probability mass function over the discrete state space), effectively encouraging the human to improve his/her behavior, in a socially acceptable manner. Results of applying the method to a real scenario of interaction between a robot and humans are presented, supporting its practical use.

1 Introduction

Social robots need to be capable of developing affective interactions and to empathize with human users [4]. This requirement involves the ability to infer and react according to latent variables: the user's affective and motivational status.

J. A. Garcia · P. U. Lima (✉)
Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon,
Lisbon, Portugal
e-mail: pedro.lima@tecnico.ulisboa.pt

J. A. Garcia
e-mail: joao.p.garcia@tecnico.ulisboa.pt

M. I. Aldinhas Ferreira et al. (eds.), *Cognitive Architectures*, Intelligent Systems,
Control and Automation: Science and Engineering 94,
https://doi.org/10.1007/978-3-319-97550-4_10

The agent acting in a Human-Robot Interaction (HRI) scenario must take into account the effects of its actions on the human user, which are uncertain, and the sensory information it receives, which is noisy. Planning under these conditions is attainable through Partially Observable Markov Decision Processes (POMDPs) [3]. POMDPs, through the transition and observation models, deal with the aforementioned uncertainty, by probabilistically modeling the possible outcomes of the agent's different actions and the accuracy of the sensory information. Furthermore, the problem of empathizing with the human user adds the goal of information gain on latent (i.e., not directly observed) state variables, which is addressed by the extensions to POMDPs introduced by Partially Observable Markov Decision Processes with Information Rewards (POMDPs-IR) [9].

Thus, this work introduces a POMDP-IR framework for planning under uncertainty in HRI problems, which allows the agent to accomplish a given task, actively infer latent state variables of interest and adapt its behavior accordingly. The aforementioned framework is implemented in a real robot system, to ensure it is capable of successfully solving HRI planning problems in practice.

2 Related Work

Among HRI scenarios, Decision-Theoretic (DT) approaches to planning based on the POMDP framework are found in assistive scenarios, such as the robot wheelchair [10], in which the goal is to recognize the intention of the user but do not include social capabilities for improving recognition. Also, in socially assistive settings, the POMDP framework models the social interaction between robot and human users in, e.g., nursing homes [7], although without taking into account the user's status. Finally, the POMDP was used to model problems with latent variables and adapt the agent's behavior accordingly in an automated hand-washing assistant [1]. However, the agent in the latter work does not actively seek to gain information on the user's status, and is, therefore, limited to reacting based on a possibly high-uncertainty belief on the hidden variables.

The traditional POMDP model does not allow for rewarding low-uncertainty beliefs. Consequently, in order to obtain a certain level of knowledge about the features of interest, the POMDP framework needs to be extended to reward information gain. This extension is provided through the POMDP-IR (POMDP with Information Reward)) framework. DT planning based on POMDP-IR has been applied to the problem of active cooperative perception [9]. The present work, however, is focused on multimodal human-robot interaction.

3 Background

POMDP-IR can be expressed as a tuple $(S, A, T, R, \Omega, O, \gamma)$, where:

- $S = S_1 \times \cdots \times S_n$ represents the environment's factored state space, defining the model of the world;
- A is a finite set of actions available to the agent that contains the domain-level action factor A_d and an Information-Reward (IR) action factor A_i for each state factor of interest ($A = A_d \times A_1 \times \cdots \times A_l$, where l is the number of IR actions);
- T is the transition function that represents the probability of reaching a particular state $s \in S$ by a given state-action pair ($T : S \times A \times S \rightarrow [0, 1]$);
- R is the reward function, which defines the numeric reward given to the agent for each state-action pair ($R : S \times A \rightarrow \mathbb{R}$), and is therefore given by $R = R_d(s, a_d) + \sum_{i=1}^{l} R_i(s_i, a_i)$, with $s \in S$, $a_d \in A_d$, $s_i \in S_i$, $a_i \in A_i$, R_d being the POMDP reward model and R_i the information reward;
- Ω is a finite set of observations that correspond to features of the environment directly perceived by the agent's sensors;
- O is the observation function that represents the probability of perceiving observation $o \in \Omega$ after performing action $a \in A$ and reaching state $s' \in S$ ($O : S \times A \times \Omega \rightarrow [0, 1]$);
- γ is the discount factor, used to weight rewards over time.

The POMDP-IR fits into the classic POMDP framework, and can, therefore, be represented as a belief-state Markov Decision Process (MDP), in which the history of executed actions and perceived observations are encoded in a probability distribution over all states: the belief state. Every time the agent performs an action $a \in A$ and observes $o \in \Omega$, the belief is updated by the Bayes' rule:

$$b^{ao}(s') = \frac{P(o|s', a)}{P(o|b, a)} \sum_{s \in S} P(s'|s, a) b(s), \tag{1}$$

where $P(s'|s, a)$ and $P(o|s', a)$ are defined by the Transition and Observation model, respectively, and

$$P(o|b, a) = \sum_{s' \in S} P(o|s', a) \sum_{s \in S} P(s'|s, a) b(s) \tag{2}$$

is a normalizing constant. Furthermore, the value function $V^{\pi}(b)$, defined as the expected future discounted reward given to the agent by following policy π, starting from belief b:

$$V^{\pi}(b) = \mathbf{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(b_t, \pi(b_t)) \middle| b_0 = b\right], \tag{3}$$

where $R(b_t, \pi(b_t)) = \sum_{s \in S} R\big(s, \pi(b_t)\big) b_t(s)$, remains approximately Piecewise Linear Convex (PWLC) in the POMDP-IR framework. This way, the most common algorithms for solving POMDPs, which exploit the PWLC representation of the value function, can also be used to solve POMDPs-IR. The optimal policy π^* is characterized by the optimal value function V^*, which satisfies the Bellman optimality equation:

$$V^*(b) = \max_{a \in A} \left[R(b, a) + \gamma \sum_{o \in O} P(o|b, a) V^*(b^{ao}) \right]. \tag{4}$$

Solution methods for POMDPs differ from exact solution algorithms (e.g., Monahan's enumeration algorithm [5]), intractable for large problems, to approximate policy optimization (e.g., Point-based Value Iteration (PBVI) [6]). The method of reference in solving POMDPs throughout this work is *PERSEUS* [8], a randomized PBVI algorithm.

4 Framework Description

The proposed framework approaches the problem of planning under uncertainty in HRI under the POMDP-IR extension. Figure 1 represents the projected POMDP-IR as a two-stage Dynamic Bayesian Network (DBN), which depicts the dynamics of the HRI problem.

4.1 States and Transitions

The agent acting in an HRI scenario considers two types of state factors: the *task* variables T and the *person* variables P. The *task* variables model the environment features that provide information on the progress of the tasks. On the other hand, the *person* variables track the human state and are inherently latent. The latter are used to gain information on the human user's affective and motivational status and adapt the robot behavior accordingly.

The number of state variables depends on the amount of features essential to represent the environment, and is, therefore, dependent on the specific task. The criteria for the selection of states involve a trade-off between operational complexity and predicted system performance, since operational complexity increases with the number of states.

Furthermore, depending on the objectives of the agent acting in an HRI setting, the *task* variables might not exist. This is the case when the single goal of the agent is to gain information on the human user.

A *person* variable can have a constant value over time if its value does not change during the task. This is the case with personal traits (e.g., *Personality* and *Prefer-*

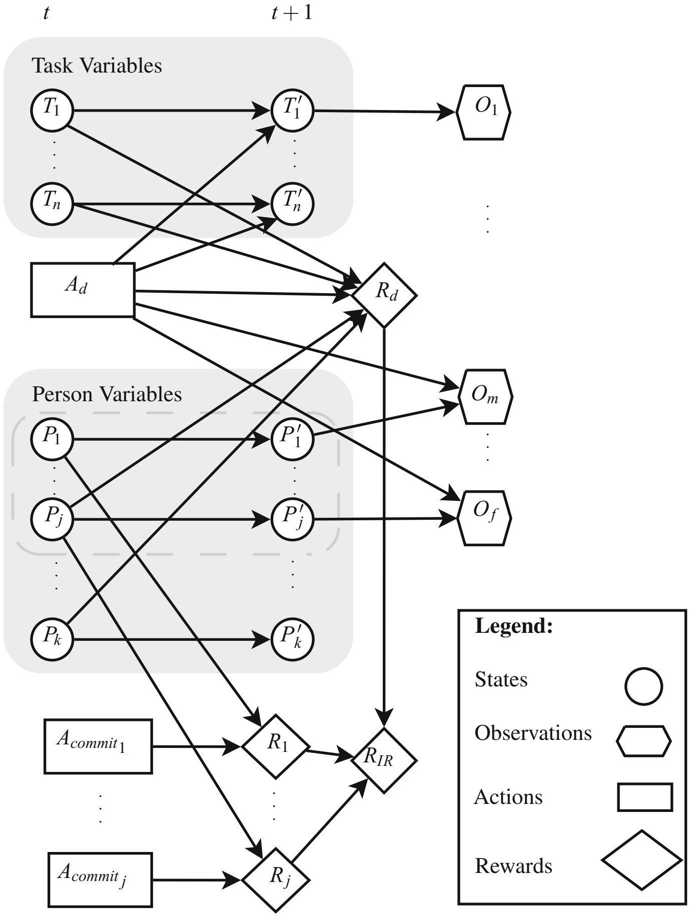


Fig. 1 DBN representation of the DT model for multimodal HRI

ences), which are relevant for the robot behavior and do not change for the duration of the interaction. In Fig. 1, P_k represents a constant *person* variable.

Otherwise, *person* variables are inferred from the user's behavior at each time step (factors P_1 to P_j in Fig. 1), which is represented in the model's observations. These state variables may consist of state factors of interest, according to the POMDP-IR framework.

4.2 Observations and Observation Model

In a social HRI setting, observations reflect the user's behavior. This behavior is used to monitor the progress of the task and infer the user's affective and motivational status.

Observations are discrete, symbolic values, classified from sensory data, which correspond to features of the environment that are observable in a given state.

The observation factors are contingent on the sensory capabilities of the robot system. Nevertheless, the correct understanding of the user's status relies on the agent being capable of recognizing human communication methods. Consequently, the robot system ought to be able to recognize speech and gestures in order to understand the human user's affective and motivational status.

The observation model is of key importance in the achievement of the information gain goals of the agent. It reflects the probability of receiving a certain observation, given the state of the environment and the action performed. Certain actions, such as questioning or approaching the user, increase the probability of perceiving certain observations. This fact is of utter importance in order to actively gain information on the user's status. The dependency on the action is represented in observations O_m to O_f in Fig. 1.

4.3 Actions

The model in Fig. 1 comprehends two sets of actions: A_d and A_{commit}. The actions in A_d have an effect on the environment and are dependent on the actuators of the agent, while the actions in A_{commit} are used to achieve the information gain goals of the agent.

Typically, the action set A_d contains the minimum set of functionalities that allow the agent to complete its tasks. Social robots need to communicate in a natural, easily understandable way with the human users. To achieve this objective, the robot must be able to express different moods and emotions. Consequently, the action set A_d of a social robot ought to include speech and/or gestural capabilities and/or graphical emotion displays.

Following the POMDP-IR framework, besides the domain-level action factor A_d, the model has additional action factors A_{commit} for each state factor of interest. The state factors of interest, in the problem under study, are included in the *person* variables, as these contain the aforementioned affective and motivational state of the human user. The actions in A_{commit} allow for rewarding the agent for decreasing the uncertainty regarding particular features of the environment.

4.4 Reward Model

In the DT model in Fig. 1, rewards are either associated with task objectives: R_d, or with the information gain goals: $R_i, i = 1, \dots, j$. The sum of these rewards, R_{IR}, constitutes the reward awarded to the agent at each time step.

The behavior of the robot consists of the sequence of domain actions A_d the agent performs. In the social HRI scenario, and in order to adapt the robot's behavior to the user's affective and motivational status, the reward assigned to an action depends not only on the *task* variables, but also on the *person* variables.

The information rewards R_i influence the behavior of the agent, with the purpose of achieving a low uncertainty regarding certain *person* variables. The value of these rewards is dependent on the threshold of knowledge required, according to the POMDP-IR framework [9].

5 Selected Application

The proposed approach was tested in a case study that considers a socially assistive task: rehabilitation therapy.

5.1 Scenario

Rehabilitation therapy includes passive or active exercises. In the first, the therapist (human or robot) physically assists the patient in moving the affected limb. On the other hand, in active exercises, the patient moves the affected limb by him/herself, while the therapist has the functions of coaching and motivating.

Up-to-date research in rehabilitation robotics mainly covers passive exercises. Nevertheless, social robots provide a way to approach active rehabilitation exercises, representing an innovative way to monitor, motivate and coach patients.

Overall, the goals of the robot therapist in the considered rehabilitation scenario are:

- To help the user in the given setting, by monitoring the patient's movements (e.g., encouraging the patient to continue if he/she stops performing the exercise);
- To adapt its behavior and, consequently, the therapy style (e.g., nurturing vs challenging the patient), in accordance with the patient's affective and motivational status.

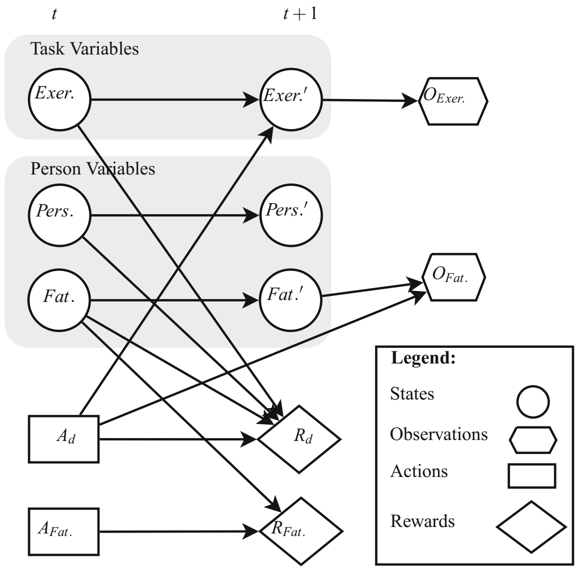


Fig. 2 DBN representation of the DT model for the robot therapist

5.2 *Decision-Theoretic Model for the Robot Therapist*

The application of the proposed framework to the robot therapist scenario results in the DT model represented in Fig. 2.

5.2.1 States

The significant features of the environment in which the robot is to operate are related to the human user. Fulfillment of the task's objectives requires that the agent keep track of the user's movements (state $Exer.$), possess knowledge regarding relevant personal traits ($Pers.$) of the user and infer his/her affective status ($Fat.$). Therefore, the proposed DT model considers the state space represented, in factored form, in Table 1.

The user's movement is encoded in the *task* state factor *Exercise* ($Exer.$). When the exercise is performed as prescribed, the state factor assumes the value *Correct*: $Exer. = Correct$. Otherwise, if the movement is inappropriately performed or not performed at all, $Exer. = Incorrect$. The state factor *Personality* ($Pers.$) is a constant *person* variable, known beforehand by the problem designer, which represents the patient's behavioral personality, as *Introverted* or *Extroverted*. Finally, the *Fatigue* state factor ($Fat.$) is a measure of the patient's weariness, caused by the physical exercise. It assumes the values *Tired* or *Energized*, depending on whether the patient shows signs of fatigue or liveliness, respectively.

Table 1 State, observation and action spaces for the robot therapist case study

	Factors	Values
States	$Exer.$	Correct, incorrect
	$Pers.$	Introverted, extroverted
	$Fat.$	Tired, energized
Observations	$O_{Exer.}$	Proper, wrong
	$O_{Fat.}$	Weary, energetic, none
Actions	A_d	Nurture, challenge, query patient, end therapy, none
	$A_{Fat.}$	Commit tired, commit energized, null

5.2.2 Observations

The observation space is represented, in factored form, in Table 1. Observations reflect the relevant behavior of the patient, in accordance with the task's goals. In the present case study, the agent ought to classify the movement performed by the patient ($O_{Exer.}$) and his/hers affective status ($O_{Fat.}$).

The gesture-related observation factor $O_{Exer.}$ is used to evaluate the exercise and assumes, as a result, the values *Proper* or *Wrong*. $O_{Exer.} = Proper$ whenever the agent perceives that the patient performed the movement as prescribed. Otherwise, $O_{Exer.} = Wrong$ if the agent perceives that the patient did not perform the movement or performed it incorrectly.

The observation factor $O_{Fat.}$, which is related to the affective status of the patient represented in state factor *Fatigue*, assumes the values *Weary*, *Energetic* or *None*. $O_{Fat.} = Weary$ or $O_{Fat.} = Energetic$ when the patient demonstrates feeling tired or lively, respectively. Otherwise, $O_{Fat.} = None$ if the agent does not perceive any relevant information regarding the affective status of the patient.

$O_{Exer.}$ is obtained by visual classification of the patient's gestures and $O_{Fat.}$ through classification of the user's verbal responses.

5.2.3 Actions

The proposed DT model considers two action factors: the *Action Domain* A_d and the *IR Action* $A_{Fat.}$ At each time step, the agent chooses one value for each action factor. The possible values for the action factors are represented in Table 1.

The IR action is defined according to the POMDP-IR framework, with a *commit* action for each value of the related state factor ($Fat.$) and a *null* action. $A_{Fat.}$ allows for rewarding the agent for reducing the uncertainty regarding the state factor $Fat.$, related to the patient's fatigue.

The *Action Domain* A_d contains the set of functionalities that allow the agent to achieve the task and information gain goals.

The therapy style, i.e., the robot's approach to the patient, changes as a function of his/her *Fatigue* and *Personality*. Dependent on these factors, the encouragement is classified as $Nurture$ or $Challenge$ if the agent opts, respectively, for a softer (e.g., "You are doing great! Keep up the good work.") or a more defiant approach (e.g., "You can do better than that!").

Since the therapy style is dependent on the *person* variables, it is important to gain information and maintain a low uncertainty regarding the state factors $Pers.$ and $Fat.$ As $Pers.$ is constant, the agent actively seeks to reduce uncertainty on the state factor $Fat.$ through the $Query\ Patient$ action. This action consists of verbally interacting with the patient to infer his/hers *Fatigue*.

Moreover, the agent ought to end the exercise ($End\ Therapy$) when the patient persistently shows he/she is not able to proceed with it. Finally, at each time step, the agent might choose to do nothing ($None$).

5.2.4 Transition, Observation and Reward Functions

The proposed framework allows us to take into account the effects of time in the states of the DT model. Namely, in the current case study, the transition function T encodes that $b(Fat. = Tired)$ increases at each time step in the absence of opposing observations ($O_{Fat.} = Energetic$). That is, the agent realistically believes that the patient is feeling more tired over time. The transition function of this case study dictates that the probability of the patient correctly performing the exercise ($Exer. = Correct$) increases with the motivation actions ($Nurture$ or $Challenge$). Moreover, Personality ($Pers.$) is modeled as a constant variable, not inferred by the agent, as its value does not change during the task.

The observation function O encodes the error in sensory data classification. This means, for instance, that even if the patient's gesture is classified as incorrect ($O_{Exer.} = Wrong$), the agent's belief about $Exer. = Incorrect$ is not 100%, and the robot might require more information before motivating the patient. Furthermore, the probabilities in O take into account that information-gathering actions (such as $Query\ Patient$) increase the probability of perceiving a verbal response from the user (e.g., $O_{Fat.} = Weary$).

The DT model in Fig. 2 rewards IR actions ($R_{Fat.}$) and A_d actions (R_d). The information rewards are defined, in accordance with the POMDP-IR framework, so that the agent actively seeks to have a certainty about $Fat.$ greater than 75% (i.e., $b(Fat. = Tired) > 0.75$ or $b(Fat. = Energized) > 0.75$). Actions in A_d are rewarded in accordance with the state of the environment: *Encouragement* actions ($Nurture$ and $Challenge$) are rewarded 0.2 whenever the patient is incorrectly performing the exercise or 0.1 when he/she shows signs of feeling tired, and penalized -0.1 otherwise. The reward given to each action also depends on the state factor $Pers.$: for an *Introverted* person, the $Nurture$ action is preferred, while the $Challenge$ action is favored for an *Extroverted* person; The $Query\ Patient$ action is penalized with -0.2; $None$ is neither rewarded nor penalized; $End\ Therapy$ receives high penalization (-1) when the patient feels energetic and a reward of 0.1

otherwise. Rewards are defined over the abstract states and actions of the DT model. The discount factor in this case study is $\gamma = 0.9$.

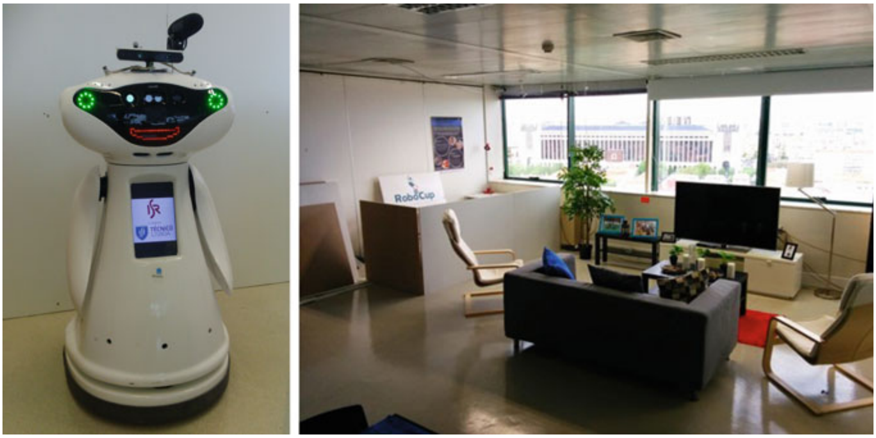
As it would be impractical to obtain the models from empirical studies, especially as the system becomes more complex, the aforementioned reward values are tuned to lead to a policy that handles different patients adequately.

6 Experiments

The robot therapist case study was implemented as a robot system consisting of a real social mobile robot networked with a RGB-D camera, which interacted, in different experiments, with distinct persons, in a realistic apartment testbed.

6.1 Experimental Setup

The networked robot system used in the present case study consists of the MOnarCH robot platform, represented in Fig. 3a, and an external Kinect camera. The robot platform provides the actuating capabilities required to implement the domain actions A_d and the sensors necessary for the speech-related observations $O_{Fat.}$. The Kinect camera is strategically located for a clear view of the patient's movements and is used, therefore, for the classification of the exercise $O_{Exer.}$.



(a) Robot Platform used in the experiments (b) Living room area of the ISRoboNet@Home testbed

Fig. 3 Experimental setup for the robot therapist case study

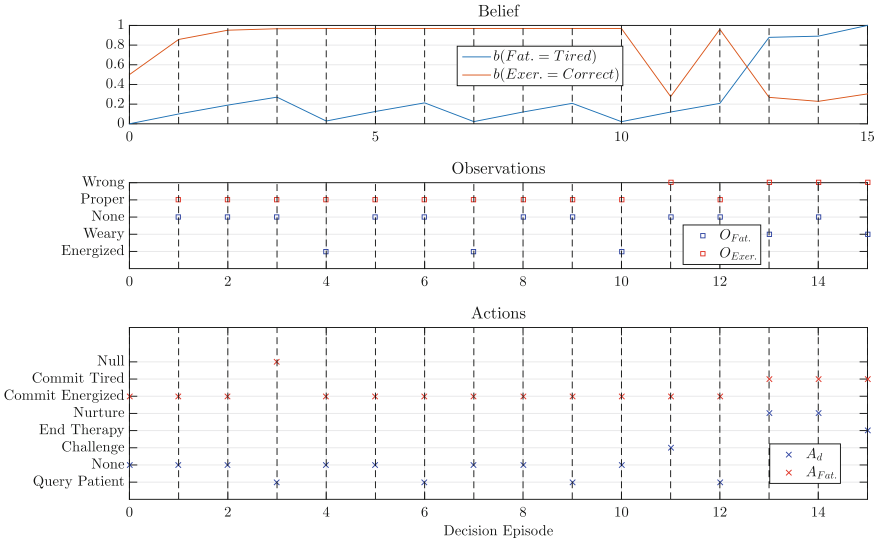


Fig. 4 Evolution of the belief about the states $Fat.$ and $Exer.$ w.r.t. the decision episode, the observations received and the actions performed, for experiment A

The experiments within this case study took place in the ISRobotNet@Home Testbed,[1] which is represented in Fig. 3b. This testbed provides the infrastructure necessary to implement networked robot systems in a domestic environment.

6.2 *Experimental Results*

Each experiment considers a different user, who is classified according to his/hers personality (i.e., as introverted or extroverted), and with regard to his/hers ability to perform the exercise (athletic or unfit).

The experiments carried out within this work were recorded, and videos are available at https://goo.gl/TlyXGT.

6.2.1 Experiment A

This experiment considers a user who is classified as extroverted ($Pers. = Extroverted$) and athletic. The user feels energetic for the first 50 s (decision step 10), approximately, and tired afterwards. Figure 4 plots the data acquired in experiment A.

At the beginning, the robot chooses not to act, since the exercise is well performed and the agent has a low uncertainty regarding the *fatigue* status of the user. This

[1] http://welcome.isr.tecnico.ulisboa.pt/isrobonet/.

uncertainty on the state factor $Fat.$, however, increases over time, driving the robot to actively seek to reduce it, by querying the user (decision step 3). The answer ($O_{Fat.} = Energetic$), informs the robot that the user is still active and motivated, increasing the certainty about $Fat. = Energized$. This behavior is repeated until the user does not perform the exercise correctly ($O_{Exer.} = Incorrect$) in decision step 11. Then, the robot motivates the person through a challenging approach according to the considered *personality* of the user and the current *fatigue* status. After receiving information that the user now feels tired ($O_{Fat.} = Weary$), the robot changes therapy style and adopts a nurturing approach. As the user continuously shows an inability to carry out the exercise and the certainty about $Fat. = Tired$ increases, the robot finally chooses to end the therapy in decision step 15.

6.2.2 Experiment B

This experiment considers a user classified as extroverted ($Pers. = Extroverted$) and unfit. The user feels energetic for the first 40 s, approximately, and tired afterwards. Figure 5 plots the data acquired in experiment B.

Figure 6 represents an episode of experiment B where the robot interacts with the user.

The behavior of the robot is similar to that in the previous experiment while the user demonstrates feeling energetic and correctly performs the exercise. Nonetheless, the user incorrectly performs the exercise more often, upon which occasions the robot

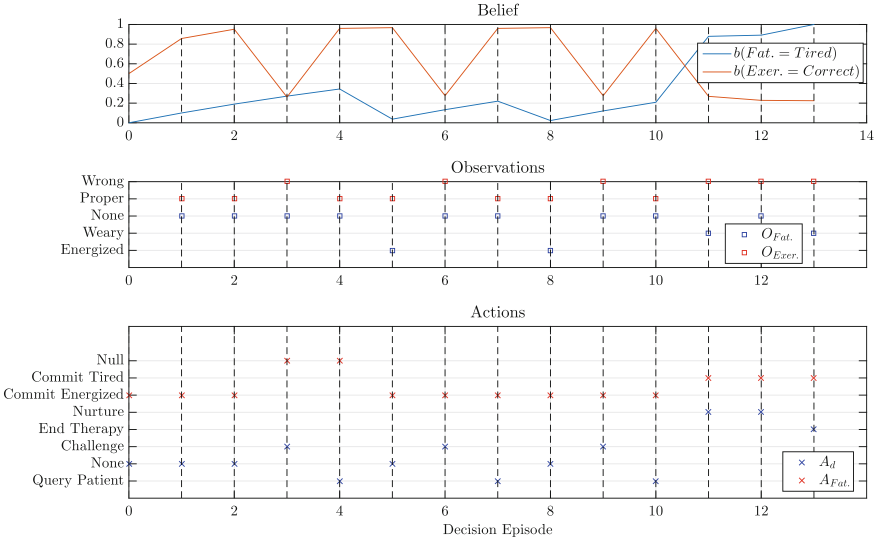


Fig. 5 Evolution of the Belief about the states $Fat.$ and $Exer.$ w.r.t. the decision episode, the observations received and the actions performed, for experiment B

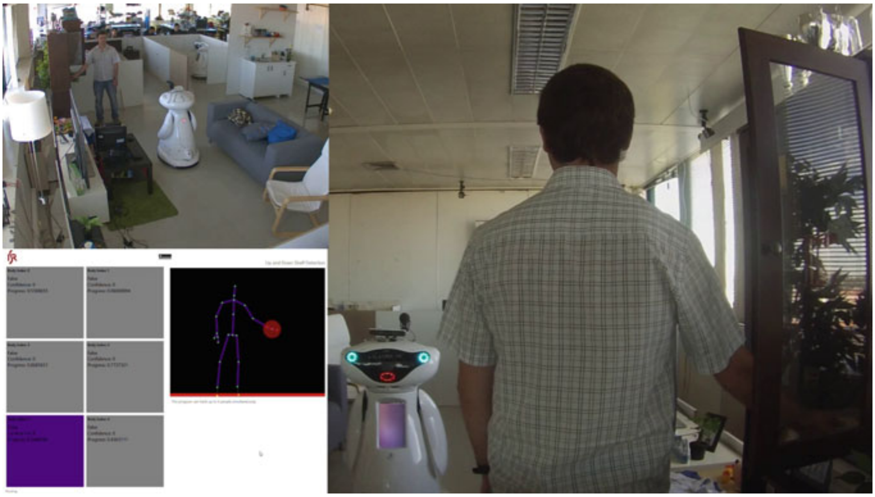


Fig. 6 Episode of the experiment B when the robot queries the user. Right and top left images show different views of the interaction between the robot and the human; bottom left image represents the interface of the gesture classification application

motivates the user with a challenging approach, while the agent believes that the user feels motivated/energetic. Despite motivating the user, the robot keeps track of his/her *fatigue* and reacts when the uncertainty about $Fat.$ is high. Finally, the agent ends the therapy once it persistently observes that the user is not performing the exercise and feels tired.

6.2.3 Experiment C

This experiment considers a user classified as introverted ($Pers. = Introverted$) and athletic. The patient feels energetic up to, approximately, 45 s (decision step 9), and tired afterwards. Figure 7 plots the data acquired in experiment C.

The behavior of the robot is heavily dependent on its knowledge regarding the fatigue status of the user. While the uncertainty about the $Fat.$ state factor is high, the robot queries the user. Since the uncertainty about $Fat.$ increases over time, the agent performs the action $Query\ Patient$ until it perceives an answer $O_{Fat} = Energetic$ or $O_{Fat} = Weary$ (decision steps 3 and 4/7 and 8). Nevertheless, the robot performs the therapy task while actively gathering information on the environment, motivating the user once the belief about $b(Fat. = Tired)$ is high, and ending the therapy appropriately.

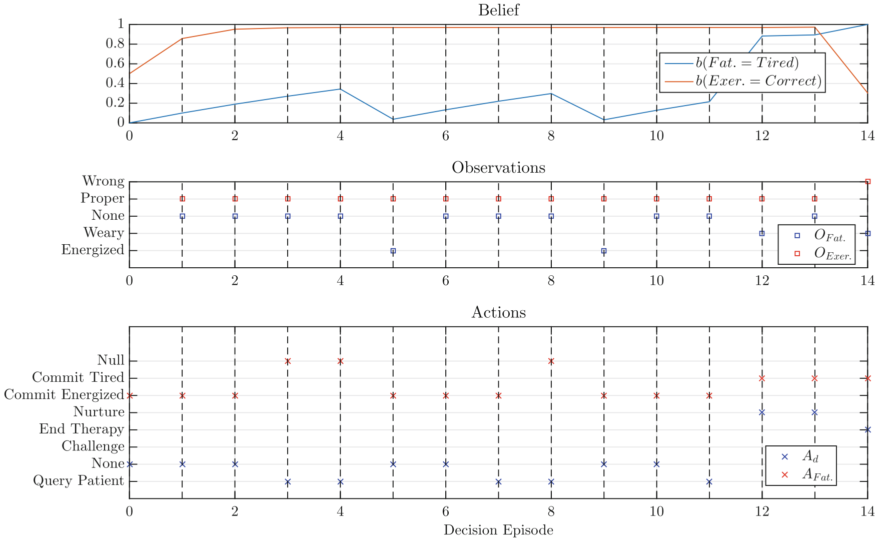


Fig. 7 Evolution of the Belief about the states $Fat.$ and $Exer.$ w.r.t. the decision episode, the observations received and the actions performed, for experiment C

6.2.4 Experiment D

This experiment considers a user who is classified as introverted ($Pers. = Introverted$) and unfit. The user feels energetic for the first 40 s (decision step 8), approximately, and tired thereon. Figure 8 plots the data acquired in experiment D.

The behavior of the robot changes in accordance with its belief about the states of the environment. In the present experiment, there is a "trade-off" between motivating or querying the user depending on the belief about the state factors $Fat.$ and $Exer.$ In decision step 3, the agent queries the agent due to the high uncertainty about $Fat.$ Afterwards, the agent perceives no answer, but observes that the user performed the movement incorrectly. This observation does not translate, however, into an absolute certainty about the exercise having been performed incorrectly ($b_4(Exer. = Correct) \approx 0.3$), since the DT framework takes into account sensor-related noise. The agent then queries the user once again (decision step 4), due to the increasing uncertainty about the *fatigue* of the user. Once again, the Network Robot System (NRS) receives no answer ($O_{Fat.} = None$), and observes that the user performed the movement incorrectly. This time, the agent's belief about $Exer. = Incorrect$ is higher ($b_5(Exer. = Incorrect) \approx 0.95$), and thus it motivates the user. Nevertheless, the uncertainty about $Fat.$ is still high in decision step 6, and the robot once again queries the user, perceiving an answer this time.

For the rest of the experiment, the robot follows a behavior similar to that of the previous experiments, until it ends the trial in decision step 14.

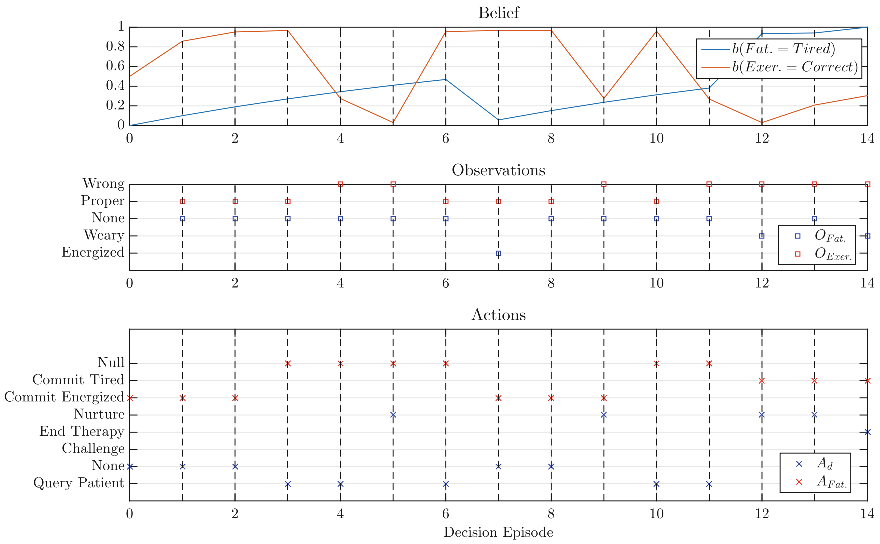


Fig. 8 Evolution of the Belief about the states *Fat.* and *Exer.* w.r.t. the decision episode, the observations received and the actions performed, for experiment D

6.3 Discussion

Table 2 details the behavior of the robot for each experiment. As expected, the number of motivation actions is higher for the users classified as unfit, who perform the exercise incorrectly more often than the athletic users; and the number of query actions is higher for the users classified as introverted.

The robot detected the fatigue status change from *Energized* to *Tired* in all of the experiments. Moreover, the agent motivated the user upon detection of faulty movements, either immediately after observing $O_{Exer.} = Wrong$ (experiments A, B and C) or after two consecutive observations (experiment D). Finally, the agent ended

Table 2 Behavior of the robot with regard to the experiment

	A	B	C	D
Motivation actions	3	5	2	4
Query actions	4	3	5	5
Time elapsed until agent detected change of user's status (s)	15	15	15	20
Time elapsed until agent ended therapy because it detected user was tired (s)	10	10	10	10
Duration of the experiment (s)	75	65	70	70

the therapy after consistently observing that the user was not capable of proceeding with the exercise.

Overall, the DT approach to planning in the robot therapist resulted in a behavior capable of achieving the task and information goals, in a manner both adaptive to the user's status and socially appealing.

7 Conclusions and Future Work

Building on the POMDP-IR framework, this work introduced a DT approach to planning under uncertainty with information rewards in social HRI. The properties of the DT framework were demonstrated in the robot therapist case study and the experiments' results validate the proposed framework for a problem involving robot systems in HRI scenarios.

Use of (PO)MDPs to model decision-making in realistic scenarios, such as the framework proposed in this work, presents an important practical difficulty, since they assume complete knowledge of the stochastic transition and observation models, meaning one needs to specify or estimate all of the probabilities involved. Moreover, any change to the parameters of these models implies a recalculation of the DT policy. Alternatively, in model-free Reinforcement Learning (RL) approaches [2], the DT policies are learned from the interaction of robot agents with their environment, without requiring full knowledge of the transition and observation models. Therefore, we plan to use RL in future applications of these methods.

To further validate the framework developed within this work, we plan its application to another health-related scenario, which we have been working with under the CMU-Portugal project INSIDE,[2] considering distinct scenarios of HRI with autistic children and their therapists. INSIDE is a research project, whose team developed a mobile robot with several interaction sensors and expressiveness skills, networked with RGB-D cameras. This networked robot system has been designed to display symbiotic autonomy when interacting with autistic children. Research has reported that autistic children are frequently willing to engage with social robots, and even create affective bonds with them. This is probably due to the predictability of the robots' behaviour. Despite the relative simplicity (when compared with a human) of the behaviours displayed by the INSIDE robot system so far, the system requires multi-modal perception systems that enable it to recognize children's activity (e.g., speech/sound, gestures, motion and location) and actuation systems so as to interact with the children using different approaches (e.g., spoken sentences, motion, "face" expressiveness). As the autonomy level of the robot system increases, autonomous decision-making methods such as the one described in this work must be included in the system.

The INSIDE robot system is composed of a mobile robot with onboard sensors, such as a LIDAR (for self-localization and obstacle avoidance), RGB-D cameras (to

[2] http://www.project-inside.pt.

detect children's faces and their emotions), and a directional microphone (to recognize children's utterances and therapist keywords), networked with four Microsoft Kinect RGB-D cameras installed on the ceiling of the room (to detect and locate the children and understand some of their gestures). Additionally, a supervision interface, comprising an actuation and a perception console, enables external operators, hidden from the children, to become aware of the interaction status and intervene in the robot decision-making process if necessary. A state machine orchestrates the sequencing of behaviours, interfacing with them through a behaviour manager. The system was developed to follow an adjustable autonomy strategy, aiming at a smooth transition from a Wizard of Oz setup (in which external operators, can override the information sensed and processed by the system, as well as the behavior selections suggested by the decision-making algorithm) to full autonomy. The symbiotic autonomy manifests itself through the fact that some of the robot behaviours consist of asking the child to help, while others make suggestions to the child as to what to do. We plan to apply our method to the development of a decision-making system that encourages the children to progress in games (e.g., building a puzzle, removing an obstacle that prevents the robot from entering an area where it can help the children during the game) by observing a child's behavior and updating the belief about his/her performance.

Acknowledgements This work was partially funded by project CMUP-ERI/HCI/0051/2013 and also by the LARSyS strategic funding from FCT, project UID/EEA/50009/2013.

References

1. Hoey, J., Poupart, P., Av, Bertoldi, Craig, T., Boutilier, C., & Mihailidis, A. (2010). Automated handwashing assistance for persons with dementia using video and a partially observable Markov decision process. *Computer Vision and Image Understanding*, *114*(5), 503–519.
2. Jaakkola, T., Singh, S. P., & Jordan, M. I. (1995). Reinforcement learning algorithm for partially observable Markov decision problems. In *Advances in neural information processing systems* (Vol. 7, pp. 345–352). MIT Press
3. Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1), 99–134.
4. Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, *5*(2), 291–308.
5. Monahan, G. E. (1982). A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, *28*(1), 1–16.
6. Pineau, J., Gordon, G., & Thrun, S. (2003). Point-based value iteration: An anytime algorithm for pomdps. In *International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 1025–1032)
7. Pineau, J., Montemerlo, M., Pollack, M., Roy, N., & Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results. *Special issue on Socially Interactive Robots, Robotics and Autonomous Systems*, *42*(3–4), 271–281.
8. Spaan, M. T. J., & Vlassis, N. (2005). Perseus: Randomized point-based value iteration for POMDPs. *Journal of Artificial Intelligence Research*, *24*(1), 195–220.

9. Spaan, M. T. J., Veiga, T. S., & Lima, P. U. (2015). Decision-theoretic planning under uncertainty with information rewards for active cooperative perception. *Autonomous Agents and Multi-Agent Systems*, *29*(6), 1157–1185.
10. Taha, T., Miro, J. V., & Dissanayake, G. (2008). POMDP-based long-term user intention prediction for wheelchair navigation. In *IEEE International Conference on Robotics and Automation, 2008, ICRA 2008* (pp. 3920–3925)