Human and robot behavior modeling for probabilistic cognition of an autonomous service robot

Sven R. Schmidt-Rohr, Martin Lösch, Rüdiger Dillmann

Abstract—This paper presents an approach to model multi-modal human-robot interaction as partially observable Markov decision processes (POMDPs) for a service robot in realistic settings. Interaction modalities include spoken dialog and non-verbal human activities like gestures and general body postures. By using POMDPs which can model uncertainties in robot perception as well as human behavior, robustness and flexibility concerning autonomous decision making are improved in real world settings. This paper presents strategies to express perception uncertainties, stochastic human behavior and typical mission objectives in explicit POMDP models. Additionally, a system is presented to compile models from more compact representations. Finally, models are actually evaluated on a physical, autonomous service robot, controlled by POMDP decision making and compared to a classical baseline controller in typical domestic missions.

I. INTRODUCTION

Autonomy and flexibility in behavior are aspects with increasing importance in contemporary robotics. This is especially true for service robots and robot companions, operating in structured and yet complex domestic environments, including natural and intuitive human-robot interaction.

Several characteristics of these environments pose a challenge to autonomous robot behavior, in particular arising from the human factor. Foremost, the environment is dynamic as events can happen without actions of the robot being the cause and the history of events is not cyclic, but ongoing. The challenge becomes more intricate by the insight that the environment is only partially observable because of sensor limitations and that environment dynamics, especially human behavior, are not fully deterministically predictable.

To be able to deal with these environments, algorithmic procedures for autonomous behavior have to regard those fundamental characteristics. A methodology enabling autonomous decision making in mission scenarios within such environments are partially observable Markov decision processes (POMDPs). Dynamic and partially observable settings with stochastic dynamics can be expressed by models describing those properties numerically from which autonomous behavior can be derived.

This paper proposes a framework to model multi-modal human-robot interaction for an autonomous service robot as POMDPs. The focus is set on being sufficiently general to model different kinds of interaction settings while at the same time being actually utilized by an existing POMDP decision making system, controlling a physical service robot in realistic settings.

In the following, related work concerning POMDP foundations, general applications as well as human-machine interaction, is analyzed. Subsequently, the capabilities of the multi-modal service robot are presented shortly together with the actual autonomous POMDP control system. The following main part describes the methodology proposed to model certain aspects in human-robot interaction as POMDPs and presents a system making knowledge management feasible. Finally, behavior of the physical robot when being controlled by POMDPs, based on the models, is compared to a baseline state machine in realistic settings and conclusions are drawn.

II. STATE OF THE ART

Probabilistic techniques have lately become popular in robotics as they take into account uncertainty in perception and control, prevalent in real world settings. These techniques cover a wide field, including perception, mission planning and actuator control in navigation, object manipulation and human-robot interaction. Concerning high-level probabilistic decision making of rational agents (e.g. autonomous robots), a promising framework are partially observable Markov decision processes (POMDPs) [1], especially the class of discrete, model based POMDPs.

A POMDP is an abstract environment model for reasoning under uncertainty [2],[3]. A POMDP models a flow of events in discrete states and discrete time. A specific POMDP model is represented by the 8-tuple \((S, A, M, T, R, O, \gamma, b_0)\). \(S\) is a finite set of states, \(A\) is a discrete set of actions and \(M\) is a discrete set of measurements. The transition model \(T(s', a, s)\) describes the probability of a transition from state \(s\) to \(s'\) when the agent has performed action \(a\). The observation model \(O(m, s)\) describes the probability of a measurement \(m\) when the intrinsic state is \(s\). The reward model \(R(s, a)\) defines the numeric reward given to the agent when being in state \(s\) and executing action \(a\). The parameter \(\gamma\) controls the time discount factor for possible future events. The initial belief state is marked by \(b_0\).

As POMDPs handle partially observable environments, there exists only an indirect representation of the intrinsic state of the world. In POMDPs, the belief state, a discrete probability distribution over all states in a scenario model, forms this representation. At each time step, the belief state is updated by Bayesian forward-filtering.

A decision about which action is most favorable for the agent when executed next, can be retrieved from a policy function which contains information about the most favorable action for any possible belief distribution. The policy incorporates balancing the expected probabilities of
the course of events into the future with the accumulated reward which has to be maximized.

Computing a policy from a POMDP model is computationally challenging and computing exact, optimal policies is highly intractable [4]. Recent approximate solutions as point based value iteration (PBVI) [5], discrete PERSEUS [6] or HSVI2 [7], however, are quite fast and yield good results for most mid-size scenarios.

The application of POMDPs in robotic domains has been stimulated by these recent algorithmic advancements. A main focus of applications is robot navigation [8], where POMDPs can deal with sensor limitations, dynamic obstacles and low-level control glitches. Another, recent application is POMDP-controlled grasping of objects [9].

A further, promising field of POMDP application is human-machine interaction, as there is usually much uncertainty in the perception of human intention and the predictability of human activities. Recent applications include dialog based human-machine interaction [10], supervision based on visual clues in the caretaker domain [11] and an early approach for spoken human-robot interaction [12].

Especially for human-machine interaction domains, it has become obvious that modeling such scenarios as POMDPs is intricate. Thus, there is a need for systematic approaches to create models containing important characteristics of multimodal interaction, including verbal and non-verbal aspects. The following work proposes such an approach for a service-robot setting.

III. CONTROL SYSTEM

The capabilities of an autonomous service robot and its control system mostly define the possible context of mission scenario settings. The service robot, used as target system and evaluation platform for the proposed methods is a typical representative. Its capability domains encompass autonomous navigation in structured indoor environments, object manipulation and natural multi-modal human-robot interaction. Each domain consists of a bunch of algorithmic components processing sensor input and controlling actuators. The perceptive low level components perform complex algorithmic procedures to present relevant environment properties in a more abstract way to a POMDP decision making module. In the system, all components deliver information about the uncertainty of their observations, to allow more specific assessment of risks by decision making. The POMDP decision making module delivers symbolic actions to a sequencer which coordinates basic commands among the available actuator components. Fig. 1 shows the robot and a rough sketch of the control system.

For human-robot interaction, some details are relevant concerning the capabilities of respective perceptive components and actuators. In case of the utilized robot these are speech recognition and human body activity recognition as well as speech synthesis and overall robot posture.

Speech recognition is realized by using an onboard microphone and the Sphinx4 [13] speech recognition engine, modified to deliver discrete probability distributions over a set of human utterances: percept human speech = \( p(utter_1), p(utter_2), ..., p(utter_n) \), \( \sum_i p(utter_i) = 1 \), e.g. \( p(“Hello robot”) = 0.2, p(“Hold position”) = 0.8 \).

Human body activity recognition is realized by using a 3D-time-of-flight camera, supported by color cameras which deliver sensory input to the human body tracking method VooDoo, employing a cylinder model of the human body posture [14]. Classification methods, based on relevance criteria and support vector machine selection of body posture over time, label symbolic human activities with certain probabilities [15]. These probabilities do not sum up to one, as some activities may occur simultaneously, e.g. sitting and waving. Thus, percept human activity = \( p(\text{act}_1), p(\text{act}_2), ..., p(\text{act}_n) \), \( \sum_i p(\text{act}_i) \geq 0 \), e.g. \( p(“Waving”) = 0.7, p(“Standing”) = 0.8 \).

Natural speech synthesis simply receives a text string and converts it into an utterance by using speakers. Overall robot posture includes finger positions, hand position and orientation, head (camera) orientation and platform orientation which are addressed by numerical coordinates and angles. Scripts addressed by the sequencer can act as symbolic placeholders for selected postures.

As shown in the system layout, POMDP decision making provides autonomy and the ability for proactive behavior in the rational agent perceive – decide – act cycle. The main control loop runs with around 20 Hz, computing a new belief state, retrieving a decision from the policy and executing a new, symbolic action in the sequencer, when applicable. First in a cycle, the belief state is updated by Bayesian filtering, using observations as delivered by perceptive components and predictions based on the last robot action, the last belief and transition probabilities in the POMDP model. Subsequently, the policy is queried using the new belief state and an action is retrieved to be executed next. It is actually executed in case the previous action has terminated in the sequencer.

The policy is a piecewise linear and convex (PWLC) maximum of a set of \( |b| \)-dimensional linear functions as

\[ \text{utter}_{0} = 0 \]

\[ \text{utter}_{2} = 8 \]

\[ \text{utter}_{8} = 0 \]

\[ \text{utter}_{7} = 1 \]

\[ \text{utter}_{5} = 0 \]

\[ \text{utter}_{6} = 0 \]

\[ \text{utter}_{4} = 7 \]

\[ \text{utter}_{3} = 0 \]
calculated by an approximately optimal value iteration algorithm, e.g. PBVI [5] or a FRTDP/HSVI [7] variant. Each policy describes approximately optimal decision making in a predefined mission scenario and can be calculated online from POMDP mission models or offline and stored in a policy knowledge database.

The next section proposes strategies to compile POMDP models for scenarios with multi-modal human-robot interaction.

IV. Model Design

As outlined, POMDPs which are discrete in states, actions and measurements are utilized in the presented system. Although recently, there has been some progress concerning the theoretical foundations of policy calculation for continuous, model based POMDPs [16], only approximate value iteration algorithms for discrete POMDPs (as outlined in sec. II) are ready for practical use. Using discrete representations, on the other hand, has the advantage to introduce efficient abstraction and improves tangibility of models and dynamics. Especially in human-robot interaction settings, low-level components, e.g. speech recognition and human activity recognition as outlined in sec. III, can perform efficient and meaningful discretization while still preserving information about uncertainties.

Because service robots are highly multi-modal, it is convenient to model the POMDP in a factored representation concerning states initially, but flatten it then. The reason is again that practical approximate value iteration algorithms exist primarily for flat representations. Additionally, inter-modality causal dependencies, especially in the stochastic transition model $T$ are easier modeled in a flat representation. Thus, the state-space $S$ is composed of a product of discrete modality-state-spaces $S_i$, each containing a finite number of discrete, symbolic states:

$s = (s_0, s_1, ..., s_n) \in S = S_1 \times S_2 \times ... \times S_n$

A. Defining states, actions and observations

Defining the semantics of states is the foundation of a model. For the modality of navigation the approach is usually straightforward, dividing the geometric space into regions of similar size and shape, where each state corresponds to the robot being in one specific region [8]. When using POMDPs for object grasping a similar, but more specific geometric discretization method can be applied [9]. On the system, as presented in sec. III, similar methods are used for those modalities. For human-robot interaction modalities, however, the state space covers a more abstract realm.

In these modalities it is reasonable to keep to the discretization and abstraction performed by low-level perceptive components. The speech recognizer translates arbitrary sound, received by a microphone, into a probability distribution over a finite set of discrete, symbolic utterances. Human activity recognition translates depth-images over some time interval into a probability distribution over a finite set of discrete, symbolic human activities.

Therefore, in the presented approach, semantics of states in these modalities keep close to the level of abstraction implied by observations as delivered by low-level components. In the modality of spoken dialog, states $s_i$ are thus defined as dialog situations, where an individual utterance leads to transitions between two states. Of course, a "not-in-dialog" anchor state $s_0$ may also be reached by transitions without utterances (e.g. human leaving). In the domain of human activity, states are defined similarly, with different symbolically classified human activities representing different states. Because the final state space is the product space of all modalities, all combinations of modality-states are possible, including arbitrary transitions. This includes non-HRI modalities, e.g. a drive action will likely move the robot to another region (transition in the navigation modality), but perhaps also break up a dialog (transition in the spoken dialog modality). In an actual scenario, the set of states is picked depending on the interaction needs in the specific setting.

Defining the semantics of actions is simple: each distinct robot utterance or gesture can lead to a state transition and has therefore to be modeled as an individual POMDP action. Concerning actions involving other modalities (e.g. "drive to location1" or "grasp cup"), individual actuator commands may be combined into a compound in the sequencer, representing a single symbolic POMDP action. Idleing has to be a distinct action, present in every scenario model, as POMDP decision making always needs an option to choose, thus explicitly choosing "do nothing now".

When coupling the semantics of states closely to the capabilities and representation of perceptive components, it is also reasonable to couple the semantics of observations closely to states. Thus, in the proposed approach, observations represent observing a specific state probabilistically, where each observation $o_i$ represents observing one specific state $s_i$.

B. Model overview

Observation model $O$, transition model $T$ and reward model $R$ reflect dynamic aspects of the scenario setting. For real world environments, any model can just approximate real dynamics which, however, is also the case for all other representations, e.g. other forms of dynamic Bayesian networks or classical finite state machines. Discretization is already a simplification of continuous domains, but usually leads to more feasible modeling. In general there is always a tradeoff between precision and feasibility for modeling environments and this is also true for POMDPs. The main goal, however, is to model perception uncertainty and stochastic action results precisely enough, to gain superior performance with POMDP decision making compared to baseline methods which do not consider uncertainty.

Considering how tailored for a certain scenario setting model aspects need to be, different conclusions need to be drawn for the respective models. $O$ models the characteristics of perceptive components and thus a more general approach can be used. $T$ is very specific, although dynamics describing parts (e.g. human behavior elements) can be partly reusable.
$R$ contains some very general constraints and some very specific parts, defining the mission objectives. In the following, these aspects are presented more closely for realistic service robot scenarios, concentrating on HRI and respective modalities.

C. Compilation of the observation model

The observation model $O(m, s)$ states the probability that the robot perceives measurement $m$ when the intrinsic state of the world is $s$. It is therefore a matrix containing discrete probability distributions which model how likely a certain measurement reflects each true state. As an example, two human utterances "hello robot" and "hold position", reflecting different dialog states may become mixed up by imperfect speech recognition. Those two will also more likely mix up than e.g. "hello robot" and "bring me the cup please" because of differences in sound similarity and length. In case both examples include just those two observations and states, the probability distribution in the first example is more evenly distributed than in the second one. By this insight, using similarity metrics is an effective, general mean to create observation models for HRI modalities. Single-modality models are created first and then combined into the product-space model. Additional effects covering inter-dependencies between modalities can be applied then. When applying such a similarity metric $\langle m_i, m_j \rangle$, each modality specific observation model entry is obtained as:

$$O_{mod}(i, j) = \frac{\langle m_i, m_j \rangle}{\sum_j \langle m_i, m_j \rangle}$$

For the modality of human utterances, depending on the trade-off between precision and feasibility, different modeling approaches can be used: syllable based similarities, complex phonetic similarity metrics or experimentally acquired similarity metrics. It has been shown that phonetic distance measures are correlated with word confusion in speech recognition engines [17]. Such methods may be utilized in scenarios with larger dialogs and extensive utterance sets. In the presented system, experimentally acquired metrics, which include scenario background noise and sensor peculiarities, are used as they are feasible in those small dialog settings.

The modality of human activity recognition can be covered by a similar approach. In this case the metric is defined over the feature space of human body postures. Because of the nature of the human body tracking system and classification method used (see sec. III) in the presented system, a numeric metric is feasible. During the training phase of the classification system, the set $F_a$ of important features of an activity and the margin $\langle a_i, a_j \rangle_{svm}$ between SVM classifications of activities $a$ can be acquired. By these means a metric with scale factor $\alpha$ can be established:

$$\langle m_i, m_j \rangle_{ActReco} = \left| F_a \cap F_{a_j} \right| \times \alpha \langle a_i, a_j \rangle_{svm}$$

This way, e.g. pointing slightly forward with the right arm and walking have a much larger distance than pointing slightly forward and pointing slightly backward with the same arm, which corresponds well with experimental observation of confusion in the actual activity recognition.

D. Compilation of the transition model

Compiling $T$ is very different from $O$, as no modality-wide metric can be used. Effects of actions which are modeled as state transitions often span several modalities, e.g. robot-say:goodbye will likely lead to an effect in both dialog and human body posture. Likewise, superimposing effects exist, spanning HRI modalities and pure robot modalities as navigation and grasping, if only caused by the duration of those actions. Furthermore, effects may have to be represented by characteristics in single transitions or many transitions including whole rows, columns or even several actions (matrices) in the third grade tensor $T(s', a, s)$. Human behavior is also a very dynamic element in the meaning as introduced in sec I, as human behavior can lead to state transitions during idle actions of the robot.

All these characteristics make the design of the transition model the most challenging aspect of model design. At this point, it appears most reasonable to design a model $T$ as a compound of individual elements describing blocks of typical human-robot interaction behavior. These building blocks can be stored in a knowledge base and combined to represent specific scenarios.

As an example, the effect of a spoken greeting by the robot towards a detected person, not yet involved in a dialog, on the intention of that person is such an individual element. It may be observed by a user study that on average 80% of the people will participate in the dialog with the robot, but 20% will ignore the robot. These values form the transition probabilities of this single building block. Other superimposing effects might exist as e.g. another human entering the scene, having an impact on the intention of the first person. However, that second effect is another individual building block. By following this approach, $T$ can be compiled from a knowledge base by combining these blocks for a specific scenario using a suitable system as explained in sec. IV-F.

E. Compilation of the reward model

The reward model matrix $R(s, a)$ can be generated easily by assigning different actions different costs (negative rewards) which leads to whole rows being modified $R(s, a)$ and assigning the few primary and secondary mission goals strong positive rewards in single entries $R(s, a)$. These rewards superimpose like in $T$. Costs in HRI depend mostly on the annoyance potential and on the time an action takes. This way, information gain reassurance actions (e.g. questions) which are a typical result of POMDP decision making are reduced to a suitable level, but not totally suppressed. Annoying the human and avoiding the risk of performing an unwanted action can be balanced this way.

F. Practical compilation system

As the models grow quickly in size, e.g. $|T| = |S|^2 \times |A|$ and superimposing effects of individual buildings blocks as used for $T$ have to be managed, a practical compilation systems is needed to calculate model matrices from more compact representations. In the presented system, a 2-tiered compilation system is used, which takes a knowledge base
and a compact representation, addressing symbolic, abstract building blocks and calculates a model from it.

The lower tier of the process compiles model matrices from compact, parameterized functions. Those functions perform arithmetic operations on the flattened model matrices in $O$, $R$ or $T$. A function may modify a single entry, row, column or matrix of a table and is addressed accordingly: $function(table, k, m, n)$ with wildcards where necessary and with $k$, $m$ being row and column number of the corresponding model and $n$ being the number of modality.

Using again the example of the greeting robot, the robot may detect a human from several locations, which are different regions in the navigation modality and thus different states. Thus when being in the intrinsic state Region1\_humanFacingRobot, there is a probability of 0.8 for a transition into Region1\_humanGreeted, from Region2\_humanFacingRobot, however, the transition with probability 0.8 is to Region2\_humanGreeted. A parameterized function can apply such an effect over all navigation states. Many other effects can be used, as in the system many arithmetic functions to perform calculations on the model matrices exist, e.g. applying gaussian functions or interpolations.

The upper tier organizes these functions in an ontology with meaningful semantics. Here, elements can be inherited and single parameters of functions modified. For example, the greeting action is of the type Introduction which is a child of Interaction which contains rules to create the according action and state transition functions to be applied on the models. The actual probability values in the case, derived from the user study are contained by Introduction0, an instance of Introduction. To actually build a scenario, instances are combined by abstract rules from which the model is compiled, e.g:

\begin{verbatim}
("Introduction0","LabRegC","","UnknownOperator0","","")
\end{verbatim}

The instance of the class Introduction0 loads typical human-robot interaction initiation transition probabilities and creates corresponding robot actions, while the location parameter restricts transitions to states, where the location modality is LabRegC. The instance derived from operator modifies the basic interaction probabilities to take into account that the human is a new interaction peer.

These semantically meaningful building blocks assemble lower tier arithmetic functions, which in turn calculate certain, related parts of the transition, reward and observation model. Addressed entries in the model matrices may not be situated close to each other in the matrix, but are related conceptually. Therefore, the presented concept makes it possible to operate on these huge, unstructured model matrices in a highly structured and organized way.

V. EXPERIMENTS AND RESULTS

A. Results

The proposed methods were evaluated on the autonomous, multi-modal domestic robot companion as outlined in sec. III and seen in fig. 2. A typical scenario was chosen for evaluation with common aspects of domestic service robot scenarios: verbal and non-verbal human-robot interaction, navigation and object delivery. The robot performs waiter duties by waiting for potentially interested persons, then tries to get into a dialog with them to finally offer to bring something. It constantly monitors if the human is facing the robot or not to have a better belief state about the dialog. When requested, it fetches a cup, then addresses the human again to ask to which destination shall be delivered, expecting an instruction with both gesture (pointing) and speech. If unsure, the robot may request a destination again. The cup is then delivered to the perceived destination and the robot returns to a suitable location to wait for further persons. Only onboard sensors are used, speech recognition is using an onboard microphone, activity recognition the onboard time-of-flight camera and navigation the onboard laser scanner. Thus, measurement uncertainties are high which favors POMDP reasoning, modeling human behavior and navigation glitches as stochastic.

Grasping was handled in the sequential programs in this case, thus the modality domains utilized were navigation, spoken dialog and human body activity. In total 8 high-level navigation states, 5 spoken dialog and 5 human body activity states lead to a $8 \times 5 \times 5$ state-space and 11 symbolic POMDP actions.

For evaluation purposes, the POMDP approach was evaluated against a purely baseline state-machine (FSM) approach and an MDP approach. For both FSM and MDP, the low level observations were in this case processed with a fixed threshold and assuming the observation with the highest probability being the correct one. The MDP, was using the same transition and reward models as the POMDP.

Both approaches were evaluated on the physical robot (see fig. 2), controlling it completely autonomously and relying solely on onboard sensors. Each time, the robot
was given exactly half an hour to perform waiter duties while the interacting human behaved stochastically, but on average according to the assumed behavior model. A human supervisor recored true states and actual behavior of the robot.

The following table presents the results of the experiment concerning some aspects of the scenario and overall results:

<table>
<thead>
<tr>
<th>Reward</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FSM</td>
</tr>
<tr>
<td>Reassurance</td>
<td>-3.0</td>
</tr>
<tr>
<td>Fetching</td>
<td>1.2</td>
</tr>
<tr>
<td>Delivery</td>
<td>1.8</td>
</tr>
<tr>
<td>Total</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Reassurance indicates penalties for asking the interacting person about wishes again, because the robot believes, it has not sufficiently understood the human. Fetching and delivery each note the total rewards of correct and incorrect fulfillment of wishes concerning those two aspects of the scenario.

The FSM shows inferior performance because it cannot sufficiently consider stochastic human behavior and imperfect observations and is thus very conservative when developing a dialog. The comparison of MDP and POMDP is interesting, because the MDP fails especially when delivering the cup. This happens because delivery to the correct location depends both on observing very similar human speech utterances and very similar gestures (pointing with the same arm into different directions). In contrast to the MDP, the POMDP has an observation model using the metrics as presented in sec. IV-C, thus it has a far better idea, when it can be sure enough to deliver the cup to the correct location. This property is a crucial argument for using POMDPs in human-robot interaction in real world settings with a lot of sensor uncertainty, even over an MDP which can already consider stochastic human behavior.

VI. CONCLUSION AND OUTLOOK

This paper proposes a general approach to create POMDP models for multi-modal service robots with human-robot interaction. Specific approaches to determine state-, action-, and observation-space as well as observation model, transition model and reward model were presented. Additionally, a rule-based system to make POMDP model compilation feasible, was shortly outlined.

A POMDP scenario with multi-modal human-robot interaction was actually created by this approach and experimentally validated on a completely autonomous, physical service robot. It was shown that the capability of POMDPs to consider stochastic human behavior and uncertainty in perception can lead to superior behavior compared to classical baseline approaches and MDPs. The POMDP method thus is promising in human-robot interaction where safety and compliant behavior even in the face of uncertainty is mandatory.

Further work will investigate creating large knowledge bases of model build blocks, including transition probabilities, directly from observing human behavior by user studies coupled to programming by demonstration and learning from observation methodologies. This way, larger dialog and interaction setting model can be acquired, which is infeasible by manual knowledge engineering, even on abstract representations.

VII. ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 126239.

REFERENCES