A Decision-Theoretic Approach to Task Assistance for Persons with Dementia

Jennifer Boger
IBBME, Univ. Toronto
jen.boger@utoronto.ca

Pascal Poupart
School of Computer Science, Univ. Waterloo
ppoupart@cs.uwaterloo.ca

Jesse Hoey
Dept. Computer Science, Univ. Toronto
jhoey@cs.toronto.edu

Craig Boutilier
Dept. of Computer Science, Univ. Toronto
cebly@cs.toronto.edu

Geoff Fernie
Toronto Rehabilitation Institute
Fernie.Geoff@torontorehab.on.ca

Alex Mihailidis
Dept. of Occup. Therapy, Univ. Toronto
alex.mihailidis@utoronto.ca

Abstract

Cognitive assistive technologies that aid people with dementia (such as Alzheimer's disease) hold the promise to provide such people with an increased level of independence. However, to realize this promise, such systems must account for the specific needs and preferences of individuals. We argue that this form of customization requires a sequential, decision-theoretic model of interaction. We describe both fully and partially observable Markov decision process (POMDP) models of a handwashing task, and show that, despite the potential computational complexity, these can be effectively solved and produce policies that are evaluated as useful by professional caregivers.

1 Introduction

Dementia is a clinical syndrome characterized by the deterioration of a person's memory and cognitive function. It is estimated that there are nearly 18 million older adults with dementia worldwide, with this number expected to reach 35 million by 2050 [1]. Alzheimer's disease is the most common form of dementia and accounts for more than half of dementia diagnoses [3]. Home care offers tremendous advantages for such individuals (e.g., economically, quality of life) over placement in a long-term healthcare facility. However, significant barriers often prevent home care from being a viable option, including the physical and emotional burden placed on family members and the cost of maintaining professional care in the home.

A pressing need for people with advanced dementia arises from the difficulty they have completing *activities of daily living (ADLs)*. A caregiver generally must guide them, step by step, through such routine activities as handwashing, toileting, dressing, moving about, taking medication, etc. Cognitive assistive technologies that help guide *users* (i.e., care recipients) through ADLs can relieve some of the stress of home care on family members and care professionals. However, current monitoring and reminding devices cannot accurately track user behavior at a step-by-step level, nor adapt to specific users. Moreover, many devices require explicit feedback (e.g., button presses), which cannot reasonably be expected of people with moderate-to-severe dementia.

Our objective is to design and develop systems that actively monitor a user attempting a task and offer assistance in the form of task guidance (e.g., prompts or reminders) when it is most appropriate and in a form that will do the most good. Such systems should tailor their guidance decisions (w.r.t. both form and timing) to specific individuals and circumstances, based on what will work best for them. In this work, we extend the prototype COACH system [12] in which behavioral monitoring of a *handwashing task* is realized with a vision system. We extend COACH by modeling the guidance decision process in a decision-theoretic fashion, more precisely, as a *partially observable Markov decision process (POMDP)*. There are several important factors that should influence guidance decisions that strongly suggest the use of this model, including perceptual noise, the stochastic nature of user behavior, the need to trade off various objective criteria (e.g., task completion, caregiver burden, user frustration and independence), and the need to tailor guidance to specific individuals and circumstances.

Our key contributions are as follows. First, we develop a POMDP model for a specific ADL (handwashing). While the model itself is of value for handwashing, the general *form* of the model is appropriate for almost any ADL guidance or intervention module. Second, we show that these models exhibit considerable structure, allowing them to be specified easily, and the underlying decision problem to be solved approximately by recent state-of-the-art algorithms. In particular, we obtain an exact solution for a fully observable MDP simplification of the POMDP; and an approximate solution is obtained for the POMDP despite its size and complexity. Third, controlled evaluation of the MDP policy by professional caregivers demonstrates its significant value as an assistive device. While MDP performance is not as good as that of professional caregivers, it is more than adequate to serve as the basis for a supplemental device which can relieve the burden on caregivers or family members. Finally, in simulation, we show that the approximate POMDP policy outperforms the exact MDP policy. This suggests that accounting for partial observability and user customization via the POMDP approach offers considerable benefits. These benefits are currently being quantified in clinical trials which will be reported in a longer version of this paper.[1]

[1] Trials are not yet complete so cannot be reported here.

2 Cognitive Assistive Technologies & COACH

Assistive technology (AT) is increasingly used to offset the impact of impairments resulting from injury, disease, and the aging process and related disorders. Typically these technologies have focused on assisting users with mobility impairments. Recent growth in the number of people with cognitive disabilities, such as dementia, has resulted in considerable research being conducted into developing *cognitive assistive technologies (CATs)* to address the difficulties that this population faces. In broad terms, CATs attempt to compensate for existing impairments by using devices, tools, or techniques that either compensate for a person's impaired cognitive ability, or translate a problem into one that matches the user's strengths. To date, computerized CATs have been of limited scope, especially in the use of AI techniques. Prototypes of cognitive aids include handheld devices that remind a person to complete basic ADLs (e.g., taking medication [8]). However, most of these systems were not designed for older adults, particularly those with moderate-to-severe dementia, and would likely not be acceptable for this population.

Several intelligent systems that use AI and ubiquitous computing techniques are currently being developed for the older adult. These include the Aware Home Project [13], the Assisted Cognition Project [7] and the Nursebot Project [15]. These new projects are similar to the work described in this paper in the sense that they attempt to incorporate AI and a decision-theoretic approach to overcome the shortcomings of current approaches. In particular, the Autominder System [16], one aspect of the Nursebot Project, applies a POMDP in the development of the planning and scheduling aspect of the system [15]. These systems do not incorporate advanced prompting techniques and algorithms, but rather are being developed as scheduling and memory aids.

COACH, our first prototype of an intelligent environment for older adults with dementia, monitors progress and provides assistance during handwashing. It uses computer vision and simple AI techniques to learn to associate hand positions (2D coordinates) with specific handwashing steps (e.g., turning on the water, using the soap), and to adjust its parameters and cuing strategies [12]. COACH uses audio prompts for each of the possible plan steps, with each step associated with a *general*, a *moderate*, or a *specific* prompt. For example, a general prompt to turn on the water simply states "Please turn on the water," while a specific prompt uses the subject's name and elaborates on the the process ("...by placing your hands on the tap in front of you ..."). Clinical trials with 10 subjects with moderate-to-severe dementia showed that the number of handwashing steps that the subjects were able to complete without assistance from the caregiver increased overall by approximately 25% when the device was used [10].

Although this prototype and its algorithms showed some success, we have identified several remaining challenges. Limitations include the assumption of full observability (i.e., that the environment and user context are fully represented by the data provided), inability to tailor its prompting strategy to specific users, and the reliance on deterministic hand-coded rules to make decisions (w.r.t. when and how to provide assistance). The ability to plan appropriate courses of action automatically, deal with partial observability, and customize behavior to specific users has led to the development of the POMDP approach described here.

3 A POMDP Approach

The decision problem faced by COACH is fraught with uncertainty, both with respect to the observability of the environment, and the effects of system actions. Furthermore, our goal is to satisfy a number of different objective criteria that often conflict and cannot be achieved with certainty. For this reason, a POMDP provides the best formal framework for modeling the decision process. POMDPs allow one to model imprecise information about the current environment state, uncertainty in the effects of actions, and multiple, conflicting objectives. Optimal policies will proffer courses of action that balance the importance of specific objectives with their odds of success; furthermore, they account for the long-term impact of decisions, including the *value of information* inherent in an action, thereby allowing to the system to actively *learn* about certain environment or user characteristics.

Many domain characteristics associated with ADL guidance (handwashing being just one example) strongly suggest the use of a POMDP. These include:

- Sensing (e.g., vision in our system) is prone to noise, hence environment variables must be estimated probabilistically (e.g., the location of a user's hands and body must be estimated due to sensor noise or occlusion). As a result, the task step the user is attempting to engage must also be estimated.
- Successful completion of a task step given a specific form of guidance is generally stochastic (e.g., a user may react appropriately to an audio prompt only some percentage of the time).
- Conflicting objectives need to be traded off against one another and their odds of success (e.g., maximizing the odds of task completion, maximizing the number of successful task steps without caregiver aid, minimizing caregiver intervention, and minimizing user frustration).
- A given form of guidance may have different odds of success for specific individuals, hence requiring customization to specific individuals based on our current estimates (e.g., a specific user may tend to get frustrated with prompting that is too frequent).

3.1 The POMDP Model

We now describe a specific POMDP model for the handwashing ADL that captures the desiderata described above. We also describe a fully observable counterpart of this model that illustrates how MDPs can be applied. The core MDP addresses stochasticity and multiple (long-term) objectives, while the POMDP extension captures partial observability, noise, and the ability to customize behavior to specific individuals by estimation of hidden user characteristics. We take pains to describe several variants of each model. This is because the efficacy study conducted using human caregivers was based on an earlier version of the MDP. We wish to evaluate our results using this baseline; but also describe

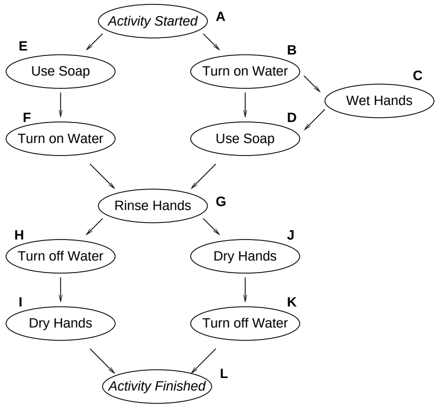

Figure 1: Plan graph for handwashing ADL.

improvements to the model which will be used in upcoming clinical trials.

A discrete-time POMDP consists of: a finite set S of states; a finite set A of actions; a stochastic transition model $\Pr : S \times A \rightarrow \Delta(S)$, with $\Pr(s, a, t)$ denoting the probability of moving from state s to t when action a is taken; a finite observation set Z; a stochastic observation model with $\Pr(s, z)$ denoting the probability of observation z in state s; and a reward function assigning reward $R(s, a, t)$ with state transition s to t induced by action a. Given a specific POMDP, our goal is to find a *policy* that maximizes the expected discounted sum of rewards attained by the system. Since the system state is not known with certainty, a policy maps either *belief states* (i.e., distributions over S) or action-observation *histories* into choices of actions. We refer to [9] for a detailed overview of POMDP concepts and algorithms.

State Variables

Our state space is characterized by four classes of variables: those that capture the state of the environment; those that summarize the ADL plan steps completed thus far; those summarizing system behavior; and those reflecting certain hidden aspects of the user's personality or mental state.

Environment variables represent the underlying physical state of the environment; these are HL (hand location: at tap, at soap, at towel, at sink, at water, away) and WO (water on: boolean).[2]

Activity status variables capture the ADL steps that the user has completed. Figure 1 shows the legitimate sequences of steps (any path from start to finish) that constitute successful handwashing. There are 4 activity status variables. PS, whose domain is the nodes (A-K) of the plan graph (L is shown only for convenience), denotes the most recent step the user has completed. MPS denotes the *maximum* plan step completed, since a user may regress in the plan graph. We use this to reward progress in the graph without rewarding duplication of steps (see reward description below); PSR denotes that the current step has been repeated. Finally, Prg indicates whether "progress" is being made within the current plan step; for example, hands moving to sink position prior to step F is indicative of progress being made towards the completion of step F despite the fact that the plan step hasn't yet been completed.

System behavior variables provide a summary of the system history relevant to prediction of user responses to a prompt. These are: NP or number of prompts issued for the current plan step (with domain 0–3+); NW or number of time steps waited since last prompt (0–4+); LP, the type of last prompt (see description of prompting actions below); PL, the specificity of the last prompt (see actions below); and Rgr, the number of times the user has regressed in the plan (0–3+), which is used as an (stochastic) indication of general responsiveness of the user.

Finally, *user variables* reflect aspects of the user's mental state that impact their response to prompts. Our current prototype uses only one variable, $Resp$ (general responsiveness), taking values low and $high$, to provide a very crude characterization of user type. However, more sophisticated user modeling can be incorporated into the POMDP, using transition and observation models of precisely the same form as those described below.[3]

Actions and Dynamics

The system has 20 actions, 18 of which comprise prompts for six different plan tasks (water on/off, use soap, wet hands, rinse hands, dry hands) at three levels of specificity (general, moderate, specific). General prompts gently cue the user, while specific prompts are designed to get the user's attention and provide a detailed description of the task. For example, a prompt to use the soap might be worded as "use the soap now" at a general specificity level, but might include references to the color or location of the soap at a specific level, such as "use the soap in the pink bottle on your left". Using the patient's name is another strategy that is effective for specific prompting. The wording of the prompts was chosen based on prior experience, and was fixed for the duration of the experiments we describe here.[4] The other two actions are the "null" action and "call caregiver." The latter action ends the process and is presumed to result in successful task completion. This is an important aspect of our system, since our aim is to develop assistive technologies that supplement rather than replace human caregivers. Our goal is to relieve stress and burden on caregivers by increasing the odds of successful "independent" task completion and reducing the need for constant (human) monitoring. However, in the efficacy study presented here, we remove this action temporarily to ensure the integrity of our evaluations (as described below). In the ongoing clinical trials, it plays a central role.

Transition probabilities describe the stochastic state changes induced by each action. These are specified by *dynamic Bayesian networks* (DBNs) over the state variables,

[2] Other relevant aspects of the environment (e.g., hands wet) can be inferred from plan step as discussed below.

[3] Other user variables might characterize auditory and visual ability, overall dementia level, general level of independence, tendency to frustration, etc. We have solved instances of the POMDP with additional user variables, but do not report these here, since they do not correspond to the MDP used in our efficacy study.

[4] Our current work includes clinical trials to assess the effectiveness of different wordings and prompting strategies. Visual prompts, played through a one-way mirror are being explored as well.

with the conditional probability tables (CPTs) represented by algebraic decision diagrams (ADDs) [4]. The considerable structure in the domain leads to a very compact specification of the dynamics. The precise parameters of the model were produced using a handcrafted prior reflecting our experience with patients in earlier clinical trials using the hand-coded prompting system and our interactions with caregivers.[5]

Space prevents a complete description of the dynamics, but we mention some of the key intuitions underlying the model. The probability of the user taking a specific "user action" such as turning on the water—corresponding to a state change in an environment variable (i.e., WO becoming true)—tends to be higher at appropriate plan steps, but the probability depends on the precise prompting history. For example, the longer the system waits for a response (modeled by NW), the less likely it is that the user will complete the current plan step independently (i.e., without prompting). Furthermore, the more "unsuccessful" prompting has been for the current step (modeled by PL and NP), the less likely it is that the patient will complete the step at all. Such "response" probabilities vary formulaicly with $Resp$ and Rgr. Higher values of Rgr are indicative of lower likelihood of eventual success, while success is more probable when $Resp$ is $high$ than when it is low.[6]

Activity status variables are updated deterministically as a function of environment variables (note that environment variables are themselves only partially observable) in the obvious way. PS may move back and forth through the plan graph (since the user can regress, e.g., by re-wetting their hands), but MPS records the maximal step in the plan reached at any point, and never regresses. System variables are updated deterministically as well (e.g., number of prompts NP is updated with each prompt for the current step, but reset to zero when the plan step changes). Finally, the user variable $Resp$ is static, and does not change value over time (though our *estimates* of its value does).[7]

Rewards

Rewards are associated with various state transitions and costs with specific actions. A large reward is given for full task completion (+300) and smaller rewards (+3) are associated with the (first) achievement of each plan step to encourage/reward even partial success (MPS is used to ensure rewards are not repeated). Action costs are also incorporated to ensure that prompting only occurs when needed. Each prompt is given a small negative reward, with the more time-consuming specific prompts penalized slightly more than simpler general prompts (the costs are -4, -5 and -7). The cost of calling the caregiver should be set so that this action only occurs if the predicted odds of completion are too low, or predicted costs of completion too high.

The design of the reward function is based on brief interactions with caregivers, earlier clinical trials with patients using hand-coded COACH, and evaluation of numerous versions of MDP (rather than POMDP) policies. While the rewards seem reasonable, fine-tuning of the reward function is an ongoing research project (see Section 5).

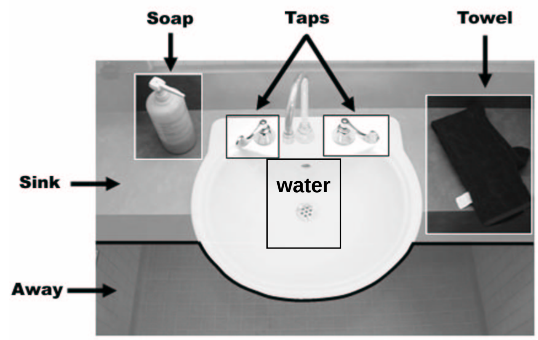

Figure 3: Possible hand locations.

Observations

The handwashing task has two observables: the *reported flow of water* (on/off), and the *reported location* of the user's hand(s). The POMDP is designed to be integrated with a computer vision system that estimates hand position (and water flow indirectly) based on skin color [11]. The vision system reports estimated hand position (x, y-coordinates) which is translated into one of six possible regions (see Fig. 3).

We assume 25% noise in the observation function that detects hand position (i.e., the correct hand position is detected with probability 0.75, while each incorrect position is detected with probability 0.05. Similarly water flow is detected correctly with probability 0.75. These probabilities are estimated based on our prior experience, but a detailed empirical study is needed to assess these accurately.

3.2 A Fully Observable Model

The POMDP model can be made fully observable with two simple changes. First, if we remove the hidden variable $Resp$, a key source of partial observability disappears, though we sacrifice the ability to customize behavior to specific individuals.[8] Second, if we assume the tracking system computes hand position and water flow without error, the system becomes fully observable. The assumption of perfect observability is not unreasonable if we use a switch sensor on the tap and additional cameras to disambiguate obscured views and reduce the inherent noise in image processing.

3.3 Computational Results

To study the feasibility of the model, we solved both the MDP and POMDP formulations of the handwashing ADL. The cost of calling the caregiver in both trials was set very high (-1000), to ensure that this action was never taken,

[5]Some versions of our model (see Section 5) include updates of this prior using data produced by coding 21 hours of video capturing 600 trials of human caregivers interacting with patients.

[6]The influence of $Resp$ on user response is estimated based on the authors' previous clinical experience. Planned clinical evaluation of the POMDP policy will be used to assess this more precisely.

[7]This is true for the time scales we currently consider, but certain aspects of a user's state, such as level of dementia, general responsiveness, etc., can change (e.g., deteriorate) over longer periods.

[8]Customization based on other fully observable attributes, such as hearing and visual ability, is still possible in an MDP.

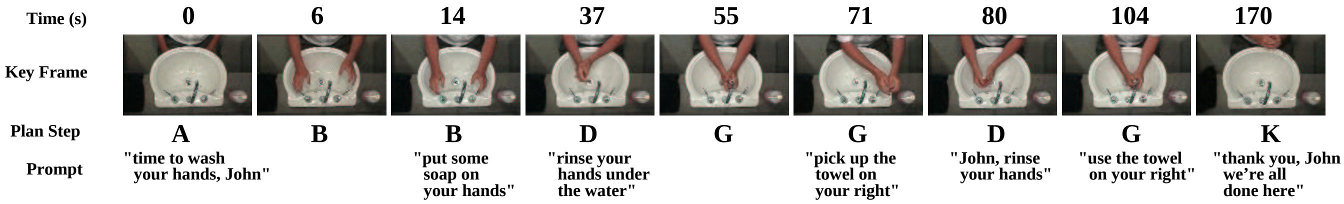

Figure 2: Example sequence during a trial in which prompts were selected by the MDP policy. Prompts were read by a human caregiver. The plan steps are those from Figure 1. The user turns the water on and off independently (at 6 and 160 seconds, respectively), but must be prompted for all other steps. The user regresses at 71 seconds by putting more soap on his hands, which the system recognizes and thus gives a second prompt to rinse hands.

as required by the evaluation procedure we describe in Section 4. The version of the fully observable MDP we solved has 12 variables (25,090,560 states) and 20 actions, rendering explicit state-based approaches impossible. However, due to the DBN structure, we were able to solve this problem exactly using the SPUDD algorithm [4]. An optimal policy was produced in 64 minutes, and was represented using an ADD with $3,284$ internal nodes and 18 leaves. The optimal value function was an ADD having $139,443$ internal nodes and $106,328$ leaves. We discuss evaluation of this policy below.

The POMDP model has one additional variable, hence a state space of size 50,181,120, the same action space, and 12 observations. It is far beyond the reach of any exact solution techniques or model-based approximation methods. As a result, we developed a new approximation algorithm which exploits the structure in the system dynamics and rewards (as represented by our ADDs). *Perseus-ADD* consists of the Perseus algorithm [18], a randomized point-based version of value iteration, reconstructed to take advantage of ADD structure [17]. The algorithm results in a policy of 36 state-value functions (α-vectors), computed in 42.7 hours.

The relative quality of the MDP and POMDP policies demonstrates the importance of accounting for noise and user characteristics. While the POMDP policy is approximate, we evaluated it in simulation from the starting state, using 500 trials of 60 steps, with $Resp$ uniformly distributed in the initial belief state, and the specific value of $Resp$ drawn randomly for each trial. The average value obtained by the POMDP policy is 100.1 over 60 steps. The value of the optimal policy for the fully observable MDP (ignoring $Resp$) is 134.2 (which provides an upper bound on the value of the unknown optimal POMDP solution). However, this MDP value cannot be realized in the partially observable environment. So we implemented the MDP policy in this environment by computing the most likely state at each stage of the simulation and applying the corresponding MDP policy choice. In simulation, this attains an average value of 95.6 over 60 steps. We see that the MDP model provides a reasonable approximation for this POMDP despite its limitations. However, the full POMDP policy (despite only being solved approximately) outperforms the MDP policy by a significant margin. The value of POMDP modeling is apparent when one considers that the computational overhead in solving the policy is borne offline—once computed, the optimal policy is applied in real-time without any significant computation. All that is required online is simple belief-state updating, the critical component in allowing the policy to choose different prompting strategies for different users as beliefs about these users evolve.

It is also interesting to note the customization of the POMDP policy based on its estimate of the *Resp* variable. When a patient tends to be slow to respond to prompts, it is better to wait several time steps before repeating the prompt. For instance, when a patient has reached plan step *G* and was prompted to dry her hands two time steps ago, the POMDP policy repeats the prompt when the probability of $Resp = high$ is sufficiently high, but waits when it is low. Similarly, when plan step *J* is reached and the patient was prompted to turn off the water two time steps ago, the prompt is repeated only when responsiveness is believed to be high. In contrast, the MDP policy doesn't have the ability to estimate the level of responsiveness since it is not directly observable. As a result, in both cases it repeats the prompt at the risk of annoying a slower, less responsive patient.

4 Caregiver Evaluation

While the evaluation of the MDP and POMDP policies in simulation is useful, the true value of the system can ultimately only be gauged in clinical trials. Clinical trials with Alzheimer's patients are currently underway to test our results. These trials are based on improved versions of the models (see Section 5). Results of these trials will be reported in a longer version of this paper.

Before undertaking clinical trials, an *efficacy study* of the MDP policy was undertaken to confirm its plausibility before applying it to dementia patients. In this study, an actor with considerable experience with dementia patients simulated the behavior of a user in the handwashing ADL during 33 trials, each of which was videotaped. This simulated patient was guided using either the prompting strategy given by the MDP policy, or by a professional caregiver with over 30 years experience. During the caregiver trials, the caregiver acted naturally, prompting using her own strategy. During the MDP trials, the same professional caregiver read (verbatim) the prompts provided by the MDP policy, in order to prevent verbal distinction between caregiver and MDP trials. The *HL* and *WO* variables (describing the location of the actor's hands and whether the water was on) were manually annotated by a researcher during the trials, thus fulfilling the perfect observability assumption of the MDP (note that using the vision system would not have allowed accurate implementation of the MDP policy). The videos were then viewed by 30 professional caregivers (different from the prompter), who evaluated the performance of the prompting in each trial. These evaluators were unaware that the prompts in some trials were selected by a computer (this is why the *call caregiver* action could not be used).

Fig. 2 depicts snapshots of one of the MDP-guided scenarios used in the efficacy study. In this case, the handwashing subject is able to complete the steps of turning on the water (step B) and turning the water off (step K) independently. The subject ignores the prompt given to him at $t = 71s$ to dry his hands and regresses in the activity to step D by applying soap instead. The planning system copes with this by identifying the regression, and prompting him to rinse his hands a second time. The majority of the time the prompting screen remained blank, as portrayed at $t = 6s$ and $t = 55s$, allowing the subject time to attempt each step on his own.

Six of the 33 scenarios (three MDP and three human) were chosen for evaluation by 30 professional caregivers.[9] Each evaluator rated the strategy employed in each of the six scenarios, using a five-point Likert scale, on five criteria: *Identification* (the prompt(s) given appropriately identified the next task); *Detail* (level of prompt detail appropriate); *Time* (an appropriate amount of time to attempt task provided before being prompted); *Repetitions* (number of prompt repetitions was appropriate); and *Overall Effectiveness* (patient was guided effectively). Free-form comments were also provided.[10] Quantitative results (see Fig. 4) clearly indicate (results are statistically significant) that the professional caregiver outperforms the MDP policy in all evaluated aspects. Indeed, we had no expectation that the MDP policy would perform as well. Of course, our goals are much more modest; we intend the system to supplement human caregivers, not replace them. Furthermore, the standard of an experienced professional caregiver sets the bar quite high.

Despite this, we were quite encouraged by the performance of the MDP. The average rating of the MDP was higher than that of the human professional in 28 of 150 (18.7%) evaluator-aspect pairs. Furthermore, qualitative comments indicate that the evaluators viewed the MDP policy as adequate (and we believe the evaluators to be rather critical given the average rating of the human caregiver); and none suspected the prompting to be computerized. These facts suggest the MDP prototype can serve as the basis for a supplemental assistive device. Finally, many of the critical comments made of the MDP policy reflected issues that could easily be remedied (and do not reflect a weakness in the MDP approach per se). For instance, many evaluators felt that turning the water on before asking a patient to use soap, or turning the water off before drying hands, provides a natural cue as to what the next activity step is (different transition priors or further data for training the transition model could readily reflect this). Suggestions were also made regarding the language construction of prompts, the use of positive feedback and a friendly voice, and checking water temperature before allowing the patient to immerse his/her hands. None of these require modification of the current MDP model.

[9] Selection criteria were based on scenario length (no more than 3 minutes) and the actor's head being not visible in scene

[10] A detailed technical report on experimental methodology and results is available elsewhere [2].

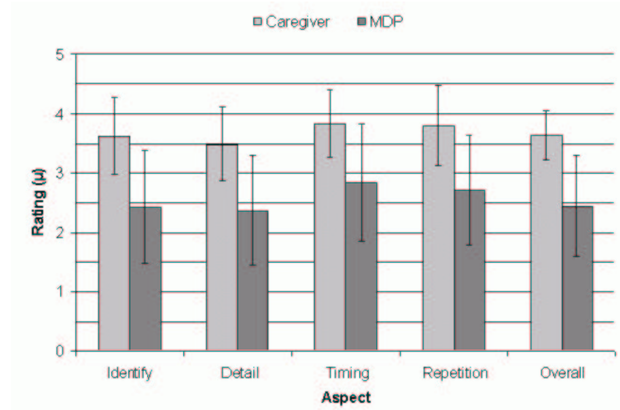

Figure 4: Caregiver ratings of MDP effectiveness for 5 criteria.

Addressing certain evaluator criticisms would require some modifications of the model (or fine-tuning of parameters). These include: using prompts that are tailored to the abilities of the user; giving the patient enough time for step completion; and incorporating visual cues (something that we have developed and tested independently). Finally, in some situations, the patient asked the caregiver a question, such as "where's the soap?". With no dialog model or speech sensors, the MDP cannot react to this (which evaluators viewed negatively). However, POMDP models of spoken dialog [19] could be used for this purpose in the future.

5 Further Model Validation & Improvement

The efficacy study supports the value of the POMDP approach. Although the study was based on a handcrafted MDP model, the policy evaluated by caregivers showed promise as a valuable means of relieving caregiver burden. Since the POMDP policy outperforms this MDP policy in simulation, we expect the POMDP model to offer additional value in a realistic settings. Ultimately, the value of the POMDP approach must be verified in a clinical setting. Clinical trials are currently underway in which both the MDP and POMDP approaches are being compared to human caregivers in guiding Alzheimer's patients through handwashing. The POMDP policy being evaluated is, in fact, somewhat different than the one reported here; we focus on the direct comparison of the POMDP described above with the MDP since it is a direct extension of the MDP (and thus differences in performance are largely due only to the power of the model to handle noise and user characteristics). The model used in the current clinical trials is significantly enhanced in several ways: a more realistic reward/cost function; a more realistic model the influence of level of responsiveness and response delays on step completion; model parameters estimated from video data of human caregivers; and a simplified and improved plan graph reflecting suggestions of the MDP policy-evaluators.

Our experience to date suggests a number of important avenues for model improvement. A difficult task is constructing reasonable reward models; while transition models can be constructed from data, rewards cannot. One future aim is to validate the reward model indirectly by having caregivers critique the policies directly. While caregivers often have difficulty quantifying the utility of, say, partial versus complete task success, or the cost of prompting, they often "recognize a good policy when they see it." Using precise caregiver policy critique, we will re-engineer the reward function so that the resulting optimal policy is consistent with the suggestions (reminiscent of revealed preference in economics or *inverse reinforcement learning* [14]).

Another important direction is learning model structure and user behaviors from sequence data, rather than imposing our own structure on tasks. Preliminary work along these lines is reported in [6]. Finally, generalizing the model to more directly account for continuous nature of both time, states (e.g., hand position) and observations in this domain is of critical importance for more accurate modeling and useful prompting. Preliminary results dealing with continuous observations in a simplified version of our domain, using a new algorithm for solving continuous observation POMDPs, are reported in [5].

6 Concluding Remarks

We have proposed a decision-theoretic view of ADL assistance for people with dementia, arguing that POMDPs provide an ideal model for such problems. While we developed this model for a specific ADL, the modeling principles embodied in our approach are broadly applicable. Despite the size and complexity of the model, we have demonstrated that the MDP and POMDP solution algorithms can successfully be used to solve these problems, thus producing (approximately) sequentially optimal policies for ADL prompting.

Considerable future work is required to apply the system in practice. Clinical trials are currently in progress to help evaluate the effectiveness of the model and to help improve it. Of particular interest are ways to to improve the (admittedly simple) user model. Sessions with caregivers are also planned to help refine our reward function. We are currently exploring many of the interesting technical issues associated with revising MDP and POMDP models (especially reward functions) based on policy critique. We are also working toward applying POMDP models to other ADLs, such as toileting, and general living patterns. Modeling these different tasks in a consistent, decision-theoretic fashion will eventually lead to hierarchies of POMDP CAT systems throughout a home or care facility, all working towards the same goal. The ability of these POMDPs to monitor *user variables*, such as $Resp$, means that changes in a patient's state of health can be assessed automatically over different time scales.

Acknowledgements

The authors thank the Bonnie Fernie for help with the experiments, and the caregivers who participated in the evaluations. The authors gratefully acknowledge the support of the Natural Sciences and Engineering Research Council (NSERC), Intel Corp., the American Alzheimer Association, and the Institute for Robotics and Intelligent Systems (IRIS).

References

[1] J. Bates, J. Boote, and C. Beverly. Psychosocial interventions for people with a dementing illness: A systematic review. *Journal of Advanced Nursing*, 45(6):644–658, 2004.

[2] Jen Boger, Jesse Hoey, Pascal Poupart, Craig Boutilier, Geoff Fernie, and Alex Mihailidis. A planning system based on markov decision processes to guide people with dementia through activities of daily living. Manuscript, 2005.

[3] D. Geldmacher and P. Whitehouse Jr. Differential diagnosis of alzheimer's disease. *Neurology*, 48(5-Suppl.6):S2–S9, 1997.

[4] J. Hoey, R. St-Aubin, A. Hu, and C. Boutilier. SPUDD: Stochastic planning using decision diagrams. In *UAI*, pages 279–288, Stockholm, 1999.

[5] Jesse Hoey and Pascal Poupart. Solving POMDPs with continuous or large discrete observation spaces. In *Proc. Intl. Joint Conf. on Artificial Intelligence*, Edinburgh, July 2005. to appear.

[6] Jesse Hoey, Pascal Poupart, Craig Boutilier, and Alex Mihailidis. Semi-supervised learning of a pomdp model of patient-caregiver interactions. unpublished manuscript, April 2005.

[7] H. Kautz, L. Arnstein, G. Borriello, O. Etzioni, and D. Fox. An overview of the assisted cognition project. In *AAAI-2002 Workshop on Automation as Caregiver: The Role of Intelligent Technology in Elder Care*, Edmonton, 2002.

[8] N. L. Kirsch. Computer-assisted interactive task guidance: Facilitating the performance of a simulated vocational task. *Journal of Head Trauma Rehabilitation*, 7(3):17–25, 1992.

[9] W. S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28:47–66, 1991.

[10] A. Mihailidis, J. C. Barbanel, and G. R. Fernie. The efficacy of an intelligent cognitive orthosis to facilitate handwashing by persons with moderate-to-severe dementia. *Neuropsychological Rehabilitation*, 14(1/2):135–171, 2003.

[11] A. Mihailidis, B. Carmichael, and J. Boger. The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home. *IEEE Trans. on Information Technology in Biomedicine (Spec. Issue on Pervasive Healthcare)*, 8(3):1–11, 2004.

[12] A. Mihailidis, G. R. Fernie, and J. C. Barbanel. The use of artificial intelligence in the design of an intelligent cognitive orthosis for people with dementia. *Assistive Technology*, 13:23–39, 2001.

[13] E. Mynatt, I. Essa, and W. Rogers. Increasing the opportunities for aging in place. In *Proceedings of ACM Universal Usability Conference*, Arlington, VA., 2000.

[14] Andrew Ng and Stuart Russell. Algorithms for inverse reinforcement learning. In *ICML*, pages 663–670, Stanford, CA, 2000.

[15] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun. Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems*, 42(3–4), 2003.

[16] M. E. Pollack. Planning technology for intelligent cognitive orthotics. In *AIPS*, pages 322–331, Toulouse, 2002.

[17] Pascal Poupart. *Exploiting Structure to Efficiently Solve Large Scale Partially Observable Markov Decision Processes*. PhD thesis, University of Toronto, Toronto, Canada, 2004.

[18] N. Vlassis and M. T. J. Spaan. A fast point-based algorithm for pomdps. In *Proc. Belgian-Dutch Conference on Machine Learning*, Brussels, Belgium, 2004.

[19] B. Zhang, Q. Cai, J. Mao, and B. Guo. Planning and acting under uncertainty: A new model for spoken dialogue system. In *UAI*, pages 572–579, Seattle, WA, 2001.