A POMDP Framework for Human-in-the-Loop System

Chi-Pang Lam and S. Shankar Sastry

Abstract— **Human operators are involved in many real world systems such as automobile systems. Traditional human-assistance features such as warning systems in the aircraft and automatic braking systems in automobile only monitor the states of the machine in order to prevent human errors and enhance safety. We believe that next generation systems should be able to monitor both the human and the machine and give an appropriate feedback to them. Although having human in the control loop has its advantage, it lacks a unified modeling framework to manage the feedback between the human and the machine. In this paper, we will present how partially observable Markov decision process (POMDP) can be used as a unified framework for the three main components in a human-in-the-loop control system—the human model, the machine dynamic model and the observation model. We use simulations to show the benefits of this framework. Finally, we outline the key challenge to advance this framework.**

I. INTRODUCTION

Most traditional manual control systems such as automobile systems belong to a low *level of autonomy* in Parasuraman's taxonomy[7]. In such systems, there is nothing to do when the human makes error, which may result in accidents. Therefore, a system with a higher level of autonomy is necessary in which the controller can monitor the human and the state of the machine and then give appropriate feedback to them. Traditional human-assistance features only monitor the state of the machine in order to prevent human errors and enhance safety. We believe that next generation systems should not only monitor the states of the machine but also the states of the human. Moreover, the automatic controller could take over human control in emergent cases. Suppose you aim to maintain a car in a single lane and your physiological state could be drowsy or awake, the system should give you alarm signals when you are drowsy. If the alarm cannot wake you up, the controller could take over your steering wheel to maintain the car in the middle of the lane. We refer to such system as human-in-the-loop (HITL) control system where there is a mechanism of interaction between the human and the controller.

From the above motivating example, we know that in order to determine when to give feedback to the human and machine, we have to estimate the human's intent and her physiological state. However, we have no way of knowing what human thinks directly. Since actually measuring the human's brain activity is too restrictive and intrusive, we want to estimate the human's intent by observing her actions. In order to design effective controls for HITL systems, we would want to use a modeling framework that have the following properties: first, the model should be probabilistic because it is impossible to measure the human's intent or physiological state exactly. This can be achieved by maintaining a probability distribution over the state of the human by observing what the human has been doing; second, it should be a sequential process that represents the long-term planning of the whole system; lastly, it should be able to handle the observation error. Given these facts, we propose to cast a HITL system as a partially observable Markov decision process (POMDP). We will show how POMDP is capable of integrating the human model and the machine model as well as their interaction in a probabilistic framework.

Pentland et al. proposed that many human behaviors can be accurately described as a set of dynamic models sequenced together by a Markov chain, called a Markov dynamic model (MDM)[8], in which they defined multiple dynamic models as internal states. Takano et al. [14] modeled the driving pattern primitives consisting of states of the environment, vehicle and driver as a hidden Markov model (HMM). Although HMM is popular for human behavior modeling [17][5][19] given the fact that it provides a stochastic framework for intent reasoning and is able to handle the uncertainty from observation, it fails to unify the effect of feedback for the human or machine. We will show that POMDP makes up for this drawback.

Most researchers used a shared control scheme to incorporate human and machine control. Chipalkatty et al. directly modified the human inputs to make the actual inputs not only conform with the human's intent but also satisfy the dynamic constraint based on a sequence of predicted human inputs[3]. Vasudevan et al. measured safety of a driving vehicle [16] to determine when to intervene. Anderson et al. used model predictive control to find a safe and optimal vehicle path and then control the vehicle via a weighted sum of human input and controller input based on threat assessment [1]. The common factor in these approaches is that they plan for future states and use a shared control scheme to make the future states satisfy certain criterion like safety and dynamic constraints along the future plan. These controllers only make use of the feedback to the machine, but do not consider incorporating the feedback to the human such as warnings. We will show that how our POMDP framework can incorporate the feedback to the machine and feedback to the human to do future planning.

POMDPs have been used in a variety of real-world sequential decision processes, including robot navigation, assistive technology, and planning under uncertainty. POMDPs have been shown to be successful in many kinds of human-

This work is supported by ONR MURI project N000141310341.

Chi-Pang Lam and S. Shankar Sastry are with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, USA. {cplam,sastry}@eecs.berkeley.edu

machine systems. Williams et al. used a POMDP to model a spoken dialog system and demonstrated significant improvement in robustness compared to existing techniques[18]. Hoey et al.[4] used a POMDP framework to implement assistance to people with dementia and showed its ability to estimate and adapt to user psychological states such as awareness and responsiveness. Broz et al.[2] modeled human-robot interaction as a time-indexed POMDP and showed that it achieves better results than simpler models that make fixed assumptions about the human's intent[2].

In this paper, we aim to address one of the three challenges for employing HITL control proposed in [6]: determining how to incorporate different models into a formal methodology of control. We propose a novel POMDP framework for human-in-the-loop control systems. We present the basic structure of HITL control system and show how the POMDP framework can incorporate all the components in a HITL control system—the human model, machine dynamics model and observation model—to determine an optimal feedback policy for when the controller should give feedback to the human (such as warning) and take over control from the human.

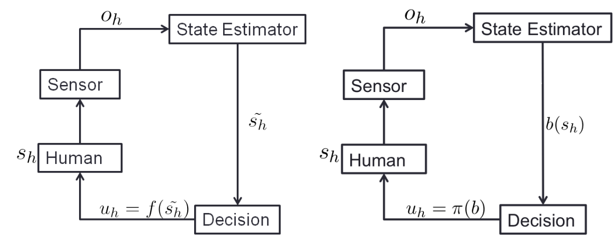
This paper is organized as follows. Section II begins with a review of POMDP and describes how a HITL system is modeled as a POMDP with factorized transition probability and observation probability. Section III shows the advantages of POMDP framework using a case study with simulation results. Finally, we conclude and highlight the key challenge of this framework in Section IV.

II. POMDP for Human-in-the-Loop System

As mentioned before, reasoning about a human's intent or physiological state is important in a HITL system. Instead of hacking into the human brain, we would want to estimate their intents by observing their actions. As shown in Fig. 1a, conventional HMM models decouple the state estimation process and the decision making process. They model the human behavior as a HMM and estimate the internal states of the human from observations[19]. Based on the estimation, a controller will give feedback to the human. As shown in Fig. 1b, POMDP estimates the probability distribution over the human's possible state $b(s_h)$ rather than just giving a single state estimation. A decision is then made based on the distribution of the hidden states. This allow the system to decide to take an action to reduce uncertainty in the human's state or provide assistance to the human. In the conventional HMM framework, the decision is made from single estimation so it will not reduce the uncertainty in the human's state. POMDPs provide an integrated model to incorporate hidden states, observations, and control actions, which perfectly describes the nature of a HITL system. The following will show how we model a HITL system as a POMDP.

A. Review of POMDP

Definition 1: A POMDP is defined as a tuple $(\mathcal{S},\mathcal{U},\mathcal{O},\mathcal{T},\Omega,R)$, where $\mathcal{S}$ is a set of hidden states; $\mathcal{U}$ is a set of actions or controls; $\mathcal{O}$ is a set of observations; $\mathcal{T}$ defines the conditional transition probability $P(s'|s,u)$; Ω defines the conditional observation probability $P(o'|s',u)$ and $R:\mathcal{S}\times\mathcal{U}\to\mathbb{R}$ is the reward function.



(a) Conventional (b) POMDP

Fig. 1: Conventional HMM model and POMDP model for human-machine interaction

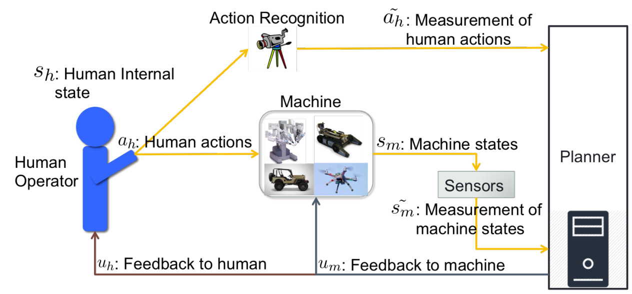


Fig. 2: Block diagram of a human-in-the-loop system

Definition 2: A belief state b is a probability distribution over $\mathcal{S}$. We let $b(s)$ denote the probability of being in a particular state $s\in\mathcal{S}$.

Definition 3: A policy π is a mapping from belief state to the set of controls $\pi(b)\in\mathcal{U}$.

At each time step, the world is in some hidden state $s\in\mathcal{S}$. The agent chooses a control $u\in\mathcal{U}$ according to her policy $u=\pi(b)$, receives a reward $R(s,u)$, and then the world state transitions to $s'\in\mathcal{S}$ with an observation $o'\in\mathcal{O}$. The belief is then updated as follows:

$$\begin{aligned} b'(s) = P(s'|o',u,b) &= \frac{P(o'|s',u,b)P(s'|u,b)}{P(o'|u,b)} \\ &= \frac{P(o'|s',u)\sum_{s\in\mathcal{S}}P(s'|s,u)b(s)}{P(o'|u,b)} \end{aligned} \tag{1}$$

where $P(o'|u,b)$ can be treated as a normalizing factor.

Given a policy π, the expected cumulative reward starting from belief state b_t is:

$$V^{\pi}(b_t) = \mathbb{E}[\sum_{n=t}^{\infty}\gamma^{(n-t)}R(s_n,u_n)]$$

where $0\le\gamma\le 1$ is a discount factor. The goal of POMDP is to find an optimal policy π^* to maximize V^{π}. Solving POMDP is often computationally intractable but there exist techniques[9][13][12] to obtain an approximate solution in practice.

B. Human-in-the-Loop Modeling

Figure 2 is the block diagram of a HITL system. The variables are defined as follow:

- $s_h \in S_h$, the set of internal states of the human, which can be the human's intent (e.g. turn right, turn left) or physiological state (e.g. fatigue, awake.)
- $a_h \in A_h$, the set of the human's actions (e.g. head pose, control to a joystick.)
- $s_m \in S_m$, the set of states of the machine(e.g. velocity, position.)
- $\tilde{a_h} \in O_{a_h}$, the set of observations of the human's actions.
- $\tilde{s_m} \in O_{s_m}$, the set of observations of the machine's states.
- $u_h \in U_h$, the set of control feedbacks for the human (e.g. warning, augmenting information.)
- $u_m \in U_m$, the set of control feedbacks for the machine (e.g. emergency brake, turning the steering wheel.)

We assume all the above sets are finite. In the diagram in Fig. 2, the human has an internal state, s_h, which could be her intent or the goal she wants to achieve, or her physiological state like fatigue, anger, being drunk, etc. Depending on her internal state, she will take an action, a_h to achieve her intent. The human, for example, may turn her head to check the left lane and then turn the steering wheel if she wants to switch to the left lane. Some human actions are control inputs to the machine and therefore the state of the machine s_m will change over time. One should note that not all human actions are control inputs of the machine. Some actions, like checking the left lane, may just be common behaviors for a specific task. There will be sensors to measure both the human's action and the state of the machine. We denote the measurement of the human's action as $\tilde{a_h}$ and the measurement of the state of the machine as $\tilde{s_m}$. The human-in-the-loop planner uses the measurements and its previous feedback as inputs, estimates a probability distribution over the hidden states, s_h, a_h and s_m and then decides an optimal feedback u_h to give to the human, and a control feedback u_m to apply to the machine.

The above process iterates for each time step and therefore the whole process can be viewed as a POMDP with a set of hidden states $S_h \times A_h \times S_m$, a control set $U_h \times U_m$, an observation set $O_{a_h} \times O_{s_m}$ and a transition probability

$$\begin{aligned}&P(s_h', a_h', s_m' | s_h, a_h, s_m, u_h, u_m)\\ =&P(s_h'|s_h, a_h, s_m, u_h, u_m) \times P(a_h'|s_h', s_h, a_h, s_m, u_h, u_m) \times\\ &P(s_m'|s_h', a_h', s_h', a_h, s_m, u_h, u_m)\end{aligned}$$

The above factorization is simply based on the chain rule in probability. Although the transition probability seems to be complicated, we can simplify it by making some reasonable conditional independence assumptions.

The first conditional independence assumption is that the internal state of the human only depends on her previous internal state, the state of the machine, and the feedback to the human. That is:

$$P(s_h'|s_h, a_h, s_m, u_h, u_m) = P(s_h'|s_h, s_m, u_h) \qquad (2)$$

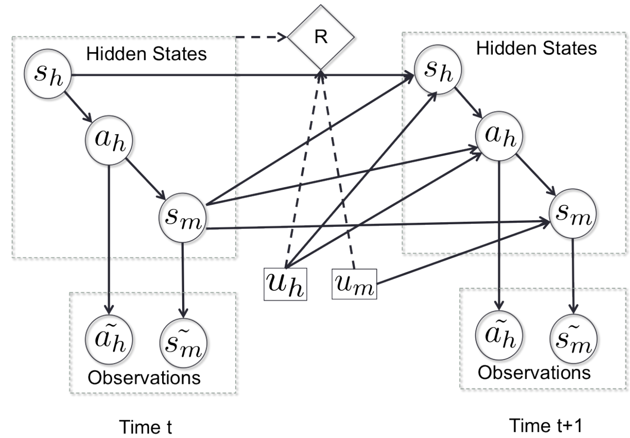


Fig. 3: Dynamic Bayesian Network representation of the HITL POMDP

which we will call it the **human internal state model**. The human internal state model describes how the human's state changes over time. One may note that the human's intent does not have to change at all time steps. For example, while controlling a robotic arm to take one of the objects on the table, the target object the user intents to take rarely changes during the whole process.

The second assumption is that the human's action is only based on her own internal state, the state of the machine and the feedback to the human, i.e.

$$P(a_h'|s_h', s_h, a_h, s_m, u_h, u_m) = P(a_h'|s_h', s_m, u_h) \qquad (3)$$

which we will call it the **human action model**. The human is taking action in order to achieve her own goal given the current machine state and our feedback.

The final assumption on the transition probability is that the state of the machine only depends on the human's action, previous machine state and the feedback to the machine, i.e.

$$P(s_m'|s_h', a_h', s_h', a_h, s_m, u_h, u_m) = P(s_m'|a_h', s_m, u_m) \qquad (4)$$

which we will call it the **machine dynamic model**. The machine dynamic model may come from machine's kinematic model or dynamics model.

In summary,

$$\begin{aligned}&P(s_h', a_h', s_m'|s_h, a_h, s_m, u_h, u_m)\\ =&P(s_h'|s_h, s_m, u_h)P(a_h'|s_h', s_m, u_h)P(s_m'|a_h', s_m, u_m) \qquad (5)\end{aligned}$$

Equation (5) defines the transition probability in the HITL POMDP.

In the **observation model**, we assume that the observations of the human's action and the state of the machine only depend on the actual action of the human and the actual state of the machine respectively:

$$P(\tilde{a_h}', \tilde{s_m}'|s_h', a_h', s_m', u_h, u_m) = P(\tilde{a_h}'|a_h')P(\tilde{s_m}'|s_m') \qquad (6)$$

Figure 3 summarizes the HITL POMDP model as a dynamic Bayesian network.

In order to obtain the models, we can either learn the models from data or handcraft them based on prior knowledge. The human internal state model and the human action model can be estimated from annotated data of sequence of interactions[11]. The machine dynamic model can be either obtained from system identification, or directly from first principles. For example, we could assume the resulting machine dynamics have the form

$$s'_m = f(s_m, a_h, u_m) + w$$

where w is the noise. Finally, the observation model comes from the accuracy of the senor system.

The design of the reward function $R(s_h, a_h, s_m, u_h, u_m)$ depends on our objective. For example, if the objective is to enhance safety, the reward in safe states should be high while the reward in unsafe states should be small. Of course a similar machinery can be applied when we would like multiple objectives in a HITL system. In our simulations next section, for example, we want to both promote safety and minimize interferences so we penalize interference from u_h and u_m while giving high rewards in safe states.

III. Simulation Results

In this section we present simulation results to illustrate the application of our proposed framework. According to the AAA Foundation for Traffic Safety, an estimated 13.1% of crashes that resulted in a person being admitted to a hospital, and 16.5% of fatal crashes involved a drowsy driver[15]. In this example, we assume the objective is to keep a car in a single lane, but the driver may be drowsy.

A. HITL POMDP for drowsy driver

The driver has two internal states:

$$S_h = \{\text{Awake, Sleepy}\}.$$

Depending on these two states, the driver's eyes could be open or closed. At the same time, the driver is driving the car to maintain the car in the middle of the lane, so we define the human actions as $A_h = A_{h1} \times A_{h2}$, where

$$\begin{aligned} A_{h1} =& \{\text{Eyes open, Eyes closed}\} \\ A_{h2} =& \{\text{Steer left = -1, Steer right = +1,} \\ & \text{Steer straight = 0, Do nothing}\}. \end{aligned}$$

We discretize the horizontal position of the car on the lane:

$$S_m = \{\text{-2, -1, 0, +1, +2, Off the lane}\}.$$

where -2 is the left most, 0 is the middle and $+2$ is the right most of the lane. The feedback to the human is a warning signal reminding the human to wake up or be careful,

$$U_h = \{\text{Warning on, Warning off}\}.$$

Assume the car has a driver assisting function that can take over the control of the steering wheel and therefore, the feedback to the machine is:

$$U_m = \{\text{Steer left = -1, Steer right = +1, Do nothing = 0}\}.$$

There are sensors to detect the human's actions and the machine states, so the observations are

$$\begin{aligned} O_{a_h} =& \{\text{Eyes open, Eyes closed}\} \times \\ & \{\text{Steer left, Steer right, Steer straight or do nothing}\} \end{aligned}$$

and $O_{s_m} = \{$-2, -1, 0, +1, +2, Off the lane$\}$.

As shown in (5), the transition probability depends on the human internal state model, the human action model and the machine dynamics model. For the sake of simplicity, we handpick the probabilities in this simulation. Although it would be more realistic to learn the models from data, the learning process is not trivial and we leave it to our future work. The human internal state model and the human action model is illustrated in Fig. 4. The nodes in Fig. 4 represent the states while the numbers on the edges represent the transition probabilities conditioned on their parent nodes.

The machine dynamics model is:

$$\begin{aligned} s'_m =& Maneuver(s_m, a_{h2}, u_m, w) \\ =& \begin{cases} s_m + a_{h2} & \text{if } a_{h2} \neq \text{Do nothing} \\ s_m + u_m & \text{if } a_{h2} = \text{Do nothing \& } u_m \neq \text{Do nothing} \\ s_m + w & \text{otherwise} \end{cases} \end{aligned}$$

where $a_{h2} \in A_{h2}$ is the human's input and $u_m \in U_m$ is the feedback to the machine. $w \in \{\text{Steer left, Steer right, Do nothing}\}$ with probability $\{0.2, 0.2, 0.6\}$ is acted as noise. Any $s_m \notin [-2, +2]$ is considered as "Off the lane". In function $Maneuver(\cdot,\cdot,\cdot,\cdot)$, a_{h2} has a higher control priority than u_m and w. To make the simulation more realistic, we use an estimation of a_{h2} from the system instead of the true a_{h2}, which is actually hidden. w takes effect as the time both the driver and controller are not maneuvering the car, where the road may have a left curve or a right curve. For example, the car entering a left curve without maneuver has the same effect as $w =$ "Steer right".

In the observation model, we assume all sensors have accuracy $P_{acc} = 0.9$

According to the safety condition, we define the reward function as

$$R(s_m, u_h, u_m) = R_1(s_m) + R_2(u_h) + R_3(u_m)$$

where $R_1(s_m)$ is as follow:

	-2	-1	0	+1	+2	Off
$R_1(s_m)$	5	10	20	10	5	0

We also penalize the intervention to human and machine:

	Warning on	Warning off
$R_2(u_h)$	-5	0

	Steer left	Steer right	Do nothing
$R_3(u_m)$	-5	-5	0

We solve the above POMDP problem with the SymbolicPerseus package[10]. Then we use the optimal policy π^* in our simulation. At each time step, we decide the optimal control $u^*_t = \pi^*(b_t)$, sample the next state and observations

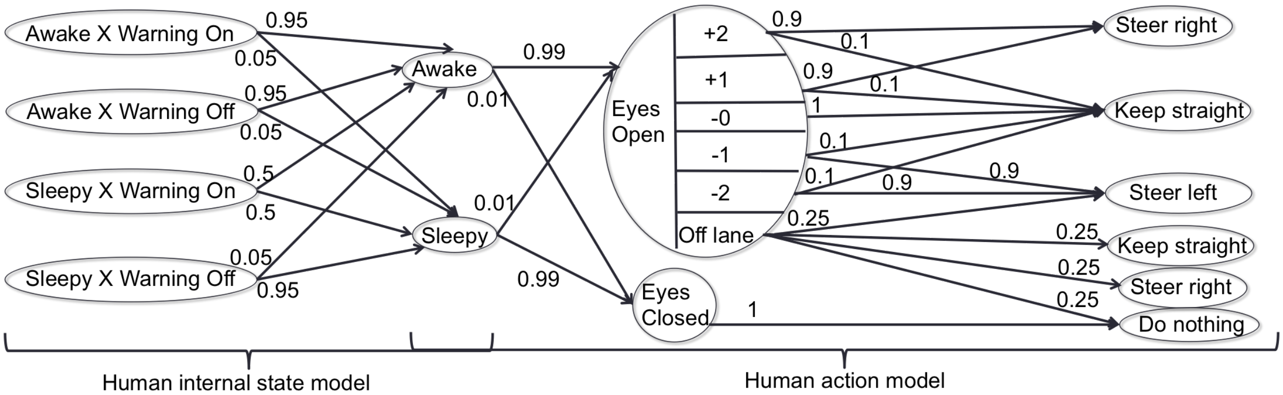


Fig. 4: A diagram representation of the transition probability of human internal state model and human action model

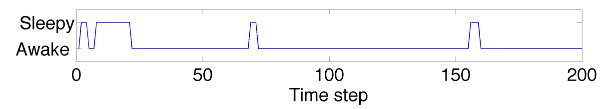
based on the transition function and observation function, and then update the current belief b_{t+1} using Eq. (1).

Figure 5 shows the simulation results. Figure 5a is the actual hidden internal state of the human and Fig. 5f shows the marginal belief of the human's internal state at each time step. Though there are false alarms because of the measurement error as shown in Fig. 5b and 5c, the probability $P(s_h = \text{Awake})$ decreases whenever the actual state is "Sleepy", which means the system is able to reason about the internal state of the human. Figure 5d shows the optimal feedback to the human, u_h, which conforms with our intuition that when $P(s_h = \text{Awake})$ goes down to some threshold, the warning system will turn on in order to keep the driver awake. Again, there are some false alarms due to the measurement error, but they are less than using a policy only based on the measurements. In this simulation, we only got 2 false alarms, whereas if we estimate the internal state of the human just relying on the sensor measurements, we will get 25 false alarms. Figure 5e shows one of the human actions, a_{h2}, and the feedback to the vehicle u_m. We can see that the optimal feedback to the vehicle obtained from our optimal policy drives the vehicle back to the middle of the lane in order to maintain safety. We can also see that given this POMDP framework, we can solve an optimal policy that automatically balances between when to give feedback to the human and when to give feedback to the machine.

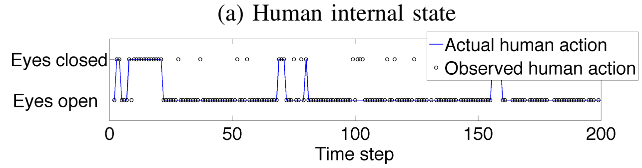
This framework also allows us to keep track of the probability of unsafe state, which is $P_{Unsafe} = \sum_{s_t \in Unsafe} b_t(s_t)$, where $s_t = (s_h^t, a_h^t, s_m^t)$ and the unsafe set $Unsafe = \{(s_h, a_h, s_m) | s_m = \text{Off the lane}\}$. Figure 5g shows the probability of the unsafe state, remaining low in the whole process.

To show the benefit of POMDP in long-term planning, we compare the optimal policy with two other policies. One is a greedy policy: when the controller observes the driver's eyes closed, the warning will be turned on. At the same time, the feedback to the vehicle will be generated to drive the vehicle towards the middle according to observed vehicle state $\tilde{s_m}$. The other one is a minimal unsafe probability policy:

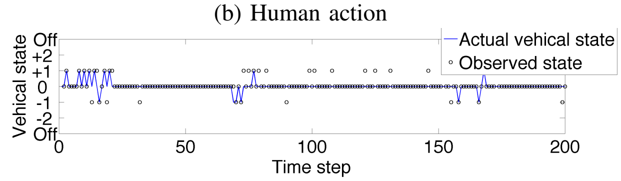
$$u^* = \arg\min_u \sum_{s_{t+1} \in Unsafe} P(s_{t+1}|s_t, u) b_t(s_t)$$



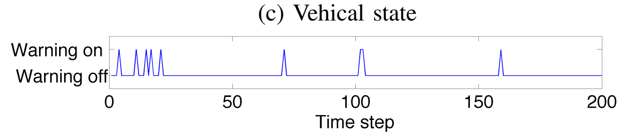
(a) Human internal state



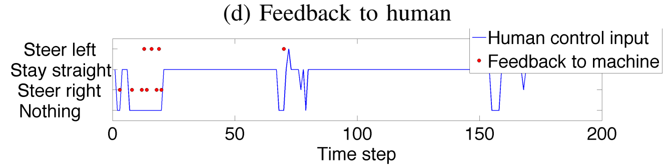
(b) Human action



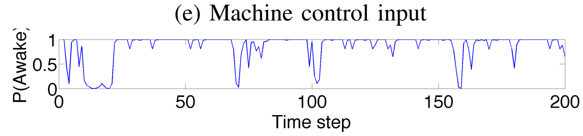
(c) Vehical state



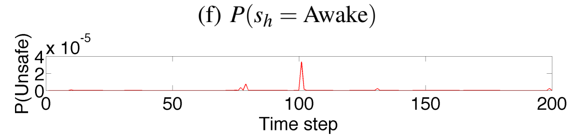
(d) Feedback to human



(e) Machine control input



(f) $P(s_h = \text{Awake})$



(g) P_{Unsafe}

Fig. 5: Simulation results

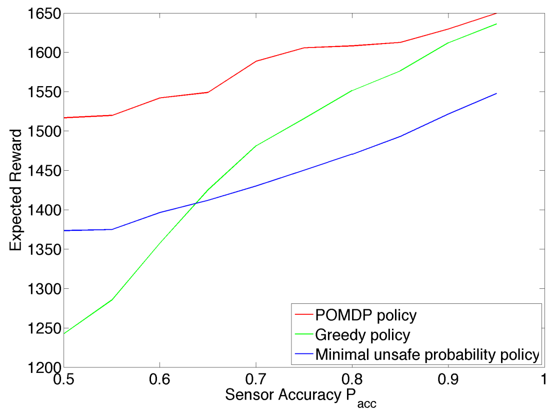


Fig. 6: Comparison of different strategy with POMDP policy

Figure 6 shows the average reward for the three different policies corresponding to the accuracy of sensors. The POMDP policy outperforms the other two policies. The difference between these policies are more in low sensor accuracy than in high sensor accuracy. This result is not surprising because the greedy policy is optimal when all states are observable, i.e. $P_{acc} = 1$. The reward of the minimal unsafe probability policy is the most conservative policy that the warning signal is turned on frequently to remind the driver, resulting in a low reward. The reward of the minimal unsafe probability policy, however, is larger than the greedy policy in low P_{acc} cases. It is because when the sensor is not accurate, it is very likely to make wrong decision just based on observations and therefore leading the vehicle into unsafe states and resulting in a low reward. POMDP policy enhances safety and minimizes intervention at the same time so it has the highest reward.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a novel POMDP framework for human-in-the-loop control systems. It is an initiating work in formalizing HITL control systems. We have shown that by imposing some reasonable conditional independent assumptions, we can succinctly unify stochastic models for the human and machine—the human internal state model, the human action model, and the machine dynamic model—into a single framework supporting global optimization for long-run planning. This paper has shown various benefits of using POMDP in HITL modeling: (1) the abilities of reasoning human internal state; (2) handling the error from observations; and (3) balancing the trade-off between the feedback to the human and the feedback to the machine.

The key challenge of advancing this framework is that POMDP can only deal with discrete states, while most machine states are described in a continuous state space. One way to handle it is to discretize the continuous state space. However, when the state space is too large or the discretizing resolution is too small, discretization is not practical because it makes solving POMDP intractable. We would like to find a method to solve the HITL POMDP problem with hybrid state in the future.

REFERENCES

[1] Sterling J Anderson, Steven C Peters, Tom E Pilutti, and Karl Iagnemma. An optimal-control-based framework for trajectory planning, threat assessment, and semi-autonomous control of passenger vehicles in hazard avoidance scenarios. *International Journal of Vehicle Autonomous Systems*, 8(2):190–216, 2010.

[2] Frank Broz, Illah Nourbakhsh, and Reid Simmons. Planning for human–robot interaction in socially situated tasks. *International Journal of Social Robotics*, 5(2):193–214, 2013.

[3] R. Chipalkatty, H. Daepp, M. Egerstedt, and W. Book. Human-in-the-loop: Mpc for shared control of a quadruped rescue robot. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 4556–4561, Sept 2011.

[4] J. Hoey, P. Poupart, A. von Bertoldi, T. Craig, C. Boutilier, and A. Mihailidis. 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, 2010.

[5] Ming Li and A.M. Okamura. Recognition of operator motions for real-time assistance using virtual fixtures. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003. Proceedings. 11th Symposium on*, pages 125–131, March 2003.

[6] Sirajum Munir, John A. Stankovic, Chieh-Jan Mike Liang, and Shan Lin. Cyber physical system challenges for human-in-the-loop control. In *Presented as part of the 8th International Workshop on Feedback Computing*, Berkeley, CA, 2013. USENIX.

[7] R. Parasuraman, T.B. Sheridan, and Christopher D. Wickens. A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 30(3):286–297, May 2000.

[8] Alex Pentland and Andrew Lin. Modeling and prediction of human behavior. *Neural Computation*, 11:229–242, 1995.

[9] Joelle Pineau, Geoffrey Gordon, and Sebastian Thrun. Point-based value iteration: An anytime algorithm for pomdps. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1025 – 1032, August 2003.

[10] Pascal Poupart. Exploiting structure to efficiently solve large scale partially observable markov decision processes, 2005.

[11] D. Sadigh, K. Driggs-Campbell, A. Puggelli, W. Li, V. Shia, R. Bajcsy, A. L. Sangiovanni-Vincentelli, S. S. Sastry, and S. A. Seshia. Data-driven probabilistic modeling and verification of human driver behavior. 2014.

[12] Trey Smith and Reid Simmons. Point-based pomdp algorithms: Improved analysis and implementation. In *in Proceedings of Uncertainty in Artificial Intelligence*, pages 542–555, 2005.

[13] Matthijs T. J. Spaan and Nikos Vlassis. Perseus: Randomized point-based value iteration for pomdps. *Journal of Artificial Intelligence Research*, 24:195–220, 2005.

[14] W. Takano, A. Matsushita, K. Iwao, and Y. Nakamura. Recognition of human driving behaviors based on stochastic symbolization of time series signal. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 167–172, Sept 2008.

[15] Brian C Tefft and AAA Foundation for Traffic Safety. Asleep at the wheel: The prevalence and impact of drowsy driving. 2010.

[16] R. Vasudevan, V. Shia, Yiqi Gao, R. Cervera-Navarro, R. Bajcsy, and F. Borrelli. Safe semi-autonomous control with enhanced driver modeling. In *American Control Conference (ACC), 2012*, pages 2896–2903, June 2012.

[17] Zheng Wang, A. Peer, and M. Buss. An hmm approach to realistic haptic human-robot interaction. In *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, pages 374–379, March 2009.

[18] Jason D Williams and Steve Young. Partially observable markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422, 2007.

[19] C. Zhu and Weihua Sheng. Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(3):569–573, May 2011.