

A Decision-Theoretic Approach to Cooperative Control and Adjustable Autonomy

Abdel-Ilhah Mouaddib¹, Shlomo Zilberstein², Aurélie Beynier³, Laurent Jeanpierre⁴

Abstract. Cooperative control can help overcome the limitations of autonomous systems (AS) by introducing a supervision unit (SU) (human or another system) into the control loop and creating adjustable autonomy. We present a decision-theoretic approach to accomplish this using Mixed Markov Decision Processes (MI-MDPs). The solution is an optimal plan that tells the AS what actions to perform as well as when to request SU attention or transfer control to the SU. This provides a varying degree of autonomy, particularly suitable for robots exploring a domain with regions that are too complex or risky for autonomous operation, or intelligent vehicles operating in heavy traffic.

1 Introduction

Despite rapid progress with the perceptual, planning and control abilities of autonomous systems, they still require supervision in many application areas. This has led to growing interest in models for *adjustable autonomy*, *human-robot teaming* or *mixed control*. We are interested in domains in which a supervision unit (SU), often a human operator, can take over control when the situation is too complex for the autonomous system (AS) [1]. While the SU may be able to perform each task by manually controlling the system, this would normally result in a time-consuming, costly operation. Our goal is therefore to develop a general framework for SU-AS teaming, to optimize performance and reduce the SU workload, costs, fatigue-driven errors and risks.

Developing a general framework for SU-AS teaming presents several challenges, most notably how to selectively decide when to use supervision and when to operate autonomously so as to optimize performance. We propose a decision-theoretic approach to this problem using a new model called Mixed Markov Decision Process (MI-MDP). In this model, an AS and a SU can each control a process using *different* sets of actions. Each action is characterized by a probabilistic transition model and cost. In addition to the state of the environment, MI-MDPs also maintain a state variable that determines the current level of autonomy. The transition between fully autonomous operation and supervision control may require multiple time steps, during which the SU increases the level of attention paid to the autonomous system. In practice, the level of attention could measure the *frequency* of attending to the system (e.g. once every five minutes) or the maximum *neglect time* (e.g. up to 10 minutes).

2 Formal Framework

2.1 The SU-AS Decision Problem

The formal framework is based on a centralized planning algorithm that creates a coordinated plan for the two possible controllers (SU and AS), only one of which controls the system at each point in time. The plan must indicate which controller is in charge, what action should be performed, and what level of attention has to be paid by the other to the process. In the case of the SU-AS interaction, we assume that control can be transferred to the AS at any time. However, SU operators need several time steps before they can control the process. The level of attention paid to the activity can depend on the complexity of the situation (the more complex the situation, the less the AS can be neglected). We generally assume that the SU must pay a high level of attention in complex situations and low level of attention when the situation is simple (the level of attention could reflect the amount of time the SU spends overseeing the AS). The level of attention is modeled by a discrete set and it can grow only by one level at a time. The SU can control the robot only at the highest level of attention. Therefore, it could take some time before the level of attention is sufficiently high for the SU to control the system.

This decision problem can be specified using two separate controllers, each defined as a standard MDP.

- **AS Control:** formalized by $\langle \mathbf{S}, \mathbf{A}, \mathbf{R}, T_r \rangle$, where $\mathbf{A}$ and T_r are respectively the actions of the robot and the transition model.
- **SU Control:** formalized by $\langle \mathbf{S}, \mathbf{H}, \mathbf{R}, T_h \rangle$, where $\mathbf{H}$ and T_h are respectively the actions of the SU and the transition model.

Note that $\mathbf{S}$ and $\mathbf{R}$ are identical in these MDPs and represent the state space and the objective function, respectively. Indeed, the AS and the SU act in the same environment (i.e., robot and surgeon perform an operation on the same patient) and have the same objective (safety of the patient and reduced costs), but they have different actions with different transition models.

From these models we can derive a pure autonomous strategy and a pure tele-operation strategy. The first strategy can lead to poor performances because of the imprecise model and limited abilities of the AS, while the second model involves more expensive actions. Consequently, it is beneficial to employ a mixed sequential decision paradigm that can switch between autonomous control with different SU attention levels and SU control.

2.2 Mixed MDPs

The mixed MDPs model is created from the two individual MDPs as follows. We have a new set of actions corresponding to the union of the AS and SU actions and a new state space representing not only

¹ GREYC/université de Caen Basse-Normandie, France, email: mouaddib@info.unicaen.fr

² Computer Science department, University of Massachusetts at Amherst USA, email: shlomo@cs.umass.edu

³ University of Paris 6, France, email: Aurelie.Beynier@lip6.fr

⁴ GREYC/université de Caen Basse-Normandie, France, email: laurent@info.unicaen.fr

to the state of the underlying system being controlled, but also the level of SU attention. The state space of the AS is represented by S while the level of SU attention is represented by $I = \{0, 1, \dots, n\}$. Consequently, the new state space is $S \times I$ and the new action space is $(A \times \{+, -, \emptyset\}) \cup (H \times \{\emptyset, 0, 1, \dots, n-1\})$

The actions in this model have two components represented as $(a, *)$ or $(h, *)$, which mean respectively that the AS controls the process (autonomy) or the SU controls the process (tele-operation). The second component, $*$, could be $+, -, k \in \{0, \dots, n-1\}$ or $\emptyset$, which represent an increase, decrease, go to level k or no change in the attention level respectively. More formally, an MI-MDP is defined by a tuple $\langle \mathbf{S}, \mathbf{A}, \mathbf{R}, \mathbf{T} \rangle$ where:

- $\mathbf{S}$ is the *state space*. An MI-MDP state (s, i) includes the system state, s , and the level of attention, i :

$$\mathbf{S} = \{(s, i) | 0 \leq i \leq n, s \in S\}.$$

The initial state is (s_0, l_0) , where l_0 is any level of attention.

- $\mathbf{A}$ is the *action space* defined as follows:

$$\mathbf{A} = \{(x, y) \in (A \times \{+, -, \emptyset\}) \cup (H \times \{\emptyset, 0, 1, \dots, n-1\})\}$$

The first component, a or h , is any AS or SU action. The second component represents the requested relative change $(\emptyset, +, -)$ or absolute $(0, \dots, n-1)$ level of SU attention.

- $\mathbf{R}$ is the combined *reward function* assigned to each state (s, i) and action (x, y) such that: $\mathbf{R}((s, i), (x, y)) = R(s) - \text{Cost}(x, y, i)$ Where R is the immediate reward received in a state, s , and Cost is the combined cost of the action and the level of SU attention.
- $\mathbf{T}_l$ is the combined *transition model* based on the transition models of the SU and the AS, and factoring the uncertainty about the change in the level of attention. Formally, it is defined as follows:

$$\mathbf{T}_l : \{0, 1, \dots, n\}^2 \times \{+, -, \emptyset, 0, 1, \dots, n-1\} \rightarrow [0, 1]$$

where $\mathbf{T}_l(i', i, *)$ represents the probability of transition from i to i' when acting with $* \in \{+, -, \emptyset, 0, 1, \dots, n-1\}$. This probability allows us to formalize situations in which SU operators may refuse to change their level of attention or cannot change it because of environmental constraints such as high workload, low level of staffing, or external causes that are not modeled by the system.

This concludes the definition of an MI-MDP. A value function over the states of the MI-MDP (equivalent to the Bellman equation) can be defined as follows:

$$V((s, i)) = \max_{(x, y)} \left[\mathbf{R}((s, i), (x, y)) + \sum_{s', i'} \mathbf{T}((s', i'), (x, y), (s, i)) V((s', i')) \right]$$

The optimal policy, π , maps each state (s, i) to the action (x, y) that maximizes the above value for that state. As with MDPs, MI-MDPs can be optimized over some finite-horizon or solved as infinite-horizon problems using standard MDP algorithms.

3 Illustration of the Approach with Mobile Robots

The development of the MI-MDP model was inspired by real-world applications such as military personnel supervising the operation of multiple mobile robots (or AUVs). The robots could operate autonomously most of the time, but may need to be tele-operated manually in certain situations.

	0	1	2	3	4	5	6
0	P						
1							
2							
3					R		
4							
5							
6			P				P

Figure 1. Illustration of the target environment

To illustrate the MI-MDP approach, we have implemented it for the scenario depicted in Figure 1. This scenario involves a robot that navigates in a “messy” environment represented as a 7×7 grid where gray cells denote areas where it is difficult for the robot to navigate autonomously. The robot can ask, via Wi-Fi communication, for the attention of a human operator. When the robot asks for more attention, a graphic interface notifies the human of the robot’s request. Human operators can acknowledge the robot request and then increase their attention, or they can ignore the robot’s request. When the human attention is at the highest level, the human can take over the control of the robot. The robot starts at cell $(4, 3)$ and it has to take pictures at some locations (cells $(0, 0)$, $(2, 6)$ and $(6, 6)$). The level of human attention is initialized to 0 (i.e., the robot is initially fully autonomous). A validation on a Koala robot with an operator has been successfully implemented where the robot has navigated autonomously in free area and asked for assistance in messy areas.

4 Discussion and Concluding Remarks

The Mixed Markov Decision Process (MI-MDP) offers a new way to formalize *mixed control problems* and create *adjustable autonomy*. It addresses effectively many of the challenges in designing systems composed of a supervision unit (SU) and autonomous unit (AS). The model is general enough to capture characteristics of SU-AS teams and the interaction between these two components. One instance of such systems is the creation of human-robot teams, for which MI-MDPs capture formally characteristics that have been presented earlier more intuitively [1]. The resulting system is robust in terms of its ability to address different degrees of complexity in the environments or varying cost of SU attention (i.e. human attention). Unlike some previous approaches that focused on the question of *when* to transfer control between the robot and human [2, 3], our approach produces an optimal policy to decide *who* should control the system as well as *what* they should do. Our approach applies to situations in which the SU does not follow the recommended actions. The SU could choose any action including transferring control back to the AS, which can immediately resume operation based on the computed plan.

REFERENCES

- [1] J.W. Crandall and M.A. Goddich, ‘Validating human-robot interaction schemes in multitasking environments’, *IEEE transaction Systems, Man, and Cybernetics, Part A : Systems and Humans*, **35**(4), 438–449, (2005).
- [2] P. Scerri, D.V. Pynadath, and M. Tambe, ‘Adjustable autonomy in real-world multi-agent environments’, in *International Conference on Autonomous Agents (Agent’01)*, pp. 300–307, (2001).
- [3] P. Scerri, D.V. Pynadath, and M. Tambe, ‘Why the elf acted autonomously: Towards a theory of adjustable autonomy’, in *AAMAS*, pp. 857–864, (2002).