Multiagent POMDPs with Asynchronous Execution

(Extended Abstract)

João V. Messias
Inst. for Systems and Robotics
Instituto Superior Técnico
Lisbon, Portugal
jmessias@isr.ist.utl.pt

Matthijs T. J. Spaan
Delft University of Technology
Delft, The Netherlands
m.t.j.spaan@tudelft.nl

Pedro U. Lima
Inst. for Systems and Robotics
Instituto Superior Técnico
Lisbon, Portugal
pal@isr.ist.utl.pt

ABSTRACT

The Multiagent POMDP (MPOMDP) framework provides well-known methods to model and solve fully communicative multiagent problems. However, the size of these models grows exponentially in the number of agents, and agents are required to act in synchrony. We show how these problems can be mitigated through an event-driven, asynchronous formulation of the MPOMDP dynamics. We can prove that the optimal value function in our framework is piecewise linear and convex, allowing us to extend a standard point-based solver to the event-driven setting. We also show how belief states can be updated at run-time in asynchronous domains. Our results show that asynchronous models scale better to larger domains than synchronous analogues, while retaining solution quality.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

General Terms

Algorithms, Theory

Keywords

Planning under uncertainty; cooperative multiagent systems

INTRODUCTION

Most existing approaches to multiagent decision making under uncertainty are grounded in the theory of Markov Decision Processes (MDPs), for example Decentralized Partially Observable MDPs (Dec-POMDPs) [2] which are intractable without communication (NEXP-Complete); and Multiagent MDPs (MMDPs) and POMDPs (MPOMDPs), which assume free communication between agents [3]. The latter framework is typical, for example, when dealing with teams of mobile robots, or in autonomous surveillance systems. However, the complexity of solving an MPOMDP is exponential in the number of agents [3]. This follows from the assumption of *synchronous* operation: at each decision step, the simultaneous observations of all agents need to be considered when selecting a new joint action.

The present work proposes Event-Driven MPOMDPs, an alternative description of the dynamics of multiagent decision making under uncertainty[1]. In our approach, agents must react to events, which are detected locally, and *asynchronously*, by each agent. Through the assumption of free communication, each local event triggers a joint observation, which is shared by the team. Since multiple events cannot occur simultaneously, the total number of joint observations in this model grows *linearly* in the number of agents (instead of exponentially), allowing these methods to scale better to larger scenarios. Furthermore, the processes through which events are detected are considered to be susceptible to false positive and false negative errors.

We can prove that the optimal value function is piecewise linear and convex (PWLC), allowing us to extend a point-based solver to the event-driven setting; and we propose a method for belief-state tracking at run-time for asynchronous agents. We evaluate our methods through simulated results, comparing the performance of an event-driven model to that of an equivalent synchronous MPOMDP.

EVENT-DRIVEN MPOMDPS

An MPOMDP is a tuple $\langle d, \mathcal{S}, \mathcal{A}, \mathcal{O}, T, O, R\rangle$ where: d is the number of agents; $\mathcal{S}$ is the state space; $\mathcal{A} = \times_i^d \mathcal{A}_i$ is a set of *joint* actions $\mathbf{a} = \langle a_1, a_2, \ldots, a_d\rangle$, and $\mathcal{A}_i$ the individual action set of agent i; $\mathcal{O} = \times_i^d \mathcal{O}_i$ is a set of joint observations $\mathbf{o} = \langle o_1, ..., o_d\rangle$; $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ is the transition function, such that $T(s, \mathbf{a}, s') = \Pr(s'|s, \mathbf{a})$; $O : \mathcal{A} \times \mathcal{S} \times \mathcal{O} \to [0, 1]$ is the observation function, such that $O(\mathbf{a}, s', \mathbf{o}) = \Pr(\mathbf{o}|\,\mathbf{a}, s')$; and $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the instantaneous reward function.

A *joint belief state* b, a probability distribution over $\mathcal{S}$, is maintained at each step in an MPOMDP, updated according to $\mathbf{a}$ and $\mathbf{o}$ [3]. Due to the influence of $\mathbf{o}$, agents must synchronize their observations before taking a joint decision. Our approach lifts this assumption of synchrony.

We view system "events" as state transition tuples $\langle s, \mathbf{a}, s'\rangle$. In this context, an Event-Driven MPOMDP is a model where decision episodes are triggered by events. Formally, it is a tuple $\langle d, \mathcal{S}, \mathcal{A}, \mathcal{O}, T, O, \mathcal{C}, R\rangle$, where: d, $\mathcal{S}$, $\mathcal{A}$, R are defined as in a standard MPOMDP; $\mathcal{O}$, the set of observa-

[1]Our definition and use of "events" differs from existing work [1], and concerns different purposes. There, events model interdependencies between agent policies in Dec-MDPs. Here, events are simply state changes: the system dynamics are driven by events.

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

tions, is defined as $\mathcal{O} = \cup_i^d \mathcal{O}_i$, implying that the local observation of any agent can be taken directly as if it were a "joint" observation, exploiting free communication; the observation function is $O : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \times \mathcal{O} \to [0, 1]$, such that $O(s, \mathbf{a}, s', \mathbf{o}) = \Pr(\mathbf{o}| s, \mathbf{a}, s')$. This allows for a description of false positive/false negative event detection rates; $\mathcal{C} : \mathcal{A} \times \mathcal{O} \to PS(A)\backslash\emptyset$, where $PS(A)$ is the power set of A, is a *constraint-generating* function which returns, for each pair $\langle \mathbf{a}, o \rangle$, a *constrained action set* $\mathcal{C}(\mathbf{a}, o) \subseteq \mathcal{A}$. This set represents the joint actions which are available to the agents at the onset of a decision episode, given that, at the previous step, the team of agents executed $\mathbf{a}$ and observed o.

The presence of $\mathcal{C}$ addresses the problem of unobservable events and false negatives—if all agents fail to detect the occurrence of an event, they won't be able to change their actions, even if the state has in fact changed. We associate the occurrence of such events with token observations f, and such that $\mathcal{C}(\mathbf{a}, f) = \mathbf{a}$. During planning, this approach allows the expected value of joint policies to remain accurate.

However, during plan execution, false negative detections are never experienced by an agent (by definition), and so agents must take into account the fact that the system can undergo several unobserved transitions between any two belief update steps. This implies that the standard belief update in an MPOMDP (c.f. [3]) cannot be directly applied, at run-time, to an event-driven model. For an infinite-horizon agent in an Event-Driven MPOMDP, given that the team is executing $\mathbf{a}$ and observing o in belief state $\hat{b}$, let $f \in \mathcal{O}$ represent false negative detections of events, and $H_o^{\mathbf{a}} : |\mathcal{S}| \times |\mathcal{S}| \to [0, 1]$ such that $H_o^{\mathbf{a}}(s', s) = T(s, \mathbf{a}, s')O(s, \mathbf{a}, s', o)$. We can show that the belief update step is then:

$$\hat{b}_o^{\mathbf{a}} = \frac{\left(H_o^{\mathbf{a}}(I - H_f^{\mathbf{a}})^{-1}\hat{b}\right)}{\mathbf{1}^T\left(H_o^{\mathbf{a}}(I - H_f^{\mathbf{a}})^{-1}\hat{b}\right)} \quad ,$$

iff for all eigenvalues λ of $H_f^{\mathbf{a}}$, $|\lambda_i| < 1$. If this isn't verified, the system has fully unobservable ($\Pr(f|\cdot) = 1$) loops, over which the belief state can't be tracked.

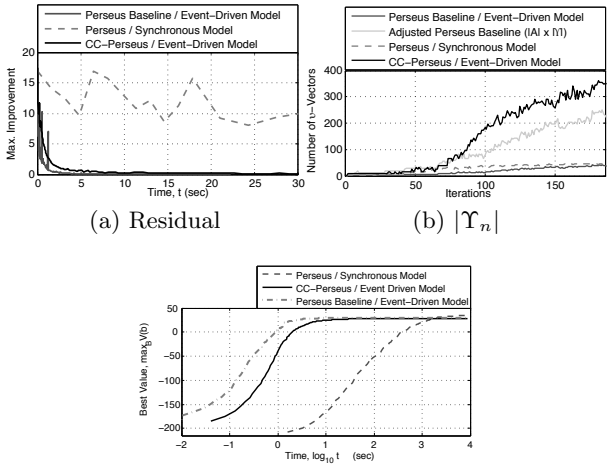
We can show that a value function for an Event-Driven MPOMDP in the presence of action constraints is PWLC, that is, for finite n, the optimal value function V_n^* can be written as:

$$V_n^*(b) = \max_{\upsilon_n \in \Upsilon_n} \upsilon_n \cdot b \quad ,$$

where Υ_n is a set of $|S|$-dimensional vectors. This enables the use of dynamic programming techniques to calculate (or approximate) an optimal policy. We propose *Constraint-Compliant* PERSEUS (CC-PERSEUS), an adaptation of the PERSEUS randomized point-based algorithm to Event-Driven MPOMDPs, with the following modifications with respect to the latter: we implement the backup stage for belief states [4], so that, at stage n, and for each $\langle \mathbf{a}, o \rangle$, only vectors in V_{n-1} associated to actions in $\mathcal{C}(\mathbf{a}, o)$ are considered; we explicitly maintain separate $Q-$value functions at each stage ($V_n = \cup_{\mathcal{A}} Q_n^{\mathbf{a}}$) and ensure that each $Q_n^{\mathbf{a}}$ is never empty.

RESULTS

The performance of CC-PERSEUS on an event-driven problem ($d = 4$, $|\mathcal{S}| = 216$, $|\mathcal{A}| = 8$, $|\mathcal{O}| = 10$) was compared to that of standard PERSEUS on an equivalent synchronous model ($d = 4$, $|\mathcal{S}| = 216$, $|\mathcal{A}| = 54$, $|\mathcal{O}| = 256$). The results,



(a) Residual (b) $|\Upsilon_n|$

(c) Value

Figure 1: (a) Residual difference $\max_{\mathcal{B}}\{V_n(b) - V_{n-1}(b)\}$**. (b) Size of the value function,** $|\Upsilon_n|$**. (c) Real time convergence.**

represented in Figure 1, show a computational advantage of the event-driven model over its alternative by more than one order of magnitude. This follows from the exponentially smaller number of observations that need to be considered at each step.

CONCLUSIONS

We propose a novel, asynchronous modeling approach for multiagent decision-making under partial observability. We adapt a common POMDP-solving algorithm to function in an event-driven paradigm, and show how agents can track belief states at run-time in the presence of false negative observations. Empirical comparison shows that event-driven MPOMDPs allow more compact representations than what is possible through standard MPOMDPs in the same domains, resulting in considerable computational savings.

ACKNOWLEDGMENTS

This work was funded in part by Fundação para a Ciência e a Tecnologia (ISR/IST pluriannual funding) through the PIDDAC Program funds and was supported by the Carnegie Mellon - Portugal Program (project CMU-PT/SIA/0023/2009). J.M. was supported by a PhD Student Scholarship, SFRH/BD/44661/2008, from the Portuguese FCT POCTI programme. M.S. is funded by the FP7 Marie Curie Actions Individual Fellowship #275217 (FP7-PEOPLE-2010-IEF).

REFERENCES

[1] R. Becker, S. Zilberstein, and V. Lesser. Decentralized Markov decision processes with event-driven interactions. In *AAMAS*, pages 302–309. IEEE Computer Society, 2004.

[2] D. Bernstein, R. Given, N. Immerman, and S. Zilberstein. The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research*, 27(4):819–840, 2002.

[3] D. Pynadath and M. Tambe. The communicative multiagent team decision problem: Analyzing teamwork theories and models. *Journal of Artificial Intelligence Research*, 16:389–423, 2002.

[4] M. T. J. Spaan and N. Vlassis. Perseus: Randomized point-based value iteration for POMDPs. *Journal of Artificial Intelligence Research*, 24(1):195–220, 2005.