Approximate Planning for Decentralized MDPs with Sparse Interactions

(Extended Abstract)

Francisco S. Melo*
GAIPS - INESC-ID
TagusPark, Edifício IST
2780-990 Porto Salvo, Portugal
fmelo@inesc-id.pt

Manuela Veloso
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213, USA
veloso@cs.cmu.edu

ABSTRACT

We explore how local interactions can simplify the process of decision-making in multiagent systems. We review decentralized sparse-interaction Markov decision process [3] that explicitly distinguishes the situations in which the agents in the team must coordinate from those in which they can act independently. We situate this class of problems within different multiagent models, such as MMDPs and transition independent Dec-MDPs [2]. We contribute new algorithm for efficient planning in this class of problems. We provide empirical comparisons between our algorithms and other existing algorithms for this class of problems.

Categories and Subject Descriptors

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

General Terms

Algorithms

Keywords

Planning under uncertainty, cooperative multiagent systems, sparse interaction.

1. LOCAL INTERACTIONS IN DEC-MDPS

We depart from the transition and reward-independent Decentralized Markov decision process (Dec-MDP) and introduce a model for multiagent decision problems that is, at the same time, more general and more specific. It is known [1] that different degrees of independence in a Dec-MDP translate in terms of reduced (worst-case) complexity. This discussion is summarized in the diagram in Fig. 1. The goal of this paper is to exploit sparse interactions among the different agents in a Dec-MDP: we are interested in Dec-MDPs in which there is some level of both transition and reward dependency, but this dependency is limited to specific regions of the state space. In the diagram of Fig. 1, our model corresponds to the blue circle.

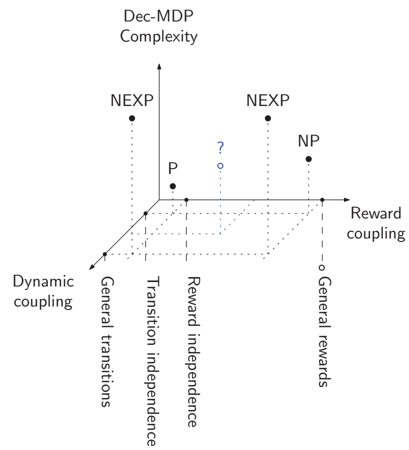

Figure 1: Currently known complexity results for different sub-classes of Dec-MDPs [1].

We represet a *decentralized sparse-interaction MDP* (Dec-SIMDP) as a tuple $\Gamma = \big(\{\mathcal{M}_k, k = 1, \ldots, N\}, \{\mathcal{M}_i^I, i = 1, \ldots, M\}\big)$, where

- Each $\mathcal{M}_k$ is an Markov decision process (MDP) $\mathcal{M}_k = (\mathcal{X}_k, \mathcal{A}_k, \mathsf{P}_k, r_k, \gamma)$ modeling agent k in the absence of other agents;
- Each $\mathcal{M}_i^I$ is an multiagent MDP (MMDP) that captures a *local interaction* between N_i agents in the states in $\mathcal{X}_i^I$ and is given by $\mathcal{M}_i^I = (K_i, \mathcal{X}_i^I, (\mathcal{A}_k), \mathsf{P}_i^I, r_i^I, \gamma)$.

Each MMDP describes the interaction between a subset K_i of agents, and the corresponding state-space $\mathcal{X}_i^I$ is an *interaction area* – a subset of the joint state-space for the agents in K_i. A Dec-SIMDP rests on the fundamental assumption that, in the interaction areas (and only in these), the agents involved in the corresponding MMDP are *able to share information* – namely, state information. In these areas communication overcomes local state perception and the agents can decide jointly on their action. Outside these areas, the agents have only a local perception of the state and must, therefore, choose the actions using only local information.

In the absence of any interaction areas, the Dec-SIMDP reduces to a set of independent MDPs that can be solved

*Part of this work was conducted while the author was a Post-Doctoral Fellow at Carnegie Mellon University.

Cite as: Approximate Planning for Decentralized MDPs with Sparse Interactions (Extended Abstract), Francisco S. Melo and Manuela Veloso, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1389-1390

Table 1: Results in different test scenarios [3].

Environment	Steps-to-goal (IDMG)	Steps-to-goal (LAPSI)
CIT	12.532	12.547
CMU	39.381	39.492
ISR	7.512	7.548
MIT	22.513	22.557
PENTAGON	5.352	5.385
SUNY	12.489	12.560

separately. On the other hand, when all agents interact in all states, the whole state-space is an interaction area and our assumption of full state observability renders this model equivalent to an MMDP. Nevertheless, the appeal of Dec-SIMDPs is that many practical situations do not fall in either of the two extreme cases. It is in these situations that the Dec-SIMDP model may bring an advantage over more general (but potentially intractable) models.

2. PLANNING IN DEC-SIMDPS

We introduce a new algorithm for Dec-SIMDPs that uses *look-ahead* during planning, taking into account possible "long-term effects" of interactions. It is henceforth named *look-ahead planning for sparse interactions* (LAPSI). Our method also provides an interesting insight into the problem of choosing the interaction areas for a given scenario, by exposing a close relation between Dec-SIMDPs and MMDPs.

Since Dec-SIMDPs are particular classes of Dec-MDPs, for any Dec-SIMDP $\mathcal{M}$ there is an *associated MMDP*, *i.e.*, the fully observable model associated with $\mathcal{M}$. This MMDP agglomerates the individual goals of all agents, creating a *joint goal* for the whole group. It also accommodates for all interactions simultaneously. Therefore, the optimal policy for this MMDP provides the optimal action choice in every joint-state and, in particular, in the states in the interaction areas. This policy also provides information about the long-term effects of these interactions in non-interaction states. The difficulty here lies in the fact that each agent does not know the state of the other agents. One possible approach is to use a "majority vote" strategy to choose the action in non-interaction states. Concretely, when at state x_k, agent k determines all joint states $x \in \mathcal{X}$ such that x_k is the kth component of x. For each such state, the agent determines all optimal actions $a \in \mathcal{A}$ using $\bar{Q}^*$, each corresponding to a "vote" in the corresponding individual action a_k. The agent computes the "voting" for each possible action $a_k \in \mathcal{A}_k$ and chooses the action with the largest number of votes.

3. RESULTS

We applied LAPSI to a collection of navigation scenarios from [3], an example of which can be found in Fig. 2(a). The reason for using of navigation scenarios is that the Dec-SIMDP model appears particularly appealing for modeling multi-robot problems. Furthermore, in this class of problems, the results can be easily visualized and interpreted.

For each of the test scenarios, we ran the LAPSI policy for $1,000$ independent trials of 250 time-steps, and compared the performance of our method to that of the IDMG algorithm [3]. Table 1 summarizes the results. Notice that, in all environments, the LAPSI algorithm performed similarly to the IDMG algorithm – which has been shown optimal

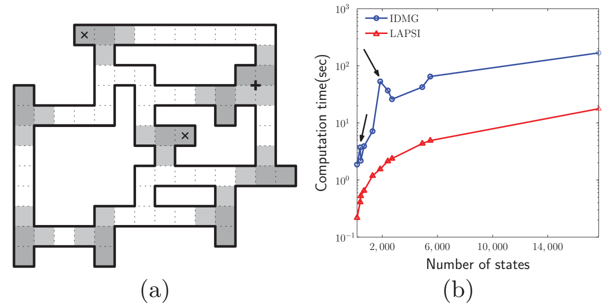

Figure 2: (a) Example of a test scenario. (b) Running times for LAPSI/IDMG.

for the used scenarios [3]. However, our method empirically exhibits running times much inferior to those of the IDMG algorithm – our method being about 10 times faster. To illustrate this point, we ran several instances of both algorithms in several problems of different sizes and the corresponding running times are reported in Fig. 2(b). The two peaks signaled in the figure correspond to two environments particularly "cluttered" with interaction areas.

4. CONCLUDING REMARKS

We conclude with several remarks on the properties of the LAPSI algorithm. First of all, our algorithm is able to "look ahead" during planning and use some information on the possibility of interaction. Preliminary results show that this overcomes one of the limitations of the IDMG algorithm. Also, the voting method is amenable to an interpretation as an expert-advice system. This interpretation opens an interesting door for future research in which regret minimization can be used to have the agent *learn* how much to trust each of the "voting states".

Finally, the associated MMDP used to compute the LAPSI policy is a "simplified version" of the Dec-SIMDP problem. However, by comparing the optimal policy in this MMDP and the optimal policies from the individual MDPs it should be possible to pinpoint those joint-states in which the joint action significantly differs from the one prescribed by the individual MDPs and in which the actions for each agent greatly depend on the state of the other agents. These differences provide one recipe for choosing the interaction states as those in which individual state-information is not sufficient to determine the best action.

Acknowledgements

This research was partially sponsored by the Portuguese Fundação para a Ciência e a Tecnologia under the Carnegie Mellon-Portugal Program and the Information and Communications Technologies Institute (ICTI), www.icti.cmu.edu. The views and conclusions contained in this document are those of the authors only.

5. REFERENCES

[1] M. Allen and S. Zilberstein. Complexity of Decentralized Control: Special Cases. In *Adv. Neural Inform. Proc. Systems* 22, pp. 19–27, 2009.

[2] R. Becker, S. Zilberstein, V. Lesser, and C. Goldman. Solving transition independent decentralized MDPs. *J. Artificial Intelligence Res.*, 22:423–455, 2004.

[3] M. Spaan and F. Melo. Interaction-driven Markov games for decentralized multiagent planning under uncertainty. In *Int. Conf. Autonomous Agents and Multiagent Systems*, pp. 525–532, 2008.