Policy gradient approach to multi-robot learning

Francisco Melo
Institute for Systems and Robotics
Instituto Superior Técnico
Lisboa, Portugal
fmelo@isr.ist.utl.pt

Extended Abstract

In theory, the formalism and methods of reinforcement learning (RL) can be applied to address any optimal control task, yielding optimal solutions while requiring very little a priori information on the system itself. However, in practice, RL methods suffer from the "curse of dimensionality" and exhibit limited applicability in complex control problems. Unfortunately, many actual control problems are inherently infinite, described in terms of continuous state variables. This is the case, for example, of optimal control of autonomous vehicles or complex robotic systems. However, the combination of value-based methods (such as Q-learning) and function approximation is far from trivial and the usefulness of the obtained solutions is still not clear. This has, perhaps, motivated the impressive advances in policy-gradient-based methods in recent years [3].

The motivation to extend these methods to multi-robot scenarios is evident. Many tasks found in practice are inherently too complex or even impossible for a single robot to execute. Furthermore, it is often the case that the use of several cheap robots is preferrable to the use of a single complex and expensive robot. On the other hand, the "traditional RL approach" makes use of game theoretic models such as Markov games. These approaches are generally unsuited to address problems envolving real robots, because they rely on several joint-observability assumptions inherent to these models that seldom hold in practice. Finally, more realistic models such as Dec-POMDPs are inherently too complex to be solved exactly.

It is in face of this inherent complexity in addressing complex multi-robot problems that policy gradient methods may prove of use. In this work, we conduct a preliminary study of policy-gradient methods in multi-robot problems. In particular, we analyze how successful policy-based approaches such as WoLF-PHC [1] can be adapted to accomodate the recent developments in policy gradient methods. The setting considered in this work is distinct from other approaches in the literature [2] in that we assume no joint-state or joint-action observability, which renders our approach more adequate to address multi-robot problems (where such assumptions seldom hold). We study how the existence of several independent learners in a common environment effects the overall learning performance of the different agents in several simple multi-robot scenarios and discuss how this approach can be extended to more complex problems.

Acknowledgements

This work was partially supported by Programa Operacional Sociedade do Conhecimento (POS_C) that includes FEDER funds. The author acknowledges the PhD grant SFRH/BD/3074/2000.

References

[1] M. Bowling and M. Veloso. Rational and convergent learning in stochastic games. In *Proc. 17th Int. Joint Conf. Artificial Intelligence*, pages 1021–1026, 2001.

[2] V. Könönen. Policy gradient method for team Markov games. In *Intelligent Data Engineering and Automated Learning*, *LNCS 3177*, pages 733–739, 2004.

[3] J. Peters, S. Vijayakumar, and S. Schaal. Policy gradient methods for robot control. Technical Report CS-03-787, University of Southern California, 2003.