Review: Accelerating Multi-agent Reinforcement Learning with Dynamic Co-learning

Steve Homer[1], Fabian Perez[1], Quinten Rosseel[1] and Matthias Humt[1]

[1]{steven.homer, fabian.perez, quinten.rosseel, matthias.humt}@vub.be

Abstract

The article being reviewed introduces a technique to accelerate the learning progress in Multi-Agent Reinforcement Learning (MARL) settings, by dynamically sharing the experience of contextually similar agents with their neighbors.
Todo: Finish abstract

Introduction

In the standard Reinforcement Learning (RL) setting, an agent is placed into an unknown environment and provided with a set of actions from which it can choose. By performing an action, the agent can change its state which is then solely defined by the previous state the agent was in before and the chosen action. In certain states, defined by the problem setting, the environment provides feedback to the agent, often called *reward* which can be either positive or negative. From this, the agent begins to approximate the underlying reward function in trying to maximize the expected future reward. Once having started to learn, the agent has to decide whether to exploit current knowledge by choosing the actions it expects to yield the highest reward or to explore new states with potentially higher rewards by trying new state-action combinations.
Todo: Add an example
Multi-Agent Reinforcement Learning (MARL) is an extension to the standard RL problem setting, where multiple, often hundreds or thousands of autonomous agents try to pursue their individual goals simultaneously in a common environment. We speak of cooperative MARL, if multiple or all agents pursue the same goal, which means they try to maximize a common reward function. If they succeed, we say that they follow a (near-optimal) joint policy where policy means a mapping from states to actions. To achieve this however has proven to be difficult for large-scale multi-agent systems as it is computationally expensive and requires a large number of update steps until it converges.
An important insight on the path to solving this problem is the observation, that an exploitable structure in the problem setting in the form of contextually similar groups of agents emerges during learning. Real world examples of this effect can be observed in Todo: Add examples. Agents working on similar tasks under comparable environmental dynamics might benefit from sharing information to accelerate their individual learning progress which is the paradigm proposed in the paper under review. Additionally, strategies to identify promising candidates for information sharing and group formation are proposed.
Todo: Add citations

Methods

Todo: Explain model as proposed in the paper
Todo: Explain our implementation of the model

Results

Todo: Describe the performed simulations and their results
Todo: Include chosen parameter settings

Discussion

Todo: Summary and explanation of our work

Conclusion

Todo: Add related work and an outlook
Todo: Make sure all questions given below are answered

1. Does the introduction explain clearly the content of the paper
2. whether there is sufficient background information to understand the relevance of the work
3. whether the methods are clearly explained (can the results be reproduced?)
4. whether the results answer the questions asked in the paper.
5. whether all questions are answered
6. whether the conclusion is sufficient
7. and whether the overall style is ok and
8. whether you believe things are missing in the discussion.

9. etc.

10. 3 positive points concerning the work, clearly specifying why you think they are well- done or interesting

11. 3 negative points, which may include missing/unclear explanations or suggestions for improvement

12. at least 3 clear and relevant questions on the content or the methods used which can be asked (next to other questions).

(Carroll and Seppi, 2005), (Garant et al., 2015),
(Ghavamzadeh et al., 2006), (Gmytrasiewicz and Doshi, 2005),
(Guestrin et al., 2002), (Kitano et al., 1999),
(Lazaric et al., 2008), (Littman, 2001),
(Nair et al., 2005), (Nedic and Ozdaglar, 2009),
(Oliehoek et al., 2008), (Price and Boutilier, 2003),
(Rényi et al., 1961), (Taylor and Stone, 2009),
(Vickrey and Koller, 2002), (Witwicki and Durfee, 2010),
(Zhang et al., 2010), (Zhang and Lesser, 2013)

References

Carroll, J. L. and Seppi, K. (2005). Task similarity measures for transfer in reinforcement learning task libraries. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 2, pages 803–808. IEEE.

Garant, D., da Silva, B. C., Lesser, V., and Zhang, C. (2015). Accelerating multi-agent reinforcement learning with dynamic co-learning. Technical report, Technical report.

Ghavamzadeh, M., Mahadevan, S., and Makar, R. (2006). Hierarchical multi-agent reinforcement learning. *Autonomous Agents and Multi-Agent Systems*, 13(2):197–229.

Gmytrasiewicz, P. J. and Doshi, P. (2005). A framework for sequential planning in multi-agent settings. *J. Artif. Intell. Res.(JAIR)*, 24:49–79.

Guestrin, C., Koller, D., and Parr, R. (2002). Multiagent planning with factored mdps. In *Advances in neural information processing systems*, pages 1523–1530.

Kitano, H., Tadokoro, S., Noda, I., Matsubara, H., Takahashi, T., Shinjou, A., and Shimada, S. (1999). Robocup rescue: Search and rescue in large-scale disasters as a domain for autonomous agents research. In *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on*, volume 6, pages 739–743. IEEE.

Lazaric, A., Restelli, M., and Bonarini, A. (2008). Transfer of samples in batch reinforcement learning. In *Proceedings of the 25th international conference on Machine learning*, pages 544–551. ACM.

Littman, M. L. (2001). Value-function reinforcement learning in markov games. *Cognitive Systems Research*, 2(1):55–66.

Nair, R., Varakantham, P., Tambe, M., and Yokoo, M. (2005). Networked distributed pomdps: A synthesis of distributed constraint optimization and pomdps. In *AAAI*, volume 5, pages 133–139.

Nedic, A. and Ozdaglar, A. (2009). Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61.

Oliehoek, F. A., Spaan, M. T., Whiteson, S., and Vlassis, N. (2008). Exploiting locality of interaction in factored dec-pomdps. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, pages 517–524. International Foundation for Autonomous Agents and Multiagent Systems.

Price, B. and Boutilier, C. (2003). Accelerating reinforcement learning through implicit imitation. *Journal of Artificial Intelligence Research*, 19:569–629.

Rényi, A. et al. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.

Taylor, M. E. and Stone, P. (2009). Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10(Jul):1633–1685.

Vickrey, D. and Koller, D. (2002). Multi-agent algorithms for solving graphical games. In *AAAI/IAAI*, pages 345–351.

Witwicki, S. J. and Durfee, E. H. (2010). Influence-based policy abstraction for weakly-coupled dec-pomdps. In *ICAPS*, pages 185–192.

Zhang, C. and Lesser, V. (2013). Coordinating multi-agent reinforcement learning with limited communication. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 1101–1108. International Foundation for Autonomous Agents and Multiagent Systems.

Zhang, C., Lesser, V., and Abdallah, S. (2010). Self-organization for coordinating decentralized reinforcement learning. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 739–746. International Foundation for Autonomous Agents and Multiagent Systems.