Literature Review: Real-Time Multi-Agent Learning for Robotic Weed Killing in Partially Observed Environments

Wyatt McAllister1

[1]University of Illinois at Urbana-Champaign

November 14, 2017

Objective

- Our research is on real-time multi-agent robotic weed killing in partially observed environments.
- We aim to ground our research in current work in the field.
- Towards this goal we will survey papers in multi-agent learning and note relevant results.

Learning for Multi-robot Cooperation in Partially Observable Stochastic Environments with Macro-actions [Liu et al., 2017]

- Author: Miao Liu1, Kavinayan Sivakumar, Shayegan Omidshafiei, Christopher Amato and Jonathan P. How
- Date: 2017
- Relevant Information: Details a framework for multi-agent reinforcement learning for DecPOMDPs utilizing macro-actions. Demonstrates the effectiveness of this framework for a search and rescue task performed by aerial and ground robots. Assumed full observation of the victim location via aerial vehicle but not the health status of victims.

Observation 1

- We note that even current research in the field [Liu et al., 2017], assumes knowledge of the positions of objects in the environment obtained from a UAV.
- We aim to complete our task with only the observations obtained from ground robots during operation in the field.

The Dynamics of Reinforcement Learning in Cooperative Multi-agent Systems [Claus and Boutilier, 1998]

- Author: Caroline Claus and Craig Boutilier
- Date: 1998
- Relevant Information: States that, in the multi-agent domain, greedy policies from all the agents may not maximize the overall reward. This motivates further work on coordination.

Markov games as a framework for multi-agent reinforcement learning [Littman, 1994]

- Author: Michael L. Littman
- Date: 1994
- Relevant Information: States that multi-agent RL converges to the optimal policy for zero sum games. Shows that the problem may be formulated in a minimax fashion.

Multi-Agent Reinforcement Learning (MARL): a critical survey [Shoham et al., 2003]

- Author: Yoav Shoham Rob Powers Trond Grenager
- Date: 2003
- Relevant Information: States that multi-agent RL also converges for common-payoff games. Presents a survey of key directions in MARL: how humans learn in context of other learners, vs. how agents in general should learn.

Observation 2

- We see that in contemporary literature [Shoham et al., 2003], our problem is framed as a robot foraging task.

Multi-Agent Reinforcement Learning (MARL): Independent Vs. Cooperative Agents [Tan, 1993]

- Author: Ming Tan
- Date: 1993
- Relevant Information: Presents an outline of strategies for coordination between agents in a wolf pack environment: coordinated sensing, experience sharing, expert training from more experienced agents.

A Comprehensive Survey of Multi-agent Reinforcement Learning (MARL) [Busoniu et al., 2008]

- Author: Lucian Busoniu, Robert Babuska, and Bart De Schutter
- Date: 2008
- Relevant Information: Presents a comprehensive overview of the taxonomy of MARL Algorithms with examples of algorithms from each category: Temporal Difference RL, Game Theory, Direct Policy Search, cooperative vs. competitive, static vs. dynamic.

A junction-tree based learning algorithm to optimize network wide traffic control: A coordinated multi-agent framework [Zhu et al., 2015]

- Author: Feng Zhu, H.M. Abdul Aziz, Xinwu Qian, Satish V. Ukkusuri
- Date: 2015
- Relevant Information: Details an algorithm for traffic control using multi-agent RL. Utilizes a graph network model for the traffic environment, and shows convergence for both cyclic and acyclic networks.

Parallel Reinforcement Learning for Traffic Signal Control [Mannion et al., 2015]

- Author: Patrick Manniona, Jim Duggana, Enda Howley
- Date: 2015
- Relevant Information: Details a solutions to a similar problem, but uses parallel hierarchical reinforcement learning. In this work, multiple agents learn in parallel, and share information with a master learner, increasing the rate of exploration, and the learning rate.

Adaptive group-based signal control by reinforcement learning [Jin and Ma, 2015]

- Author: Junchen Jin, Xiaoliang Ma
- Date: 2015
- Relevant Information: Presents another examination of signal traffic control. This work uses a group based RL approach, in which groups of agents throughout the traffic network learn coordinated policies. It is shown here that SARSA is more adaptive to dynamic conditions than Q learning for the experiments considered.

Decentralized Non-communicating Multi-agent Collision Avoidance with Deep Reinforcement Learning [Chen et al., 2017]

- Author: Yu Fan Chen, Miao Liu, Michael Everett, and Jonathan P. How
- Date: 2017
- Relevant Information: Details a real-time collision avoidance strategy for the multi-robot domain based on reinforcement learning. Here, a value network is trained offline based on generated sample trajectories, and then a real-time learning system, which can adapt to novel trajectories, is utilized. Does not assume communication between agents but does assume observation of the current position and velocity.

Multi-agent Reinforcement Learning in Sequential Social Dilemmas [Leibo et al., 2017]

- Author: Multi-agent Reinforcement Learning in Sequential Social Dilemmas
- Date: 2017
- Relevant Information: Details an analysis of modern multi-agent reinforcement learning algorithms in terms of the social aspects of the policy. Characterizes the policy of wolf pack environments in terms of cooperation and defection. It is shown that for this environment effective cooperative policies take longer to learn.

Multiagent cooperation and competition with deep reinforcement learning [Tampuu et al., 2017]

- Author: Ardi Tampuu, Tambet Matiisen, Dorian Kodelja, Ilya Kuzovkin, Kristjan Korjus, Juhan Aru, Jaan Aru, Raul Vicente
- Date: 2017
- Relevant Information: Details a deep Q learning approach to training Pong where both agents learn simultaneously. Shows that this simultaneous learning architecture improves the robustness of the trained policy to opponents with varying policy structures.

Reinforcement Learning in the Multi-Robot Domain [Matarić, 1997]

- Author: Maja J. Mataric
- Date: 1997
- Relevant Information: Details a practical implementation of Deep Q-Learning for multi-robot learning on coordinated tasks. Solves a gathering problem utilizing macro-actions and a reward function which forces the robots to spread out over the environment.

Observation 3

- We note that past work for [Matarić, 1997], multi-agent robot foraging systems have used additive rewards over the rewards of the individual agents.
- This is a good staring point for our reward model.

An algorithm for distributed reinforcement learning in cooperative multi-agent systems [Lauer and Riedmiller, 2000]

- Author: Maja J. Mataric
- Date: 2000
- Relevant Information: Outlines a distributed approach to Q learning in which all learning agents operate on the same problem. Each learner assumes the other agents take optimal actions at every state. Does not include coordinate and does not guarantee convergence when there are multiple optimal actions at each state.

Cooperative multi-agent learning: The state of the art [Panait and Luke, 2005]

- Author: Panait, Liviu and Luke, Sean
- Date: 2005
- Relevant Information: Surveys key directions in MAL including, team learning for cooperative and non-cooperative agents, learning for cooperative vs. general sum games, modeling of unknown agents, direct and indirect communications, and challenges such as scalability.

Packet routing in dynamically changing networks: A reinforcement learning approach [Boyan and Littman, 1994]

- Author: Boyan, Justin A and Littman, Michael L
- Date: 1994
- Relevant Information: This problem examines packet routing in dynamic networks, and uses a distributed Q learning algorithm to solve the problem.

Multi-agent Bidirectionally-Coordinated Nets for Learning to Play StarCraft Combat Games [Peng et al., 2017]

- Author: Peng, Peng and Yuan, Quan and Wen, Ying and Yang, Yaodong and Tang, Zhenkun and Long, Haitao and Wang, Jun
- Date: 2017
- Relevant Information: Utilizes an actor critic deep Q-Learning approach for StarCraft which treats the combat game as a zero sum problem between the teams, and uses a bi-directional neural network to allow coordinated learning. Trains one with team given a model for the other. Shows high performance for StarCraft.

Observation 4

- We note that current work in multi agent learning [Peng et al., 2017], has used actor-critic networks to achieve high performance.
- This will be an interesting approach to examine for our problem.

A distributed reinforcement learning scheme for network routing [Littman and Boyan, 1993]

- Author: Littman, Michael and Boyan, Justin
- Date: 1993
- Relevant Information: A previous version of the above algorithm for packet routing in dynamic networks.

Cooperative Multi-Agent Control Using Deep Reinforcement Learning [Gupta et al.,]

- Author: Gupta, Jayesh K and Egorov, Maxim and Kochenderfer, Mykel
- Date: 2017
- Relevant Information: Outlines a framework for cooperative control using deep reinforcement learning via deep deterministic policy gradient. Uses parameter sharing between agents and show effectiveness for a multi-walker, water world, and pursuit evasion.

Data-driven optimal consensus control for discrete-time multi-agent systems with unknown dynamics using reinforcement learning method [Zhang et al., 2017]

- Author: Zhang, Huaguang and Jiang, He and Luo, Yanhong and Xiao, Geyang
- Date: 2017
- Relevant Information: Proposes a deep Q-Learning framework using actor-critic for multi-agent coordination and shows its effectiveness in simulation for multi-agent coordination.

Multiagent reinforcement learning and self-organization in a network of agents [Abdallah and Lesser, 2007]

- Author: Abdallah, Sherief and Lesser, Victor
- Date: 2007
- Relevant Information: Demonstrates the effectiveness of multi-agent reinforcement learning for learning coordinated policies in a network of agents. Uses self organization to dynamically restructure the network and allows for effective distributed task allocation.

Planning for large-scale multiagent problems via hierarchical decomposition with applications to UAV health management [Chen et al., 2014]

- Author: Chen, Yu Fan and Ure, N Kemal and Chowdhary, Girish and How, Jonathan P and Vian, John
- Date: 2014
- Relevant Information: Proposes an Multiagent Markov Decision Processes (MMDP) algorithm which dynamically allocates tasks to a network of agents in order to maximize longterm reward. Tests the algorithm in a simulation for multi-agent UAV health monitoring.

Multi-Agent Reinforcement Learning with Reward Shaping for KeepAway Takers [Devlin et al., 2010]

- Author: Devlin, Sam and Grzes, Marek and Kudenko, Daniel
- Date: 2010
- Relevant Information: Uses reward shaping for multi agent reinforcement learning and benchmarks the performance for a simulation of robot soccer.

An Evolutionary Transfer Reinforcement Learning Framework for Multi-Agent System [Hou et al., 2017]

- Author: Hou, Yaqing and Ong, Yew-Soon and Feng, Liang and Zurada, Jacek M
- Date: 2017
- Relevant Information: Develops a novel framework for evolutionary transfer reinforcement learning and shows improved performance for learning on strategy games.

Planning and acting in partially observable stochastic domains [Kaelbling et al., 1998]s

- Author: Kaelbling, Leslie Pack and Littman, Michael L and Cassandra, Anthony R
- Date: 1998
- Relevant Information: Takes an approach to solving POMDPs which optimizes over belief spaces to prune policy trees in order to maximize the expected reward. Shows improved efficiency over previous approaches and includes experiments with a hall following robot.

Learning in behavior-based multi-robot systems: Policies, models, and other agents [Mataric, 2001]

- Author: Matarić, Maja J
- Date: 2001
- Relevant Information: Outlines a survey Behavior Based Control in the context of multi-robot learning and planning. Shows how this methodology allows effective learning of cooperative actions for a robot foraging problem. Discuss methods including reward shaping in the spatial and temporal domain, and imitation learning to improve performance.

Simultaneous adversarial multi-robot learning [Bowling and Veloso, 2003]

- Author: Bowling, Michael and Veloso, Manuela
- Date: 2003
- Relevant Information: Outlines a variant a the Wolf algorithm called GradWolf, which uses gradient based policy iteration with variable learning rate to perform multi-agent learning for adversarial tasks. Includes simulated and real experiments on an adversarial multi-robot soccer task.

Planning, learning and coordination in multiagent decision processes [Boutilier, 1996]

- Author: Boutilier, Craig
- Date: 1996
- Relevant Information: Discusses strategies for coordinated multi-agent planning. Assumes fully autonomous and decentralized agents, showing how coordination strategies can be learned by planning over beliefs.

Multiagent learning using a variable learning rate [Bowling and Veloso, 2002]

- Author: Michael Bowling, Manuela Veloso
- Date: 2002
- Relevant Information: Presents an extension to the Win or Learn Fast (WoLF) algorithm for variable learning rate in adversarial multi agent learning problems. This extension is based on Policy Hill Climbing (PHC) and is called and extends the WoLF algorithm to provide theoretical convergence guarantees.

Adaptive Load Balancing A Study Learning [Schaerf et al., 1995]

- Author: Schaerf, Andrea and Shoham, Yoav and Tennenholtz, Moshe
- Date: 1995
- Relevant Information: Creates an adaptive learning controller for load balancing in task allocation systems involving multiple agents. Uses a reinforcement learning approach assuming fully decentralized agents with no communication. Learns relevant parameters from local information to construct the autonomous controller.

Cooperative mobile robotics: Antecedents and directions [Cao et al., 1997]

- Author: Cao, Y Uny and Fukunaga, Alex S and Kahng, Andrew
- Date: 1997
- Relevant Information: Presents a survey of coordinated robotics including a taxonomy of key research directions, as well as cited works from each. Details popular problems such as foraging, traffic control, cooperative manipulation, and pursuit evasion. Talks about several methodologies in the centralized, decentralized domains, using varying communication architectures, and taking approaches inspired by biology, learning methods, and behavioral control.

Observation 5

- We note that the foraging task has long been considered a key problem in collaborative robotics [Cao et al., 1997].

If multi-agent learning is the answer, what is the question? [Shoham et al., 2007]

- Author: Shoham, Yoav and Powers, Rob and Grenager, Trond
- Date: 2007
- Relevant Information: Breaks up the field of multi-agent reinforcement learning into popular branches such as those concerned with computational efficiency and tractability, those which describe how agents behave in shared environments and how different learning methods perform for varying tasks, and those which desire to proscribe strategies for learning in competitive and non competitive environments. Provides relevant literature for each subfield and identifies key challenges and drawbacks within each.

Learning for Decentralized Control of Multiagent Systems in Large, Partially-Observable Stochastic Environments [Liu et al., 2016]

- Author: Liu, Miao and Amato, Christopher and Anesta, Emily P and Griffith, John Daniel and How, Jonathan P
- Date: 2016
- Relevant Information: Details the PoEM algorithm, which solves MacPOMDPs and has guaranteed convergence properties. Shows the effectiveness of the problem for a search and rescue task.

Decentralized control of partially observable Markov decision processes using belief space macro-actions [Omidshafiei et al., 2015]

- Author: Omidshafiei, Shayegan and Agha-Mohammadi, Ali-Akbar and Amato, Christopher and How, Jonathan P
- Date: 2015
- Relevant Information: Proposes an algorithm, Masked Monte Carlo Search, for solving Partially Observable Semi Markov Decision processes. This algorithm combines random sampling with masking of unfavorable solutions to balance exploration and exploitation. It is shown to perform well on a package delivery application.

Planning for decentralized control of multiple robots under uncertainty [Amato et al., 2015]

- Author: Amato, Christopher and Konidaris, George and Cruz, Gabriel and Maynor, Christopher A and How, Jonathan P and Kaelbling, Leslie P
- Date: 2015
- Relevant Information: Discusses MacDec-POMDPs and option-based dynamic programming as a generalized learning framework for the multi robot domain. Demonstrates the feasibility of this methods for a multi robot warehouse task in situations with no communication, communication within a local radius, and signal based communication. Emergent cooperative strategies were observed to facilitate multi-robot collaboration, motivating further work.

Planning with macro-actions in decentralized POMDPs [Amato et al., 2014]

- Author: Amato, Christopher and Konidaris, George D and Kaelbling, Leslie P
- Date: 2014
- Relevant Information: Describes a set of Mac-Dec-POMDP algorithms that use policy search methods to plan in the multi-robot domain in uncertain environments. Outlines the methodology behind this approach and details experiments for the tasks of meeting in a grid and navigating with moving obstacles.

Graph-based Cross Entropy method for solving multi-robot decentralized POMDPs [Omidshafiei et al., 2016]

- Author: Omidshafiei, Shayegan and Agha-Mohammadi, Ali-Akbar and Amato, Christopher and Liu, Shih-Yuan and How, Jonathan P and Vian, John
- Date: 2016
- Relevant Information: Proposes a probabilistic method for solving Dec-POMSDPs which iteratively samples the policy space, and targets sampled policies which maximize the estimate of the value function. Shows improved performance over previous Dec-POMSDPs for a package retrieval and delivery domain in the multi-robot setting.

Robot motor skill coordination with EM-based reinforcement learning [Kormushev et al., 2010]

- Author: Kormushev, Petar and Calinon, Sylvain and Caldwell, Darwin G
- Date: 2010
- Relevant Information: Does learning-based control via an RL approach with dynamic motion primitives. Incorporates imitation learning and demonstrates performance on a dynamic pancake flipping task.

Stick-Breaking Policy Learning in Dec-POMDPs [Liu et al., 2015]

- Author: Liu, Miao and Amato, Christopher and Liao, Xuejun and Carin, Lawrence and How, Jonathan P
- Date: 2015
- Relevant Information: Introduces a novel exploration method for bayesian reinforcement learning realized by defining an augmented MDP framework which includes the probability of exploring at each state, and optimizing over this augmented framework.

Decentralized control of partially observable Markov decision processes [Amato et al., 2013]

- Author: Amato, Christopher and Chowdhary, Girish and Geramifard, Alborz and Ure, N Kemal and Kochenderfer, Mykel J
- Date: 2015
- Relevant Information: Notes a key and relevant result, which is that in factored DecMDPs which are observation, transition, and reward independent may be optimized by optimizing each agents local reward.

Observation 6

- From previous work, [Amato et al., 2013], we note that a factored MDP approach to the foraging problem is guaranteed to be optimal in transition, observation, and reward indepenedent factorizations with additive rewards [Amato et al., 2013].

Observation 7

- Based on our review, we believe that a factored MDP actor-critic approach, where the critic network is decentralized, and the actor network is centralized, will be effective for this problem.
- We will now examine literature on actor critic methods and applications.

Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem [Vamvoudakis and Lewis, 2010]

- Author: Vamvoudakis, Kyriakos G and Lewis, Frank L
- Date: 2010
- Relevant Information: This paper uses a policy iteration scheme for optimal control of continuous time affine nonlinear systems. They demonstrate improvements over previous methods for known system dynamics.

Bayesian policy gradient and actor-critic algorithms [Ghavamzadeh et al., 2016]

- Author: Ghavamzadeh, Mohammad and Engel, Yaakov and Valko, Michal
- Date: 2016
- Relevant Information: This paper presents a survey of Bayesian Actor Critic methods, showing experiments for a random walk problem, alongside mountain car and ship-steering problems.

Bayesian actor-critic algorithms [Ghavamzadeh and Engel, 2007]

- Author: Ghavamzadeh, Mohammad and Engel, Yaakov
- Date: 2007
- Relevant Information: This paper presents a preliminary version of the previous work with the fundamental theory and the random walk experiments.

Actor-critic algorithms [Konda and Tsitsiklis, 2000]

- Author: Konda, Vijay R and Tsitsiklis, John N
- Date: 2000
- Relevant Information: This paper presents a survey of actor-critic algorithms, alongside convergence results.

Learning continuous control policies by stochastic value gradients [Heess et al., 2015]

- Author: Heess, Nicolas and Wayne, Gregory and Silver, David and Lillicrap, Tim and Erez, Tom and Tassa, Yuval
- Date: 2000
- Relevant Information: This paper presents a method learning-based control framework using policy gradient algorithms which learns model and control parameters. Experiments in robot environments with various robotic platforms are presented in simulation.

Reinforcement learning for humanoid robotics [Peters et al., 2003]

- Author: Peters, Jan and Vijayakumar, Sethu and Schaal, Stefan
- Date: 2003
- Relevant Information: This paper derives a natural actor-critic algorithm for learning-based control an tests it on a humanoid robot.

Policy gradient methods for robotics [Peters and Schaal, 2006]

- Author: Peters, Jan and Schaal, Stefan
- Date: 2006
- Relevant Information: This paper presents a survey of policy gradients for humanoid robots. Experiments towards motor primitive learning for baseball are also included.

Interactive Policy Learning through Confidence-Based Autonomy [Chernova and Veloso, 2009]

- Author: Chernova, Sonia and Veloso, Manuela
- Date: 2009
- Relevant Information: This paper presents a policy learning framework based on a confidence estimate, providing the algorithm with a means to utilize demonstration data when confidence is low.

J. 4 supervised actor-critic reinforcement learning [Barto, 2004]

- Author: Barto, MTRAG
- Date: 2004
- Relevant Information: This paper presents a survey of supervised actor-critic reinforcement learning, and experiments for a ship steering task, alongside control of a robot arm.

Deterministic policy gradient algorithms [Silver et al., 2014]

- Author: Silver, David and Lever, Guy and Heess, Nicolas and Degris, Thomas and Wierstra, Daan and Riedmiller, Martin
- Date: 2014
- Relevant Information: This paper presents a survey of deterministic policy gradient algorithms. Experiments in bandit problems, mountain car and pendulum environments, and control of an octopus arm are presented.

References I

Abdallah, S. and Lesser, V. (2007).
Multiagent reinforcement learning and self-organization in a network of agents.
In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, page 39. ACM.

Amato, C., Chowdhary, G., Geramifard, A., Ure, N. K., and Kochenderfer, M. J. (2013).
Decentralized control of partially observable markov decision processes.
In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 2398–2405. IEEE.

Amato, C., Konidaris, G., Cruz, G., Maynor, C. A., How, J. P., and Kaelbling, L. P. (2015).
Planning for decentralized control of multiple robots under uncertainty.
In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 1241–1248. IEEE.

Amato, C., Konidaris, G. D., and Kaelbling, L. P. (2014).
Planning with macro-actions in decentralized pomdps.
In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1273–1280. International Foundation for Autonomous Agents and Multiagent Systems.

Barto, M. (2004).
J. 4 supervised actor-critic reinforcement learning.
Handbook of learning and approximate dynamic programming, 2:359.

Boutilier, C. (1996).
Planning, learning and coordination in multiagent decision processes.
In *Proceedings of the 6th conference on Theoretical aspects of rationality and knowledge*, pages 195–210. Morgan Kaufmann Publishers Inc.

References II

Bowling, M. and Veloso, M. (2002).
Multiagent learning using a variable learning rate.
Artificial Intelligence, 136(2):215–250.

Bowling, M. and Veloso, M. (2003).
Simultaneous adversarial multi-robot learning.
In *IJCAI*, volume 3, pages 699–704.

Boyan, J. A. and Littman, M. L. (1994).
Packet routing in dynamically changing networks: A reinforcement learning approach.
In *Advances in neural information processing systems*, pages 671–678.

Busoniu, L., Babuska, R., and De Schutter, B. (2008).
A comprehensive survey of multiagent reinforcement learning.
IEEE Transactions on Systems, Man, And Cybernetics-Part C: Applications and Reviews, 38 (2), 2008.

Cao, Y. U., Fukunaga, A. S., and Kahng, A. (1997).
Cooperative mobile robotics: Antecedents and directions.
Autonomous robots, 4(1):7–27.

Chen, Y. F., Liu, M., Everett, M., and How, J. P. (2017).
Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning.
In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, volume 1, pages 285–292.

References III

Chen, Y. F., Ure, N. K., Chowdhary, G., How, J. P., and Vian, J. (2014).
Planning for large-scale multiagent problems via hierarchical decomposition with applications to uav health management.
In *American Control Conference (ACC), 2014*, pages 1279–1285. IEEE.

Chernova, S. and Veloso, M. (2009).
Interactive policy learning through confidence-based autonomy.
Journal of Artificial Intelligence Research, 34(1):1.

Claus, C. and Boutilier, C. (1998).
The dynamics of reinforcement learning in cooperative multiagent systems.

Devlin, S., Grzes, M., and Kudenko, D. (2010).
Multi-agent reinforcement learning with reward shaping for keepaway takers.
In *AAMAS'10 Workshop on Adaptive and Learning Agents (ALA'10)*.

Ghavamzadeh, M. and Engel, Y. (2007).
Bayesian actor-critic algorithms.
In *Proceedings of the 24th international conference on Machine learning*, pages 297–304. ACM.

Ghavamzadeh, M., Engel, Y., and Valko, M. (2016).
Bayesian policy gradient and actor-critic algorithms.
Journal of Machine Learning Research, 17(66):1–53.

Gupta, J. K., Egorov, M., and Kochenderfer, M.
Cooperative multi-agent control using deep reinforcement learning.

References IV


Heess, N., Wayne, G., Silver, D., Lillicrap, T., Erez, T., and Tassa, Y. (2015).
Learning continuous control policies by stochastic value gradients.
In *Advances in Neural Information Processing Systems*, pages 2944–2952.


Hou, Y., Ong, Y.-S., Feng, L., and Zurada, J. M. (2017).
An evolutionary transfer reinforcement learning framework for multi-agent system.
IEEE Transactions on Evolutionary Computation.


Jin, J. and Ma, X. (2015).
Adaptive group-based signal control by reinforcement learning.
Transportation Research Procedia, 10:207–216.


Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998).
Planning and acting in partially observable stochastic domains.
Artificial intelligence, 101(1):99–134.


Konda, V. R. and Tsitsiklis, J. N. (2000).
Actor-critic algorithms.
In *Advances in neural information processing systems*, pages 1008–1014.


Kormushev, P., Calinon, S., and Caldwell, D. G. (2010).
Robot motor skill coordination with em-based reinforcement learning.
In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 3232–3237. IEEE.


Lauer, M. and Riedmiller, M. (2000).
An algorithm for distributed reinforcement learning in cooperative multi-agent systems.
In *In Proceedings of the Seventeenth International Conference on Machine Learning*. Citeseer.

References V

Leibo, J. Z., Zambaldi, V., Lanctot, M., Marecki, J., and Graepel, T. (2017).
Multi-agent reinforcement learning in sequential social dilemmas.
In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 464–473. International Foundation for Autonomous Agents and Multiagent Systems.

Littman, M. and Boyan, J. (1993).
A distributed reinforcement learning scheme for network routing.
In *Proceedings of the international workshop on applications of neural networks to telecommunications*, pages 45–51. Psychology Press.

Littman, M. L. (1994).
Markov games as a framework for multi-agent reinforcement learning.
In *Proceedings of the eleventh international conference on machine learning*, volume 157, pages 157–163.

Liu, M., Amato, C., Anesta, E. P., Griffith, J. D., and How, J. P. (2016).
Learning for decentralized control of multiagent systems in large, partially-observable stochastic environments.
In *AAAI*, pages 2523–2529.

Liu, M., Amato, C., Liao, X., Carin, L., and How, J. P. (2015).
Stick-breaking policy learning in dec-pomdps.
In *IJCAI*, pages 2011–2018.

Liu, M., Sivakumar, K., Omidshafiei, S., Amato, C., and How, J. P. (2017).
Learning for multi-robot cooperation in partially observable stochastic environments with macro-actions.
CoRR, abs/1707.07399.

References VI


Mannion, P., Duggan, J., and Howley, E. (2015).
Parallel reinforcement learning for traffic signal control.
Procedia Computer Science, 52:956–961.


Matarić, M. J. (1997).
Reinforcement learning in the multi-robot domain.
Autonomous Robots, 4:73–83.


Mataric, M. J. (2001).
Learning in behavior-based multi-robot systems: Policies, models, and other agents.
Cognitive Systems Research, 2(1):81–93.


Omidshafiei, S., Agha-Mohammadi, A.-A., Amato, C., and How, J. P. (2015).
Decentralized control of partially observable markov decision processes using belief space macro-actions.
In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 5962–5969. IEEE.


Omidshafiei, S., Agha-Mohammadi, A.-A., Amato, C., Liu, S.-Y., How, J. P., and Vian, J. (2016).
Graph-based cross entropy method for solving multi-robot decentralized pomdps.
In *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pages 5395–5402. IEEE.


Panait, L. and Luke, S. (2005).
Cooperative multi-agent learning: The state of the art.
Autonomous agents and multi-agent systems, 11(3):387–434.

References VII


Peng, P., Yuan, Q., Wen, Y., Yang, Y., Tang, Z., Long, H., and Wang, J. (2017).
Multiagent bidirectionally-coordinated nets for learning to play starcraft combat games.
arXiv preprint arXiv:1703.10069.


Peters, J. and Schaal, S. (2006).
Policy gradient methods for robotics.
In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 2219–2225. IEEE.


Peters, J., Vijayakumar, S., and Schaal, S. (2003).
Reinforcement learning for humanoid robotics.
In *Proceedings of the third IEEE-RAS international conference on humanoid robots*, pages 1–20.


Schaerf, A., Shoham, Y., and Tennenholtz, M. (1995).
Adaptive load balancing: A study in multi-agent learning.
Journal of artificial intelligence research, 2:475–500.


Shoham, Y., Powers, R., and Grenager, T. (2003).
Multi-agent reinforcement learning: a critical survey.
Web manuscript.


Shoham, Y., Powers, R., and Grenager, T. (2007).
If multi-agent learning is the answer, what is the question?
Artificial Intelligence, 171(7):365–377.


Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., and Riedmiller, M. (2014).
Deterministic policy gradient algorithms.
In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 387–395.

References VIII

Tampuu, A., Matiisen, T., Kodelja, D., Kuzovkin, I., Korjus, K., Aru, J., Aru, J., and Vicente, R. (2017).
Multiagent cooperation and competition with deep reinforcement learning.
PloS one, 12(4):e0172395.

Tan, M. (1993).
Multi-agent reinforcement learning: Independent vs. cooperative agents.
In *Proceedings of the tenth international conference on machine learning*, pages 330–337.

Vamvoudakis, K. G. and Lewis, F. L. (2010).
Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem.
Automatica, 46(5):878–888.

Zhang, H., Jiang, H., Luo, Y., and Xiao, G. (2017).
Data-driven optimal consensus control for discrete-time multi-agent systems with unknown dynamics using reinforcement learning method.
IEEE Transactions on Industrial Electronics, 64(5):4091–4100.

Zhu, F., Aziz, H. A., Qian, X., and Ukkusuri, S. V. (2015).
A junction-tree based learning algorithm to optimize network wide traffic control: A coordinated multi-agent framework.
Transportation Research Part C: Emerging Technologies, 58:487–501.