A Very Condensed Survey and Critique of Multiagent Deep Reinforcement Learning

JAAMAS Track

Pablo Hernandez-Leal
Borealis AI
Edmonton, Canada

Bilal Kartal
Borealis AI
Edmonton, Canada

Matthew E. Taylor
Borealis AI
Edmonton, Canada

ABSTRACT

Deep reinforcement learning (RL) has achieved outstanding results in recent years. This has led to a dramatic increase in the number of applications and methods. Recent works have explored learning beyond single-agent scenarios and have considered multiagent learning (MAL) scenarios. Initial results report successes in complex multiagent domains, although there are several challenges to be addressed. The primary goal of this extended abstract is to provide a broad overview of current multiagent deep reinforcement learning (MDRL) literature, hopefully motivating the reader to review our 47-page JAAMAS survey article [28]. Additionally, we complement the overview with a broader analysis: (i) We revisit previous key components, originally presented in MAL and RL, and highlight how they have been adapted to multiagent deep reinforcement learning settings. (ii) We provide general guidelines to new practitioners in the area: describing lessons learned from MDRL works, pointing to recent benchmarks, and outlining open avenues of research. (iii) We take a more critical tone raising practical challenges of MDRL.

KEYWORDS

Multiagent learning; reinforcement learning; survey

ACM Reference Format:
Pablo Hernandez-Leal, Bilal Kartal, and Matthew E. Taylor. 2020. A Very Condensed Survey and Critique of Multiagent Deep Reinforcement Learning. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Almost 20 years ago Stone and Veloso's seminal survey [47] laid the groundwork for defining the area of multiagent systems (MAS) and its open problems in the context of AI. About ten years ago, Shoham, Powers, and Grenager [44] noted that the literature on multiagent learning (MAL) was growing significantly. Since then, the number of published MAL works continues to steadily rise, which led to different surveys on the area, ranging from analyzing the basics of MAL and their challenges [4, 13, 51], to addressing specific subareas: game theory and MAL [38, 44], cooperative scenarios [36, 39], and evolutionary dynamics of MAL [10]. The last couple of years three surveys related to MAL have been published: learning in non-stationary environments [27], agents modeling agents [3], and transfer learning in multiagent reinforcement learning (RL) [45].

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

While different techniques and algorithms were used in the above scenarios, in general, they are all a combination of techniques from two main areas: RL [48] and deep learning [33, 42].

RL is an area of machine learning where an agent learns by interacting (i.e., taking actions) within a dynamic environment. However, one of the main challenges to RL, and traditional machine learning in general, is the need for manually designing high-quality features on which to learn. Deep learning enables efficient representation learning, thus allowing the automatic discovery of features [33, 42].

In deep reinforcement learning (DRL) [6, 20] deep neural networks are trained to approximate the optimal policy and/or the value function. In this way the deep neural network, serving as function approximator, enables powerful generalization.

DRL has been regarded as an important component in constructing general AI systems and has been successfully integrated with other techniques, e.g., search [46], planning [50], and more recently with multiagent systems, with an emerging area of *multiagent deep reinforcement learning (MDRL)* [37, 40]. However, learning in multiagent settings is fundamentally more difficult than the single-agent case due to the presence of multiagent pathologies, e.g., the moving target problem (non-stationarity) [13], curse of dimensionality [44], multiagent credit assignment [53], global exploration [36], and relative overgeneralization [21].

2 A SURVEY OF MDRL

We identified four categories to group recent MDRL works:

- *Analysis of emergent behaviors.* These works, in general, do not propose learning algorithms — their main focus is to analyze and evaluate single-agent DRL algorithms, e.g., DQN, in a multiagent environment. In this category we found works that analyze behaviors in the three major settings: cooperative, competitive, and mixed scenarios.
- *Learning communication.* These works explore a sub-area in which agents can share information with communication protocols, for example through direct messages or via a shared memory.
- *Learning cooperation.* While learning to communicate is an emerging area, fostering cooperation in learning agents has a long history of research in MAL [36, 39]. In this category the analyzed works are evaluated in either cooperative or mixed settings.
- *Agents modeling agents.* Albrecht and Stone [3] presented a thorough survey in this topic and we have found many works that fit into this category in the MDRL setting, some taking inspiration from DRL, and others from MAL. Modeling agents is helpful not only to cooperate, but also for modeling

opponents for improved best-response, inferring goals, and accounting for the learning behavior of other agents. In this category the analyzed algorithms present their results in either a competitive setting or a mixed one (cooperative and competitive).

For each category, our survey [28] provides a full description as well as a outlines recent works. Then, we take a step back and reflect on how these new works relate to the existing literature.

3 A CRITIQUE OF MDRL

First, we address the pitfall of *deep learning amnesia*, roughly described as missing citations to the original works and not exploiting the advancements that have been made in the past, i.e., pre 2010s. We provide some specific examples of research milestones that were studied earlier, e.g., RL or MAL, and that now became highly relevant for MDRL, such as:

- Dealing with non-stationarity in independent learners [32]
- Multiagent credit assignment [2]
- Multitask learning [14]
- Auxiliary tasks [49]
- Experience replay [35]
- Double estimators [25]

Next, we take a more critical view with respect to MDRL and highlight different practical challenges that already happen in DRL and that are likely to occur in MDRL.

Reproducibility, troubling trends, and negative results. Reproducibility is a challenge in RL that is only aggravated in DRL due to different sources of stochasticity: baselines, hyperparameters, architectures, and random seeds. Moreover, DRL does not have common practices for statistical testing which has led to bad reporting practices (i.e., cherry picking [7]). We believe that together with following the advice on how to design experiments and report results, the community would also benefit from reporting *negative results* [19, 22, 43] for carefully designed hypothesis and experiments.

Implementation challenges and hyperparameter tuning. One problem is that canonical implementations of DRL algorithms often contain additional non-trivial optimizations — these are sometimes necessary for the algorithms to achieve good performance [30]. The effects of hyperparameter tuning were analyzed in more detail in DRL by Henderson et al. [26], who concluded that hyperparameters can have significantly different effects across algorithms and environments since there is an intricate interplay among them. Note that hyperparameter tuning is related to the troubling trend of *cherry picking* in that it can show a carefully picked set of parameters that make an algorithm work. Lastly, note that hyperparameter tuning is computationally very expensive, which relates to the challenge of computational demands.

Computational resources. Deep RL usually requires millions of interactions for an agent to learn [5], i.e., low sample efficiency [54], which highlights the need for large computational infrastructure in general. However, computational infrastructure is a major bottleneck for performing DRL and MDRL research, especially for those who do not have such large compute power (e.g., most companies and most academic research groups) [9, 43].

In the end, we believe that high compute based methods act as a frontier to showcase benchmarks [1, 52], i.e., they show what results are possible as data and compute is scaled up; however, lighter compute based algorithmic methods can also yield significant contributions to better tackle real-world problems.

Occam's razor and ablative analysis. Finding the simplest context that exposes the innovative research ideas is challenging, and if ignored, leads to a conflation of fundamental research (working principles in the most abstract setting) and applied research (working systems as complete as possible). In particular, some deep learning papers are presented as learning from pixels without further explanation, while object-level representations would have already exposed the algorithmic contribution [16]. This still makes sense to remain comparable with established benchmarks (e.g., OpenAI Gym), but less so if custom simulations are written without open source access, as it introduces unnecessary variance in pixel-level representations and artificially inflates computational resources.

Finally, we conclude with some open questions for MDRL.

- On the challenge of sparse and delayed rewards.
  Recent MDRL competitions and environments have complex scenarios where many actions are taken before a reward signal is available. This sparseness is already a challenge for RL [18, 48] and in MDRL this is even more problematic since the agents not only need to learn basic behaviors, but also to learn the strategic element (e.g., competitive/collaborative) embedded in the multiagent setting.
- On the role of self-play.
  Self-play is a cornerstone in MAL with impressive results [12, 15, 23, 29]. While notable results had also been shown in MDRL [11], recent works have also shown that *plain* self-play does not yield the best results. However, adding diversity, i.e., evolutionary methods [8, 34, 41] or sampling-based methods, have shown good results. A drawback of these solutions is the additional computational requirements since they need either parallel training (more CPU computation) or memory requirements.
- On the challenge of the combinatorial nature of MDRL.
  To learn complex multiagent interactions some type of abstraction [17] is often needed, for example, factored value functions [24, 31] try to exploit independence among agents through (factored) structure; however, in MDRL there are still open questions such as understanding their representational power (e.g., the accuracy of the learned Q-function approximations) and how to learn those factorizations.

4 CONCLUSIONS

Our view is that there are practical issues within MDRL that hinder its scientific progress: the necessity of high compute power, complicated reproduciblity, and the lack of sufficient encouragement for publishing negative results. However, we remain highly optimistic about the multiagent community and hope this work serves to raise those issues, promote good solutions, and ultimately take advantage of the existing literature and resources available to move the area in the most promising directions.

REFERENCES

[1] 2018. Open AI Five. https://blog.openai.com/openai-five. (2018). [Online; accessed 7-September-2018].
[2] Adrian K Agogino and Kagan Tumer. 2008. Analyzing and visualizing multiagent rewards in dynamic and stochastic domains. *Autonomous Agents and Multi-Agent Systems* 17, 2 (2008), 320–338.
[3] Stefano V. Albrecht and Peter Stone. 2018. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence* 258 (Feb. 2018), 66–95.
[4] Eduardo Alonso, Mark D'inverno, Daniel Kudenko, Michael Luck, and Jason Noble. 2002. Learning in multi-agent systems. *Knowledge Engineering Review* 16, 03 (Feb. 2002), 1–8.
[5] Dario Amodei and Danny Hernandez. 2018. AI and Compute. (2018). https://blog.openai.com/ai-and-compute
[6] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. 2017. A Brief Survey of Deep Reinforcement Learning . (2017). http://arXiv.org/abs/1708.05866v2
[7] Kamyar Azizzadenesheli. 2019. Maybe a few considerations in Reinforcement Learning Research?. In *Reinforcement Learning for Real Life Workshop*.
[8] Thomas Back. 1996. *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press.
[9] Edward Beeching, Christian Wolf, Jilles Dibangoye, and Olivier Simonin. 2019. Deep Reinforcement Learning on a Budget: 3D Control and Reasoning Without a Supercomputer. *CoRR* abs/1904.01806 (2019). arXiv:1904.01806 http://arxiv.org/abs/1904.01806
[10] Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. 2015. Evolutionary Dynamics of Multi-Agent Learning: A Survey. *Journal of Artificial Intelligence Research* 53 (2015), 659–697.
[11] Michael Bowling, Neil Burch, Michael Johanson, and O Tammelin. 2015. Heads-up limit hold'em poker is solved. *Science* 347, 6218 (2015), 145–149.
[12] Michael Bowling and Manuela Veloso. 2002. Multiagent learning using a variable learning rate. *Artificial Intelligence* 136, 2 (2002), 215–250.
[13] Lucian Busoniu, Robert Babuska, and Bart De Schutter. 2008. A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* 38, 2 (2008), 156–172.
[14] Rich Caruana. 1997. Multitask learning. *Machine learning* 28, 1 (1997), 41–75.
[15] Vincent Conitzer and Tuomas Sandholm. 2006. AWESOME: A general multiagent learning algorithm that converges in self-play and learns a best response against stationary opponents. *Machine Learning* 67, 1-2 (2006), 23–43.
[16] Giuseppe Cuccu, Julian Togelius, and Philippe Cudré-Mauroux. 2019. Playing atari with six neurons. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 998–1006.
[17] Yann-Michaël De Hauwere, Peter Vrancx, and Ann Nowe. 2010. Learning multi-agent state space representations. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*. Toronto, Canada, 715–722.
[18] Adrien Ecoffet, Joost Huizinga, Joel Lehman, Kenneth O Stanley, and Jeff Clune. 2019. Go-Explore: a New Approach for Hard-Exploration Problems. *arXiv preprint arXiv:1901.10995* (2019).
[19] Jessica Zosa Forde and Michela Paganini. 2019. The Scientific Method in the Science of Machine Learning. In *ICLR Debugging Machine Learning Models workshop*.
[20] Vincent François-Lavet, Peter Henderson, Riashat Islam, Marc G Bellemare, Joelle Pineau, et al. 2018. An Introduction to Deep Reinforcement Learning. *Foundations and Trends® in Machine Learning* 11, 3-4 (2018), 219–354.
[21] Nancy Fulda and Dan Ventura. 2007. Predicting and Preventing Coordination Problems in Cooperative Q-learning Systems. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*. Hyderabad, India, 780–785.
[22] Oguzhan Gencoglu, Mark van Gils, Esin Guldogan, Chamin Morikawa, Mehmet Süzen, Mathias Gruber, Jussi Leinonen, and Heikki Huttunen. 2019. HARK Side of Deep Learning–From Grad Student Descent to Automated Machine Learning. *arXiv preprint arXiv:1904.07633* (2019).
[23] Amy Greenwald and Keith Hall. 2003. Correlated Q-learning. In *Proceedings of 17th International Conference on Autonomous Agents and Multiagent Systems*. Washington, DC, USA, 242–249.
[24] Carlos Guestrin, Michail Lagoudakis, and Ronald Parr. 2002. Coordinated reinforcement learning. In *ICML*, Vol. 2. 227–234.
[25] Hado V Hasselt. 2010. Double Q-learning. In *Advances in Neural Information Processing Systems*. 2613–2621.
[26] Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. 2018. Deep Reinforcement Learning That Matters.. In *32nd AAAI Conference on Artificial Intelligence*.
[27] Pablo Hernandez-Leal, Michael Kaisers, Tim Baarslag, and Enrique Munoz de Cote. 2017. A Survey of Learning in Multiagent Environments - Dealing with Non-Stationarity. (2017). http://arxiv.org/abs/1707.09183
[28] Pablo Hernandez-Leal, Bilal Kartal, and Matthew E. Taylor. 2019. A Survey and Critique of Multiagent Deep Reinforcement Learning. *Journal of Autonomous Agents and Multiagent Systems* 33 (October 2019), 750–797. Issue 6.
[29] Junling Hu and Michael P. Wellman. 2003. Nash Q-learning for general-sum stochastic games. *The Journal of Machine Learning Research* 4 (Dec. 2003), 1039–1069.
[30] Andrew Ilyas, Logan Engstrom, Shibani Santurkar, Dimitris Tsipras, Firdaus Janoos, Larry Rudolph, and Aleksander Madry. 2018. Are Deep Policy Gradient Algorithms Truly Policy Gradient Algorithms? *CoRR* abs/1811.02553 (2018). arXiv:1811.02553 http://arxiv.org/abs/1811.02553
[31] Jelle R Kok and Nikos Vlassis. 2004. Sparse cooperative Q-learning. In *Proceedings of the twenty-first international conference on Machine learning*. ACM, 61.
[32] Guillaume J Laurent, Laëtitia Matignon, Le Fort-Piat, et al. 2011. The world of independent learners is not Markovian. *International Journal of Knowledge-based and Intelligent Engineering Systems* 15, 1 (2011), 55–64.
[33] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436.
[34] Joel Lehman and Kenneth O Stanley. 2008. Exploiting open-endedness to solve problems through the search for novelty. In *ALIFE*. 329–336.
[35] Long-Ji Lin. 1992. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine learning* 8, 3-4 (1992), 293–321.
[36] Laetitia Matignon, Guillaume J Laurent, and Nadine Le Fort-Piat. 2012. Independent reinforcement learners in cooperative Markov games: a survey regarding coordination problems. *Knowledge Engineering Review* 27, 1 (Feb. 2012), 1–31.
[37] Thanh Thi Nguyen, Ngoc Duy Nguyen, and Saeid Nahavandi. 2018. Deep Reinforcement Learning for Multi-Agent Systems: A Review of Challenges, Solutions and Applications. *arXiv preprint arXiv:1812.11794* (2018).
[38] Ann Nowé, Peter Vrancx, and Yann-Michaël De Hauwere. 2012. Game theory and multi-agent reinforcement learning. In *Reinforcement Learning*. Springer, 441–470.
[39] Liviu Panait and Sean Luke. 2005. Cooperative Multi-Agent Learning: The State of the Art. *Autonomous Agents and Multi-Agent Systems* 11, 3 (Nov. 2005).
[40] Georgios Papoudakis, Filippos Christianos, Arrasy Rahman, and Stefano V Albrecht. 2019. Dealing with Non-Stationarity in Multi-Agent Deep Reinforcement Learning. *arXiv preprint arXiv:1906.04737* (2019).
[41] Christopher D Rosin and Richard K Belew. 1997. New methods for competitive coevolution. *Evolutionary computation* 5, 1 (1997), 1–29.
[42] Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural networks* 61 (2015), 85–117.
[43] D Sculley, Jasper Snoek, Alex Wiltschko, and Ali Rahimi. 2018. Winner's curse? On pace, progress, and empirical rigor. In *ICLR Workshop*.
[44] Yoav Shoham, Rob Powers, and T. Grenager. 2007. If multi-agent learning is the answer, what is the question? *Artificial Intelligence* 171, 7 (2007), 365–377.
[45] Felipe L.D. Silva and Anna Helena Reali Costa. 2019. A Survey on Transfer Learning for Multiagent Reinforcement Learning Systems. *Journal of Artificial Intelligence Research* 64 (2019), 645–703.
[46] David Silver, A Huang, C J Maddison, A Guez, L Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.
[47] Peter Stone and Manuela M Veloso. 2000. Multiagent Systems - A Survey from a Machine Learning Perspective. *Autonomous Robots* 8, 3 (2000), 345–383.
[48] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
[49] Richard S Sutton, Joseph Modayil, Michael Delp, Thomas Degris, Patrick M Pilarski, Adam White, and Doina Precup. 2011. Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. 761–768.
[50] Aviv Tamar, Sergey Levine, Pieter Abbeel, Yi Wu, and Garrett Thomas. 2016. Value Iteration Networks. *NIPS* (2016), 2154–2162.
[51] Karl Tuyls and Gerhard Weiss. 2012. Multiagent learning: Basics, challenges, and prospects. *AI Magazine* 33, 3 (2012), 41–52.
[52] Oriol Vinyals, Igor Babuschkin, Junyoung Chung, Michael Mathieu, Max Jaderberg, Wojciech M. Czarnecki, Andrew Dudzik, Aja Huang, Petko Georgiev, Richard Powell, Timo Ewalds, Dan Horgan, Manuel Kroiss, Ivo Danihelka, John Agapiou, Junhyuk Oh, Valentin Dalibard, David Choi, Laurent Sifre, Yury Sulsky, Sasha Vezhnevets, James Molloy, Trevor Cai, David Budden, Tom Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Toby Pohlen, Yuhuai Wu, Dani Yogatama, Julia Cohen, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Chris Apps, Koray Kavukcuoglu, Demis Hassabis, and David Silver. 2019. AlphaStar: Mastering the Real-Time Strategy Game StarCraft II. https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/. (2019).
[53] David H Wolpert and Kagan Tumer. 2002. Optimal payoff functions for members of collectives. In *Modeling complexity in economic and social systems*. 355–369.
[54] Yang Yu. 2018. Towards Sample Efficient Reinforcement Learning.. In *IJCAI*. 5739–5743.