

REFERENCES

- [1] Pieter Abbeel, Adam Coates, Morgan Quigley, and Andrew Y Ng. 2007. An application of reinforcement learning to aerobatic helicopter flight. In *Advances in Neural Information Processing Systems*. 1–8.
- [2] Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. 2018. Safe reinforcement learning via shielding. In *AAAI*. 2669–2678.
- [3] Eitan Altman. 1999. *Constrained Markov decision processes*. Vol. 7. CRC Press.
- [4] Christel Baier, Joost-Pieter Katoen, and Kim Guldstrand Larsen. 2008. *Principles of Model Checking*. MIT Press.
- [5] Dimitri P Bertsekas and John N Tsitsiklis. 1996. *Neuro-dynamic Programming*. Vol. 1. Athena Scientific.
- [6] Suda Bharadwaj, Stéphane Le Roux, Guillermo Pérez, and Ufuk Topcu. 2017. Reduction techniques for model checking and learning in MDPs. In *IJCAI*. 4273–4279.
- [7] Rolando Cavazos-Cadena, Eugene A. Feinberg, and Raúl Montes-De-Oca. 2000. A note on the existence of optimal policies in total reward dynamic programs with compact action sets. *Mathematics of Operations Research* 25, 4 (2000), 657–666.
- [8] Richard Cheng, Gábor Orosz, Richard M Murray, and Joel W Burdick. 2019. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 3387–3395.
- [9] Yinlam Chow, Ofir Nachum, Edgar Duenez-Guzman, and Mohammad Ghavamzadeh. 2018. A Lyapunov-based approach to safe reinforcement learning. In *Advances in neural information processing systems*. 8092–8101.
- [10] Edmund M Clarke Jr, Orna Grumberg, Daniel Kroening, Doron Peled, and Helmut Veith. 2018. *Model checking*. MIT Press.
- [11] Stefano P Coraluppi and Steven I Marcus. 1999. Risk-sensitive and minimax control of discrete-time, finite-state Markov decision processes. *Automatica* 35, 2 (1999), 301–309.
- [12] Arthur P Dempster, Nan M Laird, and Donald B Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39, 1 (1977), 1–22.
- [13] Nathan Fulton. 2018. *Verifiably Safe Autonomy for Cyber-Physical Systems*. Ph.D. Dissertation. Carnegie Mellon University Pittsburgh, PA.
- [14] Nathan Fulton and André Platzer. 2018. Safe reinforcement learning via formal methods: Toward safe control through proof and learning. In *AAAI*. 6485–6492.
- [15] Nathan Fulton and Andre Platzer. 2019. Verifiably Safe Off-Model Reinforcement Learning. In *Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, Vol. 11427. Springer, 413–430.
- [16] Javier Garcia and Fernando Fernández. 2012. Safe exploration of state and action spaces in reinforcement learning. *JAIR* 45 (2012), 515–564.
- [17] Javier Garcia and Fernando Fernández. 2015. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research* 16, 1 (2015), 1437–1480.
- [18] Peter Geibel and Fritz Wysotzki. 2005. Risk-sensitive reinforcement learning applied to control under constraints. *Journal of Artificial Intelligence Research* 24 (2005), 81–108.
- [19] Mohammadhosein Hasanbeig, Alessandro Abate, and Daniel Kroening. 2018. Logically-constrained reinforcement learning. *arXiv preprint arXiv:1801.08099* (2018).
- [20] Mohammadhosein Hasanbeig, Alessandro Abate, and Daniel Kroening. 2019. Certified reinforcement learning with logic guidance. *arXiv preprint arXiv:1902.00778* (2019).
- [21] Mohammadhosein Hasanbeig, Alessandro Abate, and Daniel Kroening. 2019. Logically-Constrained Neural Fitted Q-Iteration. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2012–2014.
- [22] Mohammadhosein Hasanbeig, Yiannis Kantaros, Alessandro Abate, Daniel Kroening, George J Pappas, and Insup Lee. 2019. Reinforcement Learning for Temporal Logic Control Synthesis with Probabilistic Satisfaction Guarantees. In *Proceedings of the 58th Conference on Decision and Control*. IEEE, 5338–5343.
- [23] Mohammadhosein Hasanbeig and Laca Pavel. 2017. On Synchronous Binary Log-Linear Learning and Second Order Q-learning. In *The 20th World Congress of the International Federation of Automatic Control*. IFAC, 8987–8992.
- [24] Nils Jansen, Bettina Könighofer, Sebastian Junges, and Roderick Bloem. 2018. Shielded decision-making in MDPs. *arXiv preprint arXiv:1807.06096* (2018).
- [25] Xiao Li and Calin Belta. 2019. Temporal logic guided safe reinforcement learning using control barrier functions. *arXiv preprint arXiv:1903.09885* (2019).
- [26] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fiedjeland, Georg Ostrovski, et al. 2015. Human-level Control Through Deep Reinforcement Learning. *Nature* 518, 7540 (2015), 529–533.
- [27] Teodor Mihai Moldovan and Pieter Abbeel. 2012. Safe exploration in Markov decision processes. *arXiv preprint arXiv:1205.4810* (2012).
- [28] Martin Pecka and Tomas Svoboda. 2014. Safe exploration techniques for reinforcement learning—an overview. In *International Workshop on Modelling and Simulation for Autonomous Systems*. Springer, 357–375.
- [29] Silviu Pitis. 2019. Rethinking the discount factor in reinforcement learning: A decision theoretic approach. *arXiv preprint arXiv:1902.02893* (2019).
- [30] André Platzer. 2008. Differential dynamic logic for hybrid systems. *Journal of Automated Reasoning* 41, 2 (2008), 143–189.
- [31] Amir Pnueli. 1977. The temporal logic of programs. In *Foundations of Computer Science*. IEEE, 46–57.
- [32] Martin L. Puterman. 2014. *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons.
- [33] Salomon Sickert, Javier Esparza, Stefan Jaax, and Jan Křetínský. 2016. Limit-deterministic Büchi automata for linear temporal logic. In *CAV*. Springer, 312–332.
- [34] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.
- [35] Alexander L Strehl, Lihong Li, and Michael L Littman. 2009. Reinforcement learning in finite MDPs: PAC analysis. *Journal of Machine Learning Research* 10, Nov (2009), 2413–2444.
- [36] Richard S. Sutton and Andrew G. Barto. 1998. *Reinforcement Learning: An Introduction*. Vol. 1. MIT Press Cambridge.
- [37] Edward L. Thorndike. 1932. *The Fundamentals of Learning*. Teachers College Bureau of Publications.
- [38] Matteo Turchetta, Felix Berkenkamp, and Andreas Krause. 2016. Safe exploration in finite Markov decision processes with Gaussian processes. In *Advances in Neural Information Processing Systems*. 4312–4320.
- [39] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine Learning* 8, 3-4 (1992), 279–292.
- [40] Qingda Wei and Xianping Guo. 2011. Markov decision processes with state-dependent discount factors and unbounded rewards/costs. *Operations Research Letters* 39, 5 (2011), 369–374.
- [41] Min Wen and Ufuk Topcu. 2018. Constrained cross-entropy method for safe reinforcement learning. In *Advances in Neural Information Processing Systems*. 7450–7460.
- [42] Naoto Yoshida, Eiji Uchibe, and Kenji Doya. 2013. Reinforcement learning with state-dependent discount factor. In *ICDL*. IEEE, 1–6.
- [43] Zhenpeng Zhou, Xiaocheng Li, and Richard N. Zare. 2017. Optimizing chemical reactions with deep reinforcement learning. *ACS Central Science* 3, 12 (2017), 1337–1344.