



















- [29] David Silver, Richard S Sutton, and Martin Müller. 2008. Sample-based learning and search with permanent and transient memories. In *Proceedings of the 25th international conference on Machine learning*. 968–975.
- [30] Alexander L. Strehl, Carlos Diuk, and Michael L. Littman. 2007. Efficient structure learning in factored-state MDPs. In *AAAI*, Vol. 7. 645–650.
- [31] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Flores Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2018. Value-Decomposition Networks For Cooperative Multi-Agent Learning Based On Team Reward.. In *AAMAS*. 2085–2087.
- [32] Richard S. Sutton. 1988. Learning to predict by the methods of temporal differences. *Machine learning* 3, 1 (1988), 9–44.
- [33] Richard S. Sutton. 1990. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Machine Learning Proceedings 1990*. Elsevier, 216–224.
- [34] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- [35] Harm van Seijen and Richard S. Sutton. 2013. Efficient Planning in MDPs by Small Backups. In *Proc. 30th Int. Conf. Mach. Learn.*, Vol. 28. III–361.
- [36] Timothy Verstraeten, Eugenio Bargiacchi, Pieter J. K. Libin, Jan Helsen, Diederik M. Roijers, and Ann Nowé. 2020. Multi-Agent Thompson Sampling for Bandit Applications with Sparse Neighbourhood Structures. *Nature Scientific Reports* 10, 1 (2020), 6728.
- [37] Christopher M. Vigorito and Andrew G. Barto. 2009. Incremental structure learning in factored MDPs with continuous states and actions. *University of Massachusetts Amherst-Department of Computer Science, Tech. Rep* (2009).
- [38] Christopher J. C. H. Watkins. 1989. *Learning from delayed rewards*. Ph.D. Dissertation. King’s College, Cambridge.
- [39] Marco A. Wiering. 2000. Multi-agent reinforcement learning for traffic light control. In *Machine Learning: Proceedings of the Seventeenth International Conference (ICML’2000)*. 1151–1158.