附录A 算法总结表

在附录 A 中，我们将那些常见的强化学习算法总结成一张表格，尤其是那些在本书中介绍过的。我们希望这样能为读者寻找相关文献提供参考。

强化学习算法

强化学习算法	策略	动作空间	年份	文献	作者
Q-Learning	离线策略（Off-Policy）	离散	1992	Q-Learning (Watkins et al., 1992)	Cristopher J.C.H Watkins and Peter Dayan
SARSA	在线策略（On-Policy）	离散	1994	Online Q-Learning using Connectionist Systems (Rummery et al., 1994)	G.A.Rummery and M. Niranjan
DQN	离线策略	离散	2015	Human-level Control Through Deep Reinforcement Learning (Mnih et al., 2015)	Volodymyr Mnih, et al.
Dueling DQN	离线策略	离散	2015	Dueling Network Architectures for Deep Reinforcement Learning (Wang et al., 2015)	Ziyu Wang, et al.
Double DQN	离线策略	离散	2016	Deep Reinforcement Learning with Double Q-Learning (Van Hasselt et al., 2016)	Hado van Hasselt, et al.
Noisy DQN	离线策略	离散	2017	Noisy Networks for Exploration (Fortunato et al., 2017)	Meire Fortunato, et al.
Distributed DQN	离线策略	离散	2017	A Distributional Perspective on Reinforcement Learning (Bellemare et al., 2017)	Marc G. Bellemare, et al.
Actor-Critic (QAC)	在线策略	离散或连续	2000	Actor-Critic Algorithms (Konda et al., 2000)	Vijay R. Konda and John N. Tsitsiklis
A3C	在线策略	离散或连续	2016	Asynchronous Methods for Deep Reinforcement Learning (Mnih et al., 2016)	Volodymyr Mnih, et al.
DDPG	离线策略	连续	2016	Continuous Control With Deep Reinforcement Learning (Lillicrap et al., 2015)	Timothy P. Lillicrap, et al.

（续表一）

强化学习算法

强化学习算法	策略	动作空间	年份	文献	作者
REINFORCE	在线策略	离散或连续	1988	On the Use of Backpropagation in Associative Reinforcement Learning (Williams, 1988)	Ronald J. Williams
TD3	离线策略	连续	2018	Addressing function approximation error in actor-critic methods (Fujimoto et al., 2018)	Scott Fujimoto, et al.
SAC	离线策略	离散或连续	2018	Soft actor-critic algorithms and applications (Haarnoja et al., 2018)	Tuomas Haarnoja, et al.
TRPO	在线策略	离散或连续	2015	Trust region policy optimization (Schulman et al., 2015)	John Schulman, et al.
PPO	在线策略	离散或连续	2017	Proximal policy optimization algorithms (Schulman et al., 2017)	John Schulman, et al.
DPPO	在线策略	离散或连续	2017	Emergence of locomotion behaviours in rich environments (Heess et al., 2017)	Nicolas Heess, et al.
ACKTR	在线策略	离散或连续	2017	Scalable trust-region method for deep reinforcement learning using Kronecker-factored approximation (Wu et al., 2017)	Yuhuai Wu, et al.
CE Method	在线策略	离散或连续	2004	The cross-entropy method: A unified approach to Monte Carlo simulation, randomized optimization and machine learning (Rubinstein et al., 2004)	R. Rubinstein and D. Kroese

参考文献

BELLEMARE M G, DABNEY W, MUNOS R, 2017. A distributional perspective on reinforcement learning[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org: 449-458.

FORTUNATO M, AZAR M G, PIOT B, et al., 2017. Noisy networks for exploration[J]. arXiv preprint arXiv:1706.10295.

FUJIMOTO S, VAN HOOF H, MEGER D, 2018. Addressing function approximation error in actor-critic methods[J]. arXiv preprint arXiv:1802.09477.

HAARNOJA T, ZHOU A, HARTIKAINEN K, et al., 2018. Soft actor-critic algorithms and applications[J]. arXiv preprint arXiv:1812.05905.

HEESS N, SRIRAM S, LEMMON J, et al., 2017. Emergence of locomotion behaviours in rich environments[J]. arXiv:1707.02286.

KONDA V R, TSITSIKLIS J N, 2000. Actor-critic algorithms[C]//Advances in Neural Information Processing Systems. 1008-1014.

LILLICRAP T P, HUNT J J, PRITZEL A, et al., 2015. Continuous control with deep reinforcement learning[J]. arXiv preprint arXiv:1509.02971.

MNIH V, KAVUKCUOGLU K, SILVER D, et al., 2015. Human-level control through deep reinforcement learning[J]. Nature.

MNIH V, BADIA A P, MIRZA M, et al., 2016. Asynchronous methods for deep reinforcement learning[C]//International Conference on Machine Learning (ICML). 1928-1937.

RUBINSTEIN R Y, KROESE D P, 2004. The cross-entropy method: A unified approach to monte carlo simulation, randomized optimization and machine learning[J]. Information Science & Statistics, Springer Verlag, NY.

RUMMERY G A, NIRANJAN M, 1994. On-line q-learning using connectionist systems: volume 37[M]. University of Cambridge, Department of Engineering Cambridge, England.

SCHULMAN J, LEVINE S, ABBEEL P, et al., 2015. Trust region policy optimization[C]//International Conference on Machine Learning (ICML). 1889-1897.

SCHULMAN J, WOLSKI F, DHARIWAL P, et al., 2017. Proximal policy optimization algorithms[J]. arXiv:1707.06347.

VAN HASSELT H, GUEZ A, SILVER D, 2016. Deep reinforcement learning with double Q-learning[C]// Thirtieth AAAI conference on artificial intelligence.

WANG Z, SCHAUL T, HESSEL M, et al., 2015. Dueling network architectures for deep reinforcement learning[J]. arXiv preprint arXiv:1511.06581.

WATKINS C J, DAYAN P, 1992. Q-learning[J]. Machine learning, 8(3-4): 279-292.

WILLIAMS R J, 1988. On the use of backpropagation in associative reinforcement learning[C]// Proceedings of the IEEE International Conference on Neural Networks: volume 1. San Diego, CA.: 263-270.

WU Y, MANSIMOV E, GROSSE R B, et al., 2017. Scalable trust-region method for deep reinforcement learning using kronecker-factored approximation[C]//Advances in Neural Information Processing Systems. 5279-5288.