

Contents

1	Introduction	1
1.1	Evolution of Reinforcement Learning	1
1.1.1	Basic Reinforcement Learning	2
1.1.2	Inverse Reinforcement Learning	6
1.1.3	Multi-agent Reinforcement Learning	7
1.1.4	Meta Reinforcement Learning	9
1.1.5	Hierarchical Reinforcement Learning	10
1.1.6	Multi-Task Reinforcement Learning	11
1.2	Deep Reinforcement Learning	11
1.3	Applications of Reinforcement Learning	12
1.4	Summary	13
	References	13
2	Mathematics in Deep Reinforcement Learning	17
2.1	Reinforcement Learning Foundations	17
2.1.1	Basic Concepts	17
2.1.2	Model-based vs Model-free Methods	20
2.1.3	Tabular vs Approximate Methods	22
2.1.4	Value-based vs Policy-based Methods	24
2.1.5	Basic Reinforcement Learning Methods	24
2.2	Reinforcement Learning Branches	38
2.2.1	Multi-agent Reinforcement Learning	38
2.2.2	Meta Reinforcement Learning	38
2.2.3	Hierarchical Reinforcement Learning	38
2.2.4	Multi-Task Reinforcement Learning	38
2.2.5	Exploitation and Exploration	38
2.3	Deep Learning Foundations	38
2.3.1	Basic Concepts	39
2.3.2	Network Structure	42
2.4	Deep Reinforcement Learning Foundations	46
2.4.1	Value Networks	46
2.4.2	Policy Networks	47

2.4.3 Special Considerations	48
References	48
3 Deep Reinforcement Learning Models	51
3.1 Value Function Approximation	51
3.1.1 Deep Q-learning	51
3.1.2 Double DQN	53
3.1.3 More Advanced DQN Methods	55
3.2 Policy Approximation	58
3.2.1 Deep Policy Gradient Method	58
3.3 Hybrid Deep Reinforcement Learning Models	58
3.3.1 Deep Actor-Critic Method	58
3.4 Advanced Network Components	60
3.4.1 Deep Embeddings	60
3.4.2 Attention	60
3.4.3 Transformers in Reinforcement Learning Models	60
3.4.4 trust region policy optimisation	60
References	60
4 Empirical Deep Reinforcement Learning Systems	61
4.1 RL In Robotics	61
4.2 RL In Commerce	61
4.3 RL In Medications	61
4.4 RL In Education	61
4.5 RL In Civilization	61
4.6 RL In Simulations	61
5 Performance Evaluations	63
5.1 Offline Performance Metrics	63
5.2 Online Performance Metrics	63
5.3 A/B Testing	63
5.4 Interleaving	63
6 Advanced Topics	65
6.1 Memorization	65
6.2 Alignment of Target Network	65
6.3 Policy Drift and the Solutions	65
6.4 Sociality and Trust	65
6.5 Attack Resistance	65
6.6 Privacy Preservation	65
6.7 Personalization	65
6.8 Advanced Optimization Methods	65
6.9 Exploitation and Exploration Tradeoff	65
6.10 Learning Speedup	65

7	Empiriccal Researches Ongoing	67
7.1	Adaptations of Neuroscience	67
7.2	Future of Deep Reinforcement Learning	67
7.3	Hardware Optimization	67
A	Appendix	69
A.1	Math Appendix	69
	Glossary	71
	References	72
	Index	75