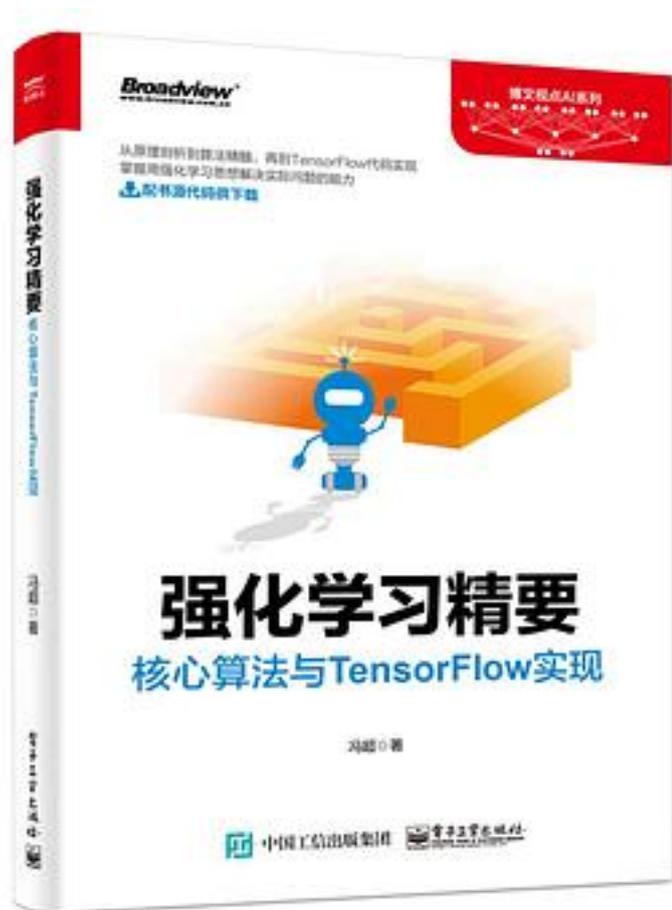


强化学习精要



[强化学习精要_下载链接1](#)

著者:冯超

出版者:电子工业出版社

出版时间:2018-6

装帧:平装

isbn:9787121340000

《强化学习精要：核心算法与TensorFlow实现》用通俗幽默的语言深入浅出地介绍了强化学习的基本算法与代码实现，为读者构建了一个完整的强化学习知识体系，同时介绍了这些算法的具体实现方式。从基本的马

尔可夫决策过程，到各种复杂的强化学习算法，读者都可以从本书中学习到。本书除了介绍这些算法的原理，还深入分析了算法之间的内在联系，可以帮助读者举一反三，掌握算法精髓。书中介绍的代码可以帮助读者快速将算法应用到实践中。

《强化学习精要：核心算法与TensorFlow实现》内容翔实，语言简洁易懂，既适合零基础的人员入门学习，也适合相关科研人员研究参考。

作者介绍:

目录: 第一部分强化学习入门与基础知识

- 1 引言2
- 1.1 强化学习的概念 2
 - 1.1.1 巴浦洛夫的狗 3
 - 1.1.2 俄罗斯方块 4
- 1.2 站在被实验者的角度看问题 5
- 1.3 强化学习效果的评估 8
 - 1.3.1 不断试错 8
 - 1.3.2 看重长期回报 8
- 1.4 强化学习与监督学习 9
 - 1.4.1 强化学习与监督学习的本质 9
 - 1.4.2 模仿学习 10
- 1.5 强化学习的实验环境 11
 - 1.5.1 Arcade Learning Environment 12
 - 1.5.2 Box2D 12
 - 1.5.3 MuJoCo 13
 - 1.5.4 Gym 14
- 1.6 本书的主要内容 15
- 1.7 参考资料 16
- 2 数学与机器学习基础17
- 2.1 线性代数基础 17
- 2.2 对称矩阵的性质 21
 - 2.2.1 特征值与特征向量 21
 - 2.2.2 对称矩阵的特征值和特征向量 22
 - 2.2.3 对称矩阵的对角化 23
- 2.3 概率论 24
 - 2.3.1 概率与分布 24
 - 2.3.2 最大似然估计 27
- 2.4 重要性采样 29
- 2.5 信息论基础 33
- 2.6 KL 散度 35
- 2.7 凸函数及其性质 39
- 2.8 机器学习的基本概念 41
- 2.9 机器学习的目标函数 43
- 2.10 总结 45
- 3 优化算法47
- 3.1 梯度下降法 47
 - 3.1.1 什么是梯度下降法 47
 - 3.1.2 优雅的步长 48
- 3.2 动量算法 53
- 3.3 共轭梯度法 59
 - 3.3.1 精妙的约束 59

- 3.3.2 共轭 60
- 3.3.3 优化步长的确定 63
- 3.3.4 Gram-Schmidt 方法 64
- 3.3.5 共轭梯度 65
- 3.4 自然梯度法 69
 - 3.4.1 基本概念 69
 - 3.4.2 Fisher 信息矩阵 71
 - 3.4.3 自然梯度法目标公式 76
- 3.5 总结 77
- 4 TensorFlow 入门78
 - 4.1 TensorFlow 的基本使用方法 78
 - 4.2 TensorFlow 原理介绍 82
 - 4.2.1 创建变量的scope 83
 - 4.2.2 创建一个Variable 背后的故事 89
 - 4.2.3 运算操作 94
 - 4.2.4 tf.gradients 96
 - 4.2.5 Optimizer 102
 - 4.2.6 TensorFlow 的反向传播技巧 106
 - 4.2.7 arg_scope 的使用 109
 - 4.3 TensorFlow 的分布式训练 113
 - 4.3.1 基于MPI 的数据并行模型 114
 - 4.3.2 MPI 的实现: mpi_adam 121
 - 4.4 基于TensorFlow 实现经典网络结构 122
 - 4.4.1 多层感知器 122
 - 4.4.2 卷积神经网络 124
 - 4.4.3 循环神经网络 126
 - 4.5 总结 129
 - 4.6 参考资料 129
- 5 Gym 与Baselines 130
 - 5.1 Gym 130
 - 5.1.1 Gym 的安装 130
 - 5.1.2 Gym 的基本使用方法 132
 - 5.1.3 利用Gym 框架实现一个经典的棋类游戏: 蛇棋 134
 - 5.2 Baselines 138
 - 5.2.1 Baselines 中的Python 3 新特性 139
 - 5.2.2 tf_util 141
 - 5.2.3 对Gym 平台的扩展 142
 - 5.3 总结 144
- 6 强化学习基本算法145
 - 6.1 马尔可夫决策过程 145
 - 6.1.1 MDP: 策略与环境模型 145
 - 6.1.2 值函数与Bellman 公式 147
 - 6.1.3 “表格式” Agent 151
 - 6.2 策略迭代 153
 - 6.2.1 策略迭代法 153
 - 6.2.2 策略提升的证明 159
 - 6.2.3 策略迭代的效果展示 160
 - 6.3 价值迭代 162
 - 6.3.1 N 轮策略迭代 162
 - 6.3.2 从动态规划的角度谈价值迭代 165
 - 6.3.3 价值迭代的实现 167
 - 6.4 泛化迭代 168
 - 6.4.1 两个极端 168
 - 6.4.2 广义策略迭代法 169

- 6.4.3 泛化迭代的实现 170
- 6.5 总结 171
- 第二部分最优价值算法
- 7 Q-Learning 基础173
- 7.1 状态转移概率：从掌握到放弃 173
- 7.2 蒙特卡罗方法 174
- 7.3 探索与利用 178
- 7.4 蒙特卡罗的方差问题 181
- 7.5 时序差分法与SARSA 183
- 7.6 Q-Learning 186
- 7.7 Q-Learning 的收敛性分析 189
- 7.8 从表格形式到价值模型 193
- 7.9 Deep Q Network 195
- 7.10 总结 202
- 7.11 参考资料 202
- 8 DQN 的改进算法203
- 8.1 Double Q-Learning 203
- 8.2 Priority Replay Buffer 204
- 8.3 Dueling DQN 209
- 8.4 解决DQN 的冷启动问题 211
- 8.5 Distributional DQN 214
- 8.5.1 输出价值分布 214
- 8.5.2 分布的更新 216
- 8.6 Noisy Network 218
- 8.7 Rainbow 221
- 8.7.1 Rainbow 的模型特点 221
- 8.7.2 Deep Q Network 的实现 223
- 8.8 总结 227
- 8.9 参考资料 227
- 第三部分基于策略梯度的算法
- 9 基于策略梯度的算法229
- 9.1 策略梯度法 229
- 9.1.1 算法推导 230
- 9.1.2 算法分析 233
- 9.1.3 算法改进 234
- 9.2 Actor-Critic 算法 236
- 9.2.1 降低算法的方差 236
- 9.2.2 A3C 算法 238
- 9.2.3 A2C 算法实战 240
- 9.3 总结 243
- 9.4 参考资料 243
- 10 使策略单调提升的优化算法244
- 10.1 TRPO 244
- 10.1.1 策略的差距 245
- 10.1.2 策略提升的目标公式 247
- 10.1.3 TRPO 的目标定义 248
- 10.1.4 自然梯度法求解 251
- 10.1.5 TRPO 的实现 254
- 10.2 GAE 256
- 10.2.1 GAE 的公式定义 256
- 10.2.2 基于GAE 和TRPO 的值函数优化 259
- 10.2.3 GAE 的实现 260
- 10.3 PPO 261
- 10.3.1 PPO 介绍 261

10.3.2 PPO 算法实践	263
10.4 总结	264
10.5 参考资料	264
11 Off-Policy 策略梯度法	265
11.1 Retrace	266
11.1.1 Retrace 的基本概念	266
11.1.2 Retrace 的算法实现	267
11.2 ACER	270
11.2.1 Off-Policy Actor-Critic	270
11.2.2 ACER 算法	272
11.2.3 ACER 的实现	276
11.3 DPG	279
11.3.1 连续空间的策略优化	279
11.3.2 策略模型参数的一致性	280
11.3.3 DDPG 算法	283
11.3.4 DDPG 的实现	286
11.4 总结	289
11.5 参考资料	289
第四部分其他强化学习算法	
12 稀疏回报的求解方法	291
12.1 稀疏回报的困难	291
12.2 层次强化学习	294
12.3 HER	298
12.3.1 渐进式学习	299
12.3.2 HER 的实现	301
12.4 总结	304
12.5 参考资料	304
13 Model-based 方法	305
13.1 AlphaZero	305
13.1.1 围棋游戏	305
13.1.2 Alpha-Beta 树	307
13.1.3 MCTS	309
13.1.4 策略价值模型	312
13.1.5 模型的对决	316
13.2 iLQR	316
13.2.1 线性模型的求解法	317
13.2.2 非线性模型的解法	322
13.2.3 iLQR 的实现	325
13.3 总结	328
13.4 参考资料	328
第五部分反向强化学习	
14 反向强化学习入门	330
14.1 基本概念	330
14.2 从最优策略求解回报	332
14.2.1 求解回报的目标函数	332
14.2.2 目标函数的约束	334
14.3 求解线性规划	335
14.3.1 线性规划的求解过程	335
14.3.2 实际案例	337
14.4 无限状态下的求解	338
14.5 从样本中学习	342
14.6 总结	344
14.7 参考资料	344
15 反向强化学习算法2.0	345

- 15.1 最大熵模型 345
 - 15.1.1 指数家族 346
 - 15.1.2 最大熵模型的推导 349
 - 15.1.3 最大熵模型的实现 354
- 15.2 最大熵反向强化学习 356
- 15.3 GAIL 361
 - 15.3.1 GAN 的基本概念 361
 - 15.3.2 GAN 的训练分析 363
- 15.4 GAIL 实现 367
- 15.5 总结 370
- 15.6 参考资料 370

• • • • • [\(收起\)](#)

[强化学习精要_下载链接1](#)

标签

强化学习

人工智能

算法

计算机

tensorflow

工程

评论

并不觉得好，国内真的没有强化学习相关的好书。

讲了一半线性代数，概率论还有tensorflow基础，讲到强化学习的时候又开始跳过不少东西了

需要一点基础，但是我觉得整本书的脉络很清晰，我推荐看一下国科大的沈华伟讲的高级人工智能的强化学习部分，再来看这本书会轻松一些。不足之处在于理论的阐述还不够清晰

忽悠初学者的书罢了，从DQN往后的内容像是论文翻译。

[强化学习精要_下载链接1](#)

书评

[强化学习精要_下载链接1](#)