

Оглавление

Об авторе	14
О научных редакторах.....	14
Предисловие	16
Для кого написана эта книга	16
Что в книге.....	16
Необходимое программное обеспечение	18
Загрузка файлов с примерами кода	19
От издательства	19
Глава 1. Введение в обучение с подкреплением.....	20
Что такое RL?	20
Алгоритм RL	22
Чем RL отличается от других парадигм машинного обучения	23
Элементы RL.....	24
Агент.....	24
Функция политики.....	24
Функция ценности.....	24
Модель.....	25
Интерфейс агента со средой.....	25
Типы сред RL.....	27
Детерминированная среда	27
Стохастическая среда.....	27
Среда с полной информацией	27

Среда с неполной информацией.....	27
Дискретная среда.....	28
Непрерывная среда.....	28
Эпизодические и неэпизодические среды.....	28
Одноагентные и многоагентные среды.....	28
Платформы RL.....	28
OpenAI Gym и Universe.....	29
DeepMind Lab.....	29
RL-Glue.....	29
Проект Malmo.....	29
ViZDoom.....	30
Практическое применение RL.....	30
Образование.....	30
Медицина и здравоохранение.....	30
Производство.....	31
Управление ресурсами.....	31
Финансы.....	31
Обработка естественного языка и машинное распознавание образов.....	31
Итоги.....	32
Вопросы.....	32
Дополнительные источники.....	32
Глава 2. Знакомство с OpenAI и TensorFlow.....	33
Подготовка системы.....	34
Установка Anaconda.....	34
Установка Docker.....	35
Установка OpenAI Gym и Universe.....	36
OpenAI Gym.....	39
Базовое моделирование.....	39
Робот учится ходить.....	41
OpenAI Universe.....	44
Построение бота для видеоигры.....	44
TensorFlow.....	48
Переменные, константы и заместители.....	49
Граф вычислений.....	50

Сеансы	51
TensorBoard	52
Итоги.....	55
Вопросы	56
Дополнительные источники.....	56
Глава 3. Марковский процесс принятия решений и динамическое программирование	57
Марковские цепи и марковские процессы.....	57
Марковский процесс принятия решений	59
Награды и возврат	60
Эпизодические и непрерывные задачи	61
Поправочный коэффициент.....	61
Функция политики.....	62
Функция ценности состояния	62
Функция ценности состояния/действия (Q-функция)	63
Уравнение Беллмана и оптимальность.....	64
Вывод уравнения Беллмана для функции ценности и Q-функции	65
Решение уравнения Беллмана	68
Динамическое программирование	68
Решение задачи о замерзшем озере	76
Итерация по ценности.....	78
Итерация по политикам.....	84
Итоги.....	87
Вопросы	88
Дополнительные источники.....	88
Глава 4. Методы Монте-Карло в играх.....	89
Метод Монте-Карло	89
Оценка значения π методом Монте-Карло	90
Прогнозирование методом Монте-Карло.....	94
Метод Монте-Карло с первым посещением.....	96
Метод Монте-Карло с каждым посещением	96
Игра в блек-джек по стратегии Монте-Карло.....	96

Управление методом Монте-Карло	105
MC-ES.....	106
Метод Монте-Карло с привязкой к политике.....	108
Метод Монте-Карло без привязки к политике.....	111
Итоги.....	112
Вопросы	113
Дополнительные источники.....	113
Глава 5. Обучение на основе временных различий	114
Обучение на основе временных различий	114
Прогнозирование на основе временных различий	115
TD-управление	118
Q-обучение	119
SARSA	127
Решение задачи о такси методом SARSA	131
Различия между Q-обучением и SARSA.....	133
Итоги.....	135
Вопросы	135
Дополнительные источники.....	135
Глава 6. Задача о многоруком бандите.....	136
Задача MAB	137
Эпсилон-жадная стратегия	139
Алгоритм softmax-исследования	140
Алгоритм верхней границы доверительного интервала	141
Алгоритм выборки Томпсона	145
Практические применения MAB	147
Выбор подходящего рекламного баннера с использованием MAB	148
Контекстные бандиты.....	151
Итоги.....	151
Вопросы	152
Дополнительные источники.....	152

Глава 7. Основы глубокого обучения.....	153
Искусственные нейроны	154
ANN	155
Входной слой	156
Скрытый слой.....	157
Выходной слой	157
Функции активации	157
Подробнее об ANN.....	159
Градиентный спуск.....	162
Нейросети в TensorFlow	168
RNN	171
Обратное распространение во времени.....	174
RNN с долгой краткосрочной памятью.....	176
Генерирование текстов песен посредством LSTM RNN.....	178
Сверточные нейросети	182
Сверточный слой.....	182
Слой подвыборки	188
Полносвязный слой	188
Архитектура CNN.....	189
Классификация предметов одежды с использованием CNN.....	189
Итоги.....	196
Вопросы	196
Дополнительные источники.....	196
Глава 8. Игры Atari с использованием Deep Q Network	197
Что такое DQN?	197
Архитектура DQN.....	199
Сверточная сеть	199
Воспроизведение опыта	200
Целевая сеть.....	201
Нормализация наград.....	202
Понимание алгоритма	202
Построение агента для игр Atari	203
Двойная сеть DQN	211

Приоритетное воспроизведение опыта	212
Архитектура дуэльных сетей.....	213
Итоги.....	215
Вопросы	215
Дополнительные источники.....	215
Глава 9. Игра Doom в глубокой рекуррентной Q-сети.....	216
DRQN.....	216
Архитектура DRQN	218
Обучение агента для игры в Doom	219
Базовая игра Doom	220
Doom с DRQN	222
DARQN.....	232
Архитектура DARQN	232
Итоги.....	233
Вопросы	234
Дополнительные источники.....	234
Глава 10. Асинхронная преимущественная сеть «актор-критик»	235
Асинхронный преимущественный алгоритм «актор-критик»	236
Три «А».....	236
Архитектура АЗС	237
Как работает АЗС	238
Подъем на гору с использованием АЗС	239
Визуализация в TensorBoard	247
Итоги.....	250
Вопросы	250
Дополнительные источники.....	250
Глава 11. Градиенты политик и оптимизация	251
Градиент политики	252
Посадка на Луну с градиентами политик	252
Глубокий детерминированный градиент политики.....	257
Раскачивание маятника.....	259

Оптимизация политики доверительной области	266
Оптимизация ближайшей политики	270
Итоги.....	272
Вопросы	273
Дополнительные источники.....	273
Глава 12. «Автогонки» с использованием DQN.....	274
Функции-обертки среды	274
Дуэльная сеть.....	278
Память воспроизведения.....	280
Обучение сети.....	281
«Автогонки»	287
Итоги.....	290
Вопросы	291
Дополнительные источники.....	291
Глава 13. Последние достижения и следующие шаги	292
Агенты, дополненные воображением.....	292
Обучение на человеческих предпочтениях.....	297
Глубокое Q-обучение на примере демонстраций	298
Ретроспективное воспроизведение опыта.....	299
Иерархическое обучение с подкреплением.....	301
Декомпозиция функции ценности MAXQ.....	302
Инвертированное обучение с подкреплением.....	305
Итоги.....	306
Вопросы	307
Дополнительные источники.....	307
Ответы	308