

УДК 004.85
ББК 32.971.3
Ш78

Шолле Ф.

Ш78 Глубокое обучение с R и Keras / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2023. – 646 с.: ил.

ISBN 978-5-93700-189-4

Прочитав эту книгу, вы получите четкое представление о том, что такое глубокое обучение, когда его следует применять и каковы его ограничения. Авторы описывают стандартный рабочий процесс поиска решения задачи машинного обучения и рассказывают, как устранять часто возникающие проблемы. Всесторонне рассматривается использование Keras для решения самых разнообразных прикладных задач, в числе которых классификация и сегментация изображений, прогнозирование временных рядов, классификация текста, машинный перевод, генерация текста и многое другое.

Издание адресовано читателям со средними навыками программирования на R. Опыт работы с Keras, TensorFlow или моделями глубокого обучения не требуется.

УДК 004.85
ББК 32.971.3

Copyright © DMC Press 2022. Authorized translation of the English edition © 2022 Manning Publications. This translation is published and sold by permission of Manning Publications, the owner of all rights to publish and sell the same.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-1-6334-3979-5 (англ.)
ISBN 978-5-93700-189-4 (рус.)

© Manning Publications, 2022
© Перевод, оформление, издание, ДМК Пресс, 2022

Содержание

Оглавление	6
Предисловие	15
Благодарности	17
Об этой книге	18
Об иллюстрации на обложке	22
Об авторах	23

1	Что такое глубокое обучение?	24
1.1	Искусственный интеллект, машинное и глубокое обучение	25
1.1.1	Искусственный интеллект	25
1.1.2	Машинное обучение	26
1.1.3	Извлечение правил и представлений из данных	28
1.1.4	«Глубина» глубокого обучения	31
1.1.5	Принцип действия глубокого обучения в трех рисунках	33
1.1.6	Каких успехов достигло глубокое обучение	35
1.1.7	Не верьте рекламной шумихе	36
1.1.8	Перспективы развития ИИ	37
1.2	Краткая история машинного обучения	38
1.2.1	Вероятностное моделирование	39
1.2.2	Первые нейронные сети	39
1.2.3	Ядерные методы	40
1.2.4	Деревья решений, случайные леса и градиентный бустинг	42
1.2.5	Назад к нейронным сетям	43
1.2.6	Отличительные черты глубокого обучения	44
1.2.7	Современный ландшафт машинного обучения	45
1.3	Почему глубокое обучение? Почему сейчас?	47
1.3.1	Оборудование	48
1.3.2	Данные	49
1.3.3	Алгоритмы	50
1.3.4	Новая волна инвестиций	51
1.3.5	Демократизация глубокого обучения	52
1.3.6	Ждать ли продолжения этой тенденции?	52

2	Математические основы нейронных сетей	55
2.1	Первое знакомство с нейронной сетью	56
2.2	Представление данных для нейронных сетей	60
2.2.1	Скаляры (тензоры нулевого ранга)	61
2.2.2	Векторы (тензоры первого ранга)	61
2.2.3	Матрицы (тензоры второго ранга)	62
2.2.4	Тензоры третьего и более высокого рангов	62
2.2.5	Ключевые атрибуты	62
2.2.6	Манипулирование тензорами в R	64
2.2.7	Пакеты данных	64
2.2.8	Практические примеры тензоров с данными	65
2.2.9	Векторные данные	65
2.2.10	Временные ряды, или последовательности данных	66
2.2.11	Изображения	67
2.2.12	Видеоданные	67
2.3	Шестеренки нейронных сетей: операции с тензорами	68
2.3.1	Поэлементные операции	69
2.3.2	Операции с тензорами разной размерности	70
2.3.3	Скалярное произведение тензоров	72
2.3.4	Изменение формы тензора	74
2.3.5	Геометрическая интерпретация операций с тензорами	75
2.3.6	Геометрическая интерпретация глубокого обучения	79
2.4	Механизм нейронных сетей: оптимизация на основе градиента	80
2.4.1	Что такое производная?	82
2.4.2	Производная операций с тензорами: градиент	83
2.4.3	Стохастический градиентный спуск	85
2.4.4	Объединение производных: алгоритм обратного распространения ошибки	88
2.5	Возвращаясь к нашему первому примеру	95
2.5.1	Повторная реализация нашего первого примера с нуля в TensorFlow	97
2.5.2	Выполнение одного шага обучения	99
2.5.3	Полный цикл обучения	101
2.5.4	Оценка модели	102
	Краткие итоги главы	103
3	Введение в Keras и TensorFlow	104
3.1	Что такое TensorFlow?	105
3.2	Что такое Keras?	106
3.3	Keras и TensorFlow: краткая история	107
3.4	Интерфейсы Python и R: краткая история	108
3.5	Настройка среды разработки для глубокого обучения	109
3.5.1	Установка Keras и TensorFlow	110
3.6	Первые шаги с TensorFlow	111
3.6.1	Тензоры TensorFlow	112
3.7	Атрибуты тензоров	113

3.7.1	Форма тензора и ее изменение	114
3.7.2	Срезы тензоров	116
3.7.3	Операции с тензорами разной размерности	117
3.7.4	Модуль <i>tf</i>	117
3.7.5	Неизменность тензоров и переменные	119
3.7.6	Математические операции в <i>TensorFlow</i>	120
3.7.7	Взгляд на <i>API GradientTape</i> с другой стороны	121
3.7.8	Полный пример: линейный классификатор в чистом <i>TensorFlow</i>	122
3.8	Анатомия нейронной сети и основы <i>API Keras</i>	127
3.8.1	Слои: строительные блоки глубокого обучения	128
3.8.2	От слоев к моделям	132
3.8.3	Этап «компиляции»: настройка процесса обучения	134
3.8.4	Выбор функции потерь	137
3.8.5	Использование метода <i>fit()</i>	138
3.8.6	Отслеживание потерь и показателей на контрольных данных	139
3.8.7	Использование модели после обучения	140
	Краткие итоги главы	141

4 Примеры работы с нейросетью: классификация и регрессия 143

4.1	Классификация отзывов к фильмам: пример бинарной классификации	145
4.1.1	Набор данных <i>IMDB</i>	145
4.1.2	Подготовка данных	147
4.1.3	Создание модели	148
4.1.4	Проверка вашего выбора	151
4.1.5	Использование обученной сети для прогнозирования на новых данных	154
4.1.6	Продолжаем эксперименты	155
4.1.7	Промежуточные итоги	155
4.2	Классификация новостных лент: пример многоклассовой классификации	156
4.2.1	Набор данных <i>Reuters</i>	156
4.2.2	Подготовка данных	158
4.2.3	Построение модели	158
4.2.4	Проверка модели	159
4.2.5	Предсказания на новых данных	161
4.2.6	Другой способ обработки меток и потерь	162
4.2.7	Важность использования достаточно больших промежуточных слоев	162
4.2.8	Дальнейшие эксперименты	163
4.2.9	Промежуточные итоги	163
4.3	Предсказание цен на дома: пример регрессии	164
4.3.1	Набор данных с ценами на жилье в Бостоне	164
4.3.2	Подготовка данных	165
4.3.3	Построение модели	165

4.3.4	Оценка качества модели методом K -кратной перекрестной проверки	166
4.3.5	Выдача прогнозов на новых данных	171
4.3.6	Промежуточные выводы	171
	Краткие итоги главы.....	171

5	Основы машинного обучения	173
5.1	Обобщение – цель машинного обучения	173
5.1.1	Недообучение и переобучение	174
5.1.2	Базовые принципы обобщения в глубоком обучении	180
5.2	Оценка моделей машинного обучения	187
5.2.1	Наборы данных для обучения, проверки и контроля.....	187
5.2.2	Использование критериев, основанных на здравом смысле.....	191
5.2.3	Что следует помнить об оценке модели	192
5.3	Улучшение качества обучения модели	193
5.3.1	Настройка ключевых параметров градиентного спуска.....	193
5.3.2	Использование лучшей априорно обоснованной архитектуры	196
5.3.3	Увеличение емкости модели.....	197
5.4	Как улучшить обобщение.....	199
5.4.1	Подготовка набора данных	199
5.4.2	Конструирование признаков	200
5.4.3	Использование ранней остановки	202
5.4.4	Регуляризация модели	202
	Краткие итоги главы.....	213

6	Обобщенный рабочий процесс машинного обучения	215
6.1	Постановка задачи.....	217
6.1.1	Уточнение задачи	217
6.1.2	Получение исходных данных.....	219
6.1.3	Добейтесь понимания данных.....	223
6.1.4	Выберите меру успеха.....	224
6.2	Разработка модели.....	225
6.2.1	Подготовка данных	225
6.2.2	Выбор протокола оценки	227
6.2.3	Как превзойти простой базовый уровень.....	228
6.2.4	Масштабирование: разработка модели, способной к переобучению	229
6.2.5	Регуляризация и настройка модели.....	230
6.3	Развертывание модели	231
6.3.1	Представление модели заказчику.....	231
6.3.2	Передача модели заказчику	232
6.3.3	Мониторинг модели в рабочей среде.....	236
6.3.4	Поддержка и обновление модели.....	236
	Краткие итоги главы.....	237

7	Работа с Keras: углубленные навыки	239
7.1	Широкий спектр рабочих процессов Keras	240
7.2	Различные способы построения моделей Keras	240
7.2.1	Sequential API	241
7.2.2	Functional API	244
7.2.3	Создание подкласса класса Model	251
7.2.4	Смешивание и сочетание разных компонентов	255
7.2.5	Используйте правильные инструменты	256
7.3	Использование встроенных циклов обучения и оценки	256
7.3.1	Разработка собственных метрик	257
7.3.2	Использование обратных вызовов	260
7.3.3	Разработка собственных обратных вызовов	262
7.3.4	Мониторинг и визуализация с помощью TensorBoard	264
7.4	Разработка собственных циклов обучения и оценки	266
7.4.1	Обучение или логический вывод	267
7.4.2	Использование метрик на низком уровне	268
7.4.3	Полный цикл обучения и оценки	269
7.4.4	Увеличьте быстродействие с помощью tf.function()	272
7.4.5	Использование fit() с пользовательским циклом обучения	273
	Краткие итоги главы	276
8	Глубокое обучение в компьютерном зрении	277
8.1	Введение в сверточные нейронные сети	278
8.1.1	Операция свертки	281
8.1.2	Выбор максимального значения из соседних (max-pooling)	286
8.2	Обучение сверточной нейронной сети с нуля на небольшом наборе данных	289
8.2.1	Целесообразность глубокого обучения для решения задач с небольшими наборами данных	290
8.2.2	Загрузка данных	290
8.2.3	Построение сети	293
8.2.4	Предварительная обработка данных	295
8.2.5	Расширение данных	301
8.3	Использование предварительно обученной сверточной нейронной сети	305
8.3.1	Выделение признаков	306
8.3.2	Дообучение ранее обученной модели	316
	Краткие итоги главы	320
9	Глубокое обучение для компьютерного зрения	321
9.1	Три основные задачи компьютерного зрения	322
9.2	Пример сегментации изображения	323
9.3	Современные стандартные архитектуры сверточных сетей	333
9.3.1	Модульность, иерархия и повторное использование	334
9.3.2	Остаточные связи	337

9.3.3	Пакетная нормализация	341
9.3.4	Разделяемые по глубине свертки	344
9.3.5	Применим знания на практике: мини-модель, подобная Xception	347
9.4	Интерпретация знаний сверточной нейросети	350
9.4.1	Визуализация промежуточных активаций	351
9.4.2	Визуализация сетевых фильтров	357
9.4.3	Визуализация тепловых карт активации класса	363
	Краткие итоги главы	369

10 Глубокое обучение и временные ряды.....370

10.1	Различные виды задач временных рядов.....	370
10.2	Пример прогнозирования температуры.....	372
10.2.1	Подготовка данных	376
10.2.2	Простое решение задачи без привлечения машинного обучения.....	380
10.2.3	Решение с использованием базовой модели машинного обучения.....	382
10.2.4	Эксперимент с одномерной сверточной сетью	384
10.2.5	Первый вариант простой рекуррентной модели	387
10.3	Рекуррентные нейронные сети	388
10.3.1	Рекуррентный слой в Keras.....	391
10.4	Продвинутое применение рекуррентных нейронных сетей	396
10.4.1	Использование рекуррентного прореживания для борьбы с переобучением.....	397
10.4.2	Наложение рекуррентных слоев	400
10.4.3	Использование двунаправленных рекуррентных сетей	402
10.4.4	Что дальше	405
	Краткие итоги главы.....	407

11 Глубокое обучение в обработке текстов.....408

11.1	Обработка естественного языка: обзор отрасли	408
11.2	Подготовка текстовых данных.....	411
11.2.1	Стандартизация текста	412
11.2.2	Разделение текста (токенизация).....	413
11.2.3	Индексация словаря.....	414
11.2.4	Использование слоя <code>layer_text_vectorization</code>	416
11.3	Два подхода к представлению групп слов: наборы и последовательности	420
11.3.1	Подготовка данных обзоров фильмов IMDB.....	421
11.3.2	Обработка слов без учета порядка	424
11.3.3	Обработка последовательности слов	432
11.4	Архитектура Transformer.....	446
11.4.1	Механизм самовнимания	446
11.4.2	Многоголовое внимание	452
11.4.3	Кодировщик в архитектуре Transformer	453
11.4.4	Когда следует использовать модели последовательности, а не модели мешка слов.....	463

11.5	Помимо классификации текста: обучение преобразованию последовательностей.....	464
11.5.1	Пример машинного перевода	466
11.5.3	Рекуррентная модель преобразования последовательностей	469
11.5.4	Преобразование последовательностей с Transformer	476
	Краткие итоги главы.....	482

12	Генеративные модели глубокого обучения	484
12.1	Генерирование текста с помощью Keras.....	486
12.1.1	Краткая история генеративных сетей.....	486
12.1.2	Как генерируют последовательности данных?	488
12.1.3	Важность стратегии выбора	488
12.1.4	Реализация генерации текста с помощью Keras.....	491
12.1.5	Обратный вызов генерации текста с выборкой при разной температуре	495
12.1.6	Подведение итогов	502
12.2	DeepDream	502
12.2.1	Реализация DeepDream в Keras.....	503
12.2.2	Подведение итогов	511
12.3	Нейронный перенос стиля.....	512
12.3.1	Функция потерь содержания.....	513
12.3.2	Функция потерь стиля	513
12.3.3	Реализация переноса стиля в Keras.....	514
12.3.4	Подведение итогов	522
12.4	Генерация изображений с помощью вариационных автокодировщиков	522
12.4.1	Выбор шаблонов из скрытых пространств изображений	523
12.4.2	Концептуальные векторы для редактирования изображений	524
12.4.3	Вариационные автокодировщики.....	525
12.4.4	Реализация VAE с помощью Keras	528
12.4.5	Подведение итогов	534
12.5	Введение в генеративно-состязательные сети	534
12.5.1	Реализация генеративно-состязательной сети	536
12.5.2	Полезные технические приемы	537
12.5.3	Получение набора данных CelebA.....	538
12.5.4	Дискриминатор.....	540
12.5.5	Генератор.....	541
12.5.6	Состязательная сеть	542
12.5.7	Подведение итогов	545
	Краткие итоги главы.....	546

13	Глубокое обучение в реальной жизни	547
13.1	Получение максимальной отдачи от ваших моделей	548
13.1.1	Оптимизация гиперпараметров	548
13.1.2	Ансамблирование моделей.....	557
13.2	Масштабируемое обучение моделей.....	559

13.2.1	Ускорение обучения на GPU со смешанной точностью	560
13.2.2	Обучение модели на нескольких GPU	563
13.2.3	Обучение модели на TPU	568
Краткие итоги главы		570

14	Заключение	571
14.1	Краткий обзор ключевых понятий	572
14.1.1	Различные подходы к ИИ	572
14.1.2	Что выделяет глубокое обучение среди других подходов к машинному обучению	573
14.1.3	Как правильно воспринимать глубокое обучение	573
14.1.4	Ключевые технологии глубокого обучения	575
14.1.5	Обобщенный рабочий процесс машинного обучения	576
14.1.6	Основные архитектуры сетей	577
14.1.7	Пространство возможностей	582
14.2	Ограничения глубокого обучения	584
14.2.1	Риск очеловечивания моделей глубокого обучения	585
14.2.2	Принципиальное различие между автоматом и интеллектом	587
14.2.3	Различие между локальным и экстремальным обобщением	589
14.2.4	Предназначение интеллекта	592
14.2.5	Восхождение по уровням обобщения	593
14.3	Курс на большую универсальность в ИИ	594
14.3.1	О важности постановки правильной цели: правило короткого пути	594
14.3.2	Новая цель	597
14.4	Реализация интеллекта: недостающие ингредиенты	599
14.4.1	Построение и использование абстрактных аналогий	599
14.4.2	Два полюса абстракции	601
14.4.3	Сочетание двух полюсов абстракции	604
14.4.4	Недостающая половина картинки	604
14.5	Будущее глубокого обучения	606
14.5.1	Модели как программы	606
14.5.2	Машинное обучение и синтез программ	608
14.5.3	Сочетание глубокого обучения и синтеза программ	608
14.5.4	Непрерывное обучение и повторное использование модульных подпрограмм	611
14.5.5	Долгосрочная перспектива	612
14.6	Как не отстать от прогресса в быстро развивающейся отрасли	614
14.6.1	Решения реальных задач на сайте Kaggle	614
14.6.2	Знакомство с последними разработками на сайте arXiv	614
14.6.3	Исследование экосистемы Keras	615
14.7	Заключительное слово	616
Приложение. Введение в Python для пользователей R		617
Предметный указатель		641