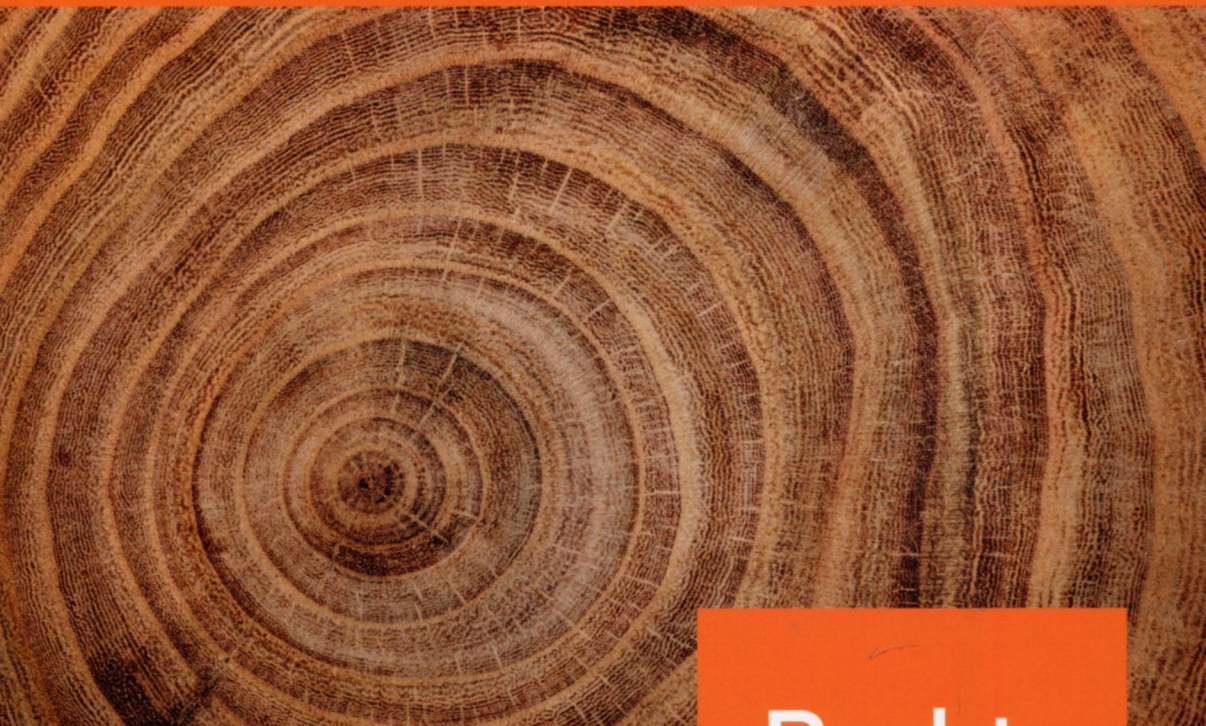


40 алгоритмов, которые должен знать каждый программист на Python

Имран Ахмад



Packt>

 ПИТЕР®

40 алгоритмов, которые должен знать каждый программист на Python

Имран Ахмад



Санкт-Петербург · Москва · Минск

2023

ББК 32.973.2-018.1
УДК 004.421+004.43
А95

Ахмад Имран

А95 40 алгоритмов, которые должен знать каждый программист на Python. — СПб.: Питер, 2023. — 368 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-4461-1908-0

Понимание работы алгоритмов и умение применять их для решения прикладных задач — must-have для любого программиста или разработчика. Эта книга поможет вам не только развить навыки использования алгоритмов, но и разобраться в принципах их функционирования, в их логике и математике.

Вы начнете с введения в алгоритмы, от поиска и сортировки перейдете к линейному программированию, ранжированию страниц и графам и даже поработаете с алгоритмами машинного обучения. Теории не бывает без практики, поэтому вы займетесь прогнозами погоды, кластеризацией твитов, механизмами рекомендаций фильмов. И, наконец, освоите параллельную обработку, что даст вам возможность решать задачи, требующие большого объема вычислений.

Дойдя до конца, вы превратитесь в эксперта по решению реальных вычислительных задач с применением широкого спектра разнообразных алгоритмов.

16+ (В соответствии с Федеральным законом от 29 декабря 2010 г. № 436-ФЗ.)

ББК 32.973.2-018.1
УДК 004.421+004.43

Права на издание получены по соглашению с Packt Publishing. Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав.

Информация, содержащаяся в данной книге, получена из источников, рассматриваемых издательством как надежные. Тем не менее, имея в виду возможные человеческие или технические ошибки, издательство не может гарантировать абсолютную точность и полноту приводимых сведений и не несет ответственности за возможные ошибки, связанные с использованием книги. Издательство не несет ответственности за доступность материалов, ссылки на которые вы можете найти в этой книге. На момент подготовки книги к изданию все ссылки на интернет-ресурсы были действующими.

В книге возможны упоминания организаций, деятельность которых запрещена на территории Российской Федерации, таких как Meta Platforms Inc., Facebook, Instagram и др.

ISBN 978-1789801217 англ.

ISBN 978-5-4461-1908-0

© Packt Publishing 2020.

First published in the English language under the title '40 Algorithms Every Programmer Should Know — (9781789801217)'

© Перевод на русский язык ООО «Прогресс книга», 2022

© Издание на русском языке, оформление ООО «Прогресс книга», 2022

© Серия «Библиотека программиста», 2022

Краткое содержание

Об авторе	16
Предисловие	17

ЧАСТЬ I Основы и базовые алгоритмы

Глава 1. Обзор алгоритмов	24
Глава 2. Структуры данных, используемые в алгоритмах	50
Глава 3. Алгоритмы сортировки и поиска	74
Глава 4. Разработка алгоритмов	94
Глава 5. Графовые алгоритмы	120

ЧАСТЬ II Алгоритмы машинного обучения

Глава 6. Алгоритмы машинного обучения без учителя	150
Глава 7. Традиционные алгоритмы обучения с учителем	187
Глава 8. Алгоритмы нейронных сетей	233

Глава 9. Алгоритмы обработки естественного языка	262
Глава 10. Рекомендательные системы	278

ЧАСТЬ III

Расширенные возможности

Глава 11. Алгоритмы обработки данных	294
Глава 12. Криптография	307
Глава 13. Крупномасштабные алгоритмы	332
Глава 14. Практические рекомендации	347

Оглавление

Об авторе	16
Предисловие	17
Для кого эта книга	17
О чем эта книга	18
Что вам потребуется при чтении этой книги	21
Условные обозначения	21
От издательства	22

ЧАСТЬ I **Основы и базовые алгоритмы**

Глава 1. Обзор алгоритмов	24
Что такое алгоритм	25
Этапы алгоритма	25
Определение логики алгоритма	27
Псевдокод	27
Использование сниппетов	30
Создание плана выполнения	30
Введение в библиотеки Python	31
Библиотеки Python	32
Реализация Python с помощью Jupyter Notebook	34
Методы разработки алгоритмов	35
Параметры данных	36
Параметры вычислений	37

Анализ производительности	38
Анализ пространственной сложности	39
Анализ временной сложности	39
Оценка эффективности	40
Выбор алгоритма	41
«О-большое»	42
Проверка алгоритма	46
Точные, приближенные и рандомизированные алгоритмы	46
Объяснимость алгоритма	48
Резюме	49
Глава 2. Структуры данных, используемые в алгоритмах	50
Структуры данных в Python	51
Список	51
Кортеж	56
Словарь	57
Множество	59
DataFrame	61
Матрица	63
Абстрактные типы данных	64
Вектор	65
Стек	65
Очередь	68
Базовый принцип использования стеков и очередей	70
Дерево	70
Резюме	73
Глава 3. Алгоритмы сортировки и поиска	74
Алгоритмы сортировки	75
Обмен значений переменных в Python	75
Сортировка пузырьком	76
Сортировка вставками	78
Сортировка слиянием	80
Сортировка Шелла	82
Сортировка выбором	84

Алгоритмы поиска	86
Линейный поиск	87
Бинарный поиск	88
Интерполяционный поиск	89
Практическое применение	90
Резюме	93
Глава 4. Разработка алгоритмов	94
Знакомство с основными концепциями разработки алгоритма	95
Вопрос 1. Даст ли разработанный алгоритм ожидаемый результат?	96
Вопрос 2. Является ли данный алгоритм оптимальным способом получения результата?	96
Вопрос 3. Как алгоритм будет работать с большими наборами данных? ..	100
Понимание алгоритмических стратегий	100
Стратегия «разделяй и властвуй»	101
Стратегия динамического программирования	103
Жадные алгоритмы	104
Практическое применение — решение задачи коммивояжера	105
Использование стратегии полного перебора	107
Использование жадного алгоритма	110
Алгоритм PageRank	111
Постановка задачи	112
Реализация алгоритма PageRank	112
Знакомство с линейным программированием	115
Формулировка задачи линейного программирования	115
Практическое применение — планирование производства с помощью линейного программирования	116
Резюме	119
Глава 5. Графовые алгоритмы	120
Представление графов	121
Типы графов	122
Особые типы ребер	125
Эгоцентрические сети	126
Анализ социальных сетей	126

Введение в теорию сетевого анализа	128
Кратчайший путь	128
Создание окрестностей	129
Показатели центральности	130
Вычисление показателей центральности с помощью Python	132
Понятие обхода графа	133
BFS — поиск в ширину	135
DFS — поиск в глубину	137
Практический пример — выявление мошенничества	140
Простой анализ мошенничества	144
Анализ мошенничества методом сторожевой башни	144
Резюме	148

ЧАСТЬ II

Алгоритмы машинного обучения

Глава 6. Алгоритмы машинного обучения без учителя	150
Обучение без учителя	151
Обучение без учителя в жизненном цикле майнинга данных	151
Современные тенденции исследований в области обучения без учителя ..	154
Практические примеры	155
Алгоритмы кластеризации	156
Количественная оценка сходства	157
Иерархическая кластеризация	164
Оценка кластеров	166
Применение кластеризации	167
Снижение размерности	168
Метод главных компонент (PCA)	168
Ограничения PCA	171
Поиск ассоциативных правил	171
Примеры использования	172
Анализ рыночной корзины	172
Ассоциативные правила	174
Оценка качества правила	176
Алгоритмы анализа ассоциаций	177

Практический пример — объединение похожих твитов в кластеры	184
Тематическое моделирование	184
Кластеризация	185
Алгоритмы обнаружения выбросов (аномалий)	185
Использование кластеризации	186
Обнаружение аномалий на основе плотности	186
Метод опорных векторов	186
Резюме	186
Глава 7. Традиционные алгоритмы обучения с учителем	187
Машинное обучение с учителем	188
Терминология машинного обучения с учителем	189
Благоприятные условия	191
Различие между классификаторами и регрессорами	192
Алгоритмы классификации	192
Задача классификации	193
Оценка классификаторов	197
Этапы классификации	201
Алгоритм дерева решений	203
Ансамблевые методы	207
Логистическая регрессия	211
Метод опорных векторов (SVM)	214
Наивный байесовский алгоритм	216
Среди алгоритмов классификации победителем становится...	219
Алгоритмы регрессии	220
Задача регрессии	220
Линейная регрессия	223
Алгоритм дерева регрессии	228
Алгоритм градиентного бустинга для регрессии	229
Среди алгоритмов регрессии победителем становится...	230
Практический пример — как предсказать погоду	230
Резюме	232
Глава 8. Алгоритмы нейронных сетей	233
Введение в ИНС	234
Эволюция ИНС	236

Обучение нейронной сети	238
Анатомия нейронной сети	238
Градиентный спуск	239
Функции активации	242
Инструменты и фреймворки	247
Keras	248
Знакомство с TensorFlow	251
Типы нейронных сетей	254
Перенос обучения	256
Практический пример — использование глубокого обучения для выявления мошенничества	257
Методология	257
Резюме	261
Глава 9. Алгоритмы обработки естественного языка	262
Знакомство с <i>NLP</i>	263
Терминология <i>NLP</i>	263
Библиотека NLTK	266
Мешок слов (BoW)	266
Эмбединги слов	269
Окружение слова	270
Свойства эмбедингов слов	270
Рекуррентные нейросети в <i>NLP</i>	271
Использование <i>NLP</i> для анализа эмоциональной окраски текста	272
Практический пример — анализ тональности в отзывах на фильмы	274
Резюме	277
Глава 10. Рекомендательные системы	278
Введение в рекомендательные системы	279
Типы рекомендательных систем	279
Рекомендательные системы на основе контента	279
Рекомендательные системы на основе коллаборативной фильтрации	282
Гибридные рекомендательные системы	284
Ограничения рекомендательных систем	286
Проблема холодного старта	287
Требования к метаданным	287

Проблема разреженности данных	287
Предвзятость из-за социального влияния	287
Ограниченные данные	288
Области практического применения	288
Практический пример — создание рекомендательной системы	288
Резюме	291

ЧАСТЬ III

Расширенные возможности

Глава 11. Алгоритмы обработки данных	294
Знакомство с алгоритмами обработки данных	294
Классификация данных	295
Алгоритмы хранения данных	296
Стратегии хранения данных	296
Алгоритмы потоковой передачи данных	299
Применение потоковой передачи	299
Алгоритмы сжатия данных	300
Алгоритмы сжатия без потерь	300
Практический пример — анализ тональности твитов в режиме реального времени	303
Резюме	306
Глава 12. Криптография	307
Введение в криптографию	307
Понимание важности самого слабого звена	308
Основная терминология	309
Требования безопасности	309
Базовое устройство шифров	312
Типы криптографических методов	315
Криптографические хеш-функции	315
Симметричное шифрование	319
Асимметричное шифрование	321
Практический пример — проблемы безопасности при развертывании модели МО	325
Атака посредника (MITM)	326

Избегание маскардинга	328
Шифрование данных и моделей	328
Резюме	331
Глава 13. Крупномасштабные алгоритмы	332
Введение в крупномасштабные алгоритмы	333
Определение эффективного крупномасштабного алгоритма	333
Терминология	333
Разработка параллельных алгоритмов	334
Закон Амдала	334
Гранулярность задачи	337
Балансировка нагрузки	338
Проблема расположения	338
Запуск параллельной обработки на Python	339
Разработка стратегии мультипроцессорной обработки	339
Введение в CUDA	340
Кластерные вычисления	343
Гибридная стратегия	346
Резюме	346
Глава 14. Практические рекомендации	347
Введение в практические рекомендации	348
Печальная история ИИ-бота в Твиттере	348
Объяснимость алгоритма	349
Алгоритмы машинного обучения и объяснимость	350
Этика и алгоритмы	353
Проблемы обучающихся алгоритмов	354
Понимание этических аспектов	355
Снижение предвзятости в моделях	356
Решение NP-трудных задач	357
Упрощение задачи	358
Адаптация известного решения аналогичной задачи	358
Вероятностный метод	359
Когда следует использовать алгоритмы	359
Практический пример — события типа «черный лебедь»	360
Резюме	362