

УДК 004.891
ББК 32.972.13
Р21

Рамсундар Б., Истман П., Уолтерс П., Панде В.
Р21 Глубокое обучение в биологии и медицине / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2020. – 200 с.: ил.

ISBN 978-5-97060-791-6

Глубокое обучение добилось впечатляющих успехов во многих отраслях. Сейчас оно все глубже проникает в прикладные научные исследования, в частности биологию и смежные дисциплины. Эта книга рассказывает о применении глубокого обучения в геномике, химии, биофизике, микроскопии, медицине и других направлениях современных исследований всего, что связано с живыми организмами.

Издание будет полезно широкому кругу специалистов, связанных с анализом данных в химии, биологии и медицине, а также разработчикам ПО для них и студентам вузов.

УДК 004.891
ББК.32.972.13

Original English language edition published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. Copyright © 2019 Bharath Ramsundar, Peter Eastman, Patrick Walters, and Vijay Pande. Russian-language edition copyright © 2020 by DMK Press. All rights reserved.

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 978-1-492-03983-9 (анг.)

Copyright © 2019 Bharath Ramsundar,
Peter Eastman, Patrick Walters, and Vijay Pande

ISBN 978-5-97060-791-6 (рус.)

© Оформление, издание, перевод, ДМК Пресс, 2020

Содержание

Предисловие	8
Об авторах	11
Глава 1. Глубокое обучение и науки о жизни	13
Почему все говорят о глубоком обучении?	13
Современные науки о жизни – это науки о данных	14
О чем эта книга?	15
Глава 2. Введение в глубокое обучение	19
Линейные модели	20
Многослойные перцептроны	21
Обучение модели	24
Проверка модели	26
Регуляризация	27
Оптимизация гиперпараметров	28
Другие типы моделей	29
Сверточные нейронные сети	30
Рекуррентные нейронные сети	31
Дополнительное чтение	32
Глава 3. Машинное обучение с DeepChem	33
Наборы данных DeepChem	34
Обучение модели для предсказания токсичности молекул	35
Пример: обучение модели MNIST	42
Набор данных распознавания цифр MNIST	42
Сверточная архитектура для набора MNIST	43
Заключение	47
Глава 4. Машинное обучение и молекулы	48
Что такое молекула?	48
Что такое внутримолекулярные связи?	50
Ковалентные связи	51
Нековалентные связи	51
Молекулярные графы	52
Конформации молекулы	53
Хиральность молекул	54
Фичеризация молекулы	55
Строки SMILES и пакет RDKit	55
Расширенные отпечатки связей	56
Молекулярные дескрипторы	57
Графовые свертки	57

Обучение модели для прогнозирования растворимости	58
MoleculeNet	60
Строки SMARTS.....	60
Заключение	63
Глава 5. Глубокое обучение и биофизика	64
Белковые структуры	65
Белковые последовательности	67
Общие принципы связывания с белками.....	70
Биофизическая фичеризация	71
Координатная фичеризация	71
Атомная фичеризация	76
Пример использования PDBBind.....	76
PDBBind Dataset	76
Представление набора данных PDBBind.....	79
Заключение	82
Глава 6. Глубокое обучение и геномика	85
ДНК, РНК и белки.....	85
Реальное положение дел	87
Сайты связывания и факторы транскрипции	90
Сверточная модель связывания TF	90
Доступность хроматина	93
РНК-интерференция	95
Заключение	98
Глава 7. Машинное обучение и микроскопия	99
Краткое введение в микроскопию	101
Современная оптическая микроскопия	102
Дифракционный предел	104
Электронная и атомно-силовая микроскопия	105
Микроскопия сверхвысокого разрешения	107
Глубокое обучение и дифракционный предел	109
Подготовка биологических препаратов для микроскопии	109
Окрашивание	109
Фиксация препаратов	110
Секционирование препаратов	111
Флуоресцентная микроскопия	111
Артефакты пробоподготовки	113
Применение глубокого обучения в микроскопии.....	114
Подсчет клеток.....	114
Клеточная сегментация	117
Вычислительные анализы.....	121
Заключение	121
Глава 8. Глубокое обучение в медицине	123
Компьютерная диагностика.....	123
Вероятностные диагнозы с байесовскими сетями.....	124

Данные электронных медицинских карт	126
В чем опасность больших баз данных ЭМК пациентов?.....	128
Глубокая радиология	129
Рентгенография и компьютерная томография	131
Гистология.....	133
Магниторезонансная томография	133
Модель глубокого обучения в качестве лечебного средства	134
Диабетическая ретинопатия.....	135
Перспективы глубокого обучения в медицине	139
Этические соображения	139
Потеря работы	140
Заключение	140
Глава 9. Генеративные модели	141
Вариационные автоэнкодеры.....	141
Генеративные состязательные сети	143
Применение генеративных моделей в науках о жизни.....	144
Генерация новых идей для соединений-прототипов	144
Разработка белков	145
Инструменты для научного поиска.....	145
Будущее генеративного моделирования	146
Работа с генеративными моделями	146
Анализ вывода генеративной модели	148
Заклучение	151
Глава 10. Интерпретация глубоких моделей.....	154
Как объяснить предсказания?	154
Оптимизация входов.....	158
Прогнозирование неопределенности	161
Интерпретируемость, объяснимость и последствия для реального мира	165
Заклучение	165
Глава 11. Практический пример виртуального скрининга.....	166
Подготовка набора данных для прогнозного моделирования.....	167
Обучение прогностической модели	172
Подготовка набора данных для прогнозирования.....	177
Применение прогностической модели	180
Заклучение	186
Глава 12. Ожидания и перспективы	188
Медицинская диагностика.....	188
Персонализированная медицина.....	190
Фармацевтические исследования	191
Биологические исследования	193
Заклучение	194
Колофон.....	195
Предметный указатель.....	196