

Оглавление

Часть I ■ ПЕРВЫЕ ШАГИ	26
1 ■ Введение в машинное обучение с участием человека	27
2 ■ Начало работы с машинным обучением с участием человека (human-in-the-loop).....	52
Часть II ■ АКТИВНОЕ ОБУЧЕНИЕ	82
3 ■ Выборка неопределенности	84
4 ■ Выборка разнообразия	124
5 ■ Расширенное активное обучение	173
6 ■ Активное обучение для решения различных задач машинного обучения.....	208
Часть III ■ АННОТИРОВАНИЕ.....	250
7 ■ Работа с людьми, аннотирующими ваши данные	252
8 ■ Контроль качества при аннотировании данных	285
9 ■ Углубленное аннотирование и дополнение данных	325
10 ■ Качественные аннотации для различных задач машинного обучения.....	373
Часть VI ■ ВЗАИМОДЕЙСТВИЕ ЧЕЛОВЕКА И КОМПЬЮТЕРА ПРИ МАШИННОМ ОБУЧЕНИИ	415
11 ■ Интерфейсы для аннотирования данных	417
12 ■ Продукты машинного обучения с участием человека.....	453

Содержание

Предисловие	16
Введение	18
Благодарности.....	19
Об этой книге	21
Об авторе	25

Часть I ПЕРВЫЕ ШАГИ

26

1 Введение в машинное обучение с участием человека.....

27

1.1	Базовые принципы машинного обучения с участием человека	28
1.2	Введение в аннотирование.....	30
1.2.1	Простые и более сложные стратегии аннотирования	30
1.2.2	Устранение пробелов в области научных знаний о данных	30
1.2.3	Качество аннотирования человеком: почему это трудно?	31
1.3	Введение в активное обучение: повышение скорости и снижение стоимости обучающих данных.....	33
1.3.1	Три широкие стратегии отбора активного обучения: неопределенность, разнообразие и случайность.....	33
1.3.2	Что такое случайный выбор оценочных данных?.....	37
1.3.3	Когда использовать активное обучение?.....	38
1.4	Машинное обучение и взаимодействие человек–компьютер	40
1.4.1	Пользовательские интерфейсы: как вы создаете обучающие данные?	40
1.4.2	Прайминг: что может повлиять на человеческое восприятие?.....	42
1.4.3	Плюсы и минусы создания меток путем оценки прогнозов машинного обучения	43
1.4.4	Основные принципы проектирования интерфейсов аннотации	43
1.5	Машинное обучение в помощь человеку или машинное обучение с участием человека	43
1.6	Перенос обучения для запуска ваших моделей	44
1.6.1	Перенос обучения в компьютерном зрении	46
1.6.2	Перенос обучения при обработке естественного языка	46
1.7	Чего ожидать от этого текста	49
	Резюме.....	50

2	Начало работы с машинным обучением с участием человека (human-in-the-loop)	52
2.1	За пределами активного обучения: ваш первый алгоритм активного обучения	53
2.2	Архитектура вашей первой системы	55
2.3	Интерпретация прогнозов модели и данных для активного обучения.....	59
2.3.1	Ранжирование достоверности	60
2.3.2	Выявление выбросов	61
2.3.3	Чего можно ожидать в процессе итераций.....	64
2.4	Построение интерфейса для сбора меток человека	66
2.4.1	Простой интерфейс для маркировки текста.....	66
2.4.2	Управление данными машинного обучения	69
2.5	Развертывание вашей первой системы машинного обучения с участием человека	69
2.5.1	Всегда в первую очередь собирайте данные для оценки	72
2.5.2	Каждая точка данных получает шанс	75
2.5.3	Выбор правильных стратегий для ваших данных.....	76
2.5.4	Переобучение модели и итерации.....	79
	Резюме.....	80

Часть II АКТИВНОЕ ОБУЧЕНИЕ..... 82

3	Выборка неопределенности	84
3.1	Интерпретация неопределенности в модели машинного обучения....	85
3.1.1	Для чего искать неопределенность в вашей модели?	86
3.1.2	Softmax и распределения вероятностей	88
3.1.3	Интерпретация успешности активного обучения	90
3.2	Алгоритмы для выборки неопределенности.....	90
3.2.1	Выборка с наименьшим доверием	92
3.2.2	Выборка по пределу уверенности	94
3.2.3	Соотношение выборок	95
3.2.4	Энтропия (энтропия классификации)	97
3.2.5	Глубокое погружение в энтропию	100
3.3	Определение случаев запутанности различных типов моделей	101
3.3.1	Выборка неопределенности с помощью логистической регрессии и моделей MaxEnt	101
3.3.2	Выборка неопределенности с помощью метода опорных векторов (SVM)	103
3.3.3	Выборка неопределенности с помощью байесовских моделей	104
3.3.4	Выборка неопределенности с помощью деревьев решений и случайных лесов.....	105
3.4	Измерение неопределенности по нескольким прогнозам	106
3.4.1	Выборка неопределенности с помощью ансамбля моделей	106
3.4.2	Запрос по комитету и отсеивание.....	108
3.4.3	Разница между алеаторной и эпистемической неопределенностями	110
3.4.4	Классификация с несколькими метками и непрерывными значениями	111
3.5	Определение правильного числа элементов для проверки человеком	112

3.5.1	Выборка неопределенности с ограниченным бюджетом	113
3.5.2	Выборка неопределенности с временными ограничениями	114
3.5.3	Когда остановиться, если нет ограничений по времени или бюджету?	115
3.6	Оценка успешности активного обучения	115
3.6.1	Нужны ли мне новые тестовые данные?	115
3.6.2	Нужны ли мне новые данные для проверки?	116
3.7	Памятка по выборке неопределенности	118
3.8	Дополнительная литература	120
3.8.1	Дополнительная литература по наименее достоверной выборке	121
3.8.2	Дополнительная литература по выборке с пределом достоверности	121
3.8.3	Дополнительная литература по доверительной выборке	121
3.8.4	Дополнительная литература по выборке на основе энтропии	121
3.8.5	Дополнительная литература по другим моделям машинного обучения	122
3.8.6	Дополнительная литература по выборке неопределенности на основе ансамблей	122
	Резюме	123

4	Выборка разнообразия	124
4.1	Осознание того, чего вы не знаете: выявление пробелов в знаниях вашей модели	126
4.1.1	Пример данных для выборки разнообразия	129
4.1.2	Интерпретация нейронных моделей для выборки разнообразия	130
4.1.3	Получение информации из скрытых слоев в PyTorch	132
4.2	Выборка выбросов на основе модели	135
4.2.1	Использование данных проверки для ранжирования активаций	136
4.2.2	Какие слои следует использовать для расчета выбросов модели?	140
4.2.3	Ограничения выбросов на данных моделей	141
4.3	Кластерная выборка	142
4.3.1	Состав кластера, центроиды и выбросы	143
4.3.2	Любой из существующих во вселенной алгоритмов кластеризации	144
4.3.3	Кластеризация k -средних с косинусным сходством	146
4.3.4	Уменьшение размерности параметров с помощью вложений или анализа главных компонент	149
4.3.5	Другие алгоритмы кластеризации	151
4.4	Репрезентативная выборка	153
4.4.1	Репрезентативная выборка нечасто используется обособленно	154
4.4.2	Простая репрезентативная выборка	156
4.4.3	Адаптивная репрезентативная выборка	157
4.5	Выборка для получения реального разнообразия	159
4.5.1	Распространенные проблемы разнообразия обучающих данных	160
4.5.2	Стратифицированная выборка для обеспечения разнообразия демографических данных	162
4.5.3	Представленный и представляющий: что важно?	163
4.5.4	Демографическая точность	164
4.5.5	Ограничения выборки для определения реального разнообразия	165
4.6	Выборка разнообразия с различными типами моделей	166

4.6.1	Выбросы на основе различных типов моделей.....	166
4.6.2	Кластеризация с использованием различных типов моделей	166
4.6.3	Репрезентативная выборка с различными типами моделей	167
4.6.4	Выборка для реального разнообразия с различными типами моделей	167
4.7	Краткая памятка по выборке разнообразия.....	167
4.8	Дополнительная литература.....	169
4.8.1	Дополнительная литература по выбросам на основе моделей	169
4.8.2	Дополнительная литература по кластерной выборке	169
4.8.3	Дополнительная литература по репрезентативной выборке.....	170
4.8.4	Дополнительная литература по выборке для реального разнообразия	170
Резюме	171

5	Расширенное активное обучение	173
5.1	Сочетание выборки неопределенности и выборки разнообразия	173
5.1.1	Выборка наименьшего доверия с выборкой на основе кластеров ...	174
5.1.2	Выборка неопределенности с выбросами по модели.....	177
5.1.3	Выборка неопределенности с выбросами по модели и кластеризацией	179
5.1.4	Репрезентативная выборка на основе кластерной выборки.....	179
5.1.5	Выборка из кластера с наибольшей энтропией	182
5.1.6	Другие комбинации стратегий активного обучения.....	185
5.1.7	Сочетание результатов активного обучения	186
5.1.8	Выборка для уменьшения предполагаемой ошибки	187
5.2	Активный перенос обучения для выборки неопределенности	189
5.2.1	Учим модель предсказывать собственные ошибки	190
5.2.2	Применение активного переноса обучения	191
5.2.3	Активный перенос обучения с большим количеством слоев	194
5.2.4	Плюсы и минусы активного переноса обучения.....	195
5.3	Применение активного переноса обучения к репрезентативной выборке	196
5.3.1	Использование модели для предсказания неизвестного	196
5.3.2	Активный перенос обучения для адаптивной репрезентативной выборки.....	198
5.3.3	Плюсы и минусы активного переноса обучения для репрезентативной выборки.....	199
5.4	Активный перенос обучения для адаптивной выборки.....	200
5.4.1	Адаптация выборки неопределенности посредством прогнозирования неопределенности.....	200
5.4.2	Плюсы и минусы метода ATLAS	203
5.5	Краткие памятки по расширенному активному обучению	204
5.6	Дополнительная литература по активному переносу обучения	206
Резюме	207

6	Активное обучение для решения различных задач машинного обучения	208
6.1	Использование активного обучения для обнаружения объектов.....	209
6.1.1	Точность выявления объектов: достоверность меток и локализация	211

6.1.2	Выборка неопределенности для оценки достоверности меток и локализации при выявлении объектов	213
6.1.3	Выборка разнообразия для достоверности меток и локализации при выявлении объектов	215
6.1.4	Активный перенос обучения для распознавания объектов	219
6.1.5	Низкий порог распознавания объектов во избежание закрепления необъективности	219
6.1.6	Создание образцов обучающих данных для репрезентативной выборки, схожих с прогнозами	221
6.1.7	Выборка разнообразия по изображениям при распознавании объектов	222
6.1.8	Создание более точных масок при использовании многоугольников	223
6.2	Использование активного обучения для семантической сегментации	224
6.2.1	Точность семантической сегментации	225
6.2.2	Выборка неопределенности для семантической сегментации	227
6.2.3	Выборка разнообразия для семантической сегментации	228
6.2.4	Активный перенос обучения для семантической сегментации	229
6.2.5	Выборка разнообразия по изображениям для семантической сегментации	229
6.3	Применение активного обучения для маркировки последовательностей	230
6.3.1	Точность маркировки последовательностей	231
6.3.2	Выборка неопределенности для маркировки последовательностей	232
6.3.3	Выборка разнообразия для маркировки последовательностей	233
6.3.4	Активный перенос обучения для маркировки последовательностей	236
6.3.5	Стратифицированная выборка по достоверности и токенам	237
6.3.6	Создание образцов обучающих данных для репрезентативной выборки, похожих на ваши прогнозы	237
6.3.7	Маркировка всей последовательности	237
6.3.8	Выборка разнообразия по документу при маркировке последовательностей	238
6.4	Применение активного обучения для генерации языка	238
6.4.1	Вычисление точности для систем генерации языка	239
6.4.2	Выборка неопределенности для генерации языка	240
6.4.3	Выборка разнообразия для генерации языка	241
6.4.4	Активный перенос обучения для генерации языка	242
6.5	Применение активного обучения к другим задачам машинного обучения	242
6.5.1	Активное обучение для поиска информации	243
6.5.2	Активное обучение для видео	245
6.5.3	Активное обучение для речи	246
6.6	Выбор подходящего количества элементов для проверки человеком	247
6.6.1	Активная разметка полностью или частично аннотированных данных	247
6.6.2	Совмещение машинного обучения с аннотированием	248
6.7	Дополнительная литература	248
	Резюме	249

Часть III АННОТИРОВАНИЕ.....250

7	Работа с людьми, аннотирующими ваши данные	252
7.1	Введение в аннотирование	254
7.1.1	Три правила хорошего аннотирования данных	255
7.1.2	Аннотирование данных и проверка прогнозов модели	256
7.1.3	Аннотации человека, полученные в процессе машинного обучения	256
7.2	Штатные эксперты	257
7.2.1	Зарботная плата для штатных сотрудников.....	258
7.2.2	Защищенность штатных сотрудников.....	259
7.2.3	Вовлеченность штатных сотрудников	259
7.2.4	Совет: всегда проводите сеансы аннотирования своими силами.....	261
7.3	Сотрудники на аутсорсинге	263
7.3.1	Зарплата для аутсорсинговых работников.....	264
7.3.2	Защищенность аутсорсинговых работников	266
7.3.3	Вовлеченность аутсорсинговых работников	266
7.3.4	Совет: общайтесь с вашими аутсорсинговыми сотрудниками	267
7.4	Краудсорсинговые работники	268
7.4.1	Зарплата для сотрудников краудсорсинга	270
7.4.2	Защищенность краудсорсинговых работников	271
7.4.3	Вовлеченность краудсорсинговых работников	272
7.4.4	Совет: создайте условия для стабильной работы и карьерного роста	273
7.5	Другие виды рабочей силы	273
7.5.1	Конечные пользователи	274
7.5.2	Волонтеры.....	275
7.5.3	Любители игр	277
7.5.4	Прогноз модели в качестве аннотации	278
7.6	Оценка требуемого объема аннотирования.....	280
7.6.1	Уравнение порядка количества необходимых аннотаций	280
7.6.2	От одной до четырех недель на обучение аннотированию и уточнение заданий	282
7.6.3	Для оценки затрат используйте пилотные аннотации и показатели точности.....	283
7.6.4	Сочетание разных типов трудовых ресурсов	283
	Резюме.....	284
8	Контроль качества при аннотировании данных	285
8.1	Сравнение аннотаций с истинными значениями ответов	286
8.1.1	Согласие аннотатора с базовыми истинными данными	289
8.1.2	Какой базовый уровень использовать для ожидаемой точности?.....	292
8.2	Межаннотаторское согласие.....	293
8.2.1	Введение в межаннотаторское согласие	294
8.2.2	Преимущества вычисления межаннотаторского согласия	296
8.2.3	Согласие по набору данных с помощью альфы Криппендорфа.....	299
8.2.4	Для чего, помимо маркировки, применима альфа Криппендорфа	303
8.2.5	Индивидуальное согласие аннотаторов	304

8.2.6	Согласие по каждой метке и каждому демографическому показателю	308
8.2.7	Повышение точности с помощью согласия для реального разнообразия	309
8.3	Агрегирование аннотаций для создания обучающих данных	309
8.3.1	Агрегирование аннотаций при общем согласии	310
8.3.2	Математический расчет для несогласных аннотаторов и низкого уровня согласия	311
8.3.3	Агрегирование аннотаций при несогласии аннотаторов	312
8.3.4	Достоверность с подачи аннотатора	314
8.3.5	Решаем, каким меткам доверять: неопределенность аннотации	315
8.4	Контроль качества посредством экспертной оценки	318
8.4.1	Набор и обучение квалифицированных сотрудников	319
8.4.2	Обучение персонала до уровня экспертов	320
8.4.3	Экспертиза с помощью машинного обучения	320
8.5	Многоэтапные рабочие процессы и задачи рецензирования	321
8.6	Дополнительная литература	323
	Резюме	324

9 Углубленное аннотирование и дополнение данных

9.1	Качественное аннотирование для субъективных задач	326
9.1.1	Выяснение предположений аннотаторов	329
9.1.2	Определение приемлемых меток для субъективных задач	330
9.1.3	Доверие к аннотатору для анализа разнообразия ответов	332
9.1.4	Байесовская сыворотка правды для субъективных суждений	334
9.1.5	Встраивание простых задач в более сложные	336
9.2	Машинное обучение для контроля качества аннотаций	337
9.2.1	Расчет достоверности аннотации как задачи оптимизации	338
9.2.2	Согласование достоверности меток при разногласиях аннотаторов	339
9.2.3	Прогнозирование достоверности отдельной аннотации	342
9.2.4	Прогнозирование согласованности для отдельной аннотации	344
9.2.5	Определение аннотатора как бота	344
9.3	Предсказание модели в качестве аннотаций	345
9.3.1	Доверие к аннотациям на основе достоверных предсказаний модели	346
9.3.2	Использование прогнозов модели в качестве единого аннотатора	349
9.3.3	Перекрестная валидация для поиска ошибочно маркированных данных	350
9.4	Вложения и контекстуальные отображения	350
9.4.1	Обучение переноса из существующей модели	353
9.4.2	Представления из смежных легко аннотируемых задач	354
9.4.3	Метод самоконтроля: использование меток, присущих данным	355
9.5	Системы на основе поиска и системы на основе правил	357
9.5.1	Фильтрация данных с помощью правил	358
9.5.2	Поиск обучающих данных	359
9.5.3	Маскированная фильтрация характеристик	359
9.6	Легкий надзор над неконтролируемыми моделями	360

9.6.1	Адаптация неконтролируемой модели к контролируемой модели	360
9.6.2	Исследовательский анализ данных под контролем человека	362
9.7	Синтетические данные, создание данных и их дополнение	362
9.7.1	Синтетические данные	362
9.7.2	Создание данных	363
9.7.3	Дополнение данных	365
9.8	Внедрение информации об аннотациях в модели машинного обучения	365
9.8.1	Фильтрация, или взвешивание элементов по доверию к их меткам	366
9.8.2	Включение идентификации аннотатора во входные данные	366
9.8.3	Внедрение неопределенности в функцию потерь	367
9.9	Дополнительная литература по расширенному аннотированию	368
9.9.1	Дополнительная литература по субъективным данным	368
9.9.2	Дополнительная литература по машинному обучению для контроля качества аннотаций	368
9.9.3	Дополнительная литература по вложениям / контекстным представлениям	369
9.9.4	Дополнительная литература по системам на основе правил	370
9.9.5	Дополнительная литература по включению неопределенности аннотаций в последующие модели	370
	Резюме	371

10 Качественные аннотации для различных задач машинного обучения

10.1	Качество аннотаций для непрерывных задач	374
10.1.1	Базовая истина для непрерывных задач	374
10.1.2	Соглашение для непрерывных задач	375
10.1.3	Субъективность в непрерывных задачах	376
10.1.4	Агрегирование непрерывных оценок для создания обучающих данных	377
10.1.5	Машинное обучение для агрегирования непрерывных задач с целью создания обучающих данных	379
10.2	Качество аннотаций для задач распознавания объектов	381
10.2.1	Базовая истина для распознавания объектов	382
10.2.2	Согласие при распознавании объектов	384
10.2.3	Размерность и точность при распознавании объектов	385
10.2.4	Субъективность при распознавании объектов	386
10.2.5	Агрегирование аннотаций объектов для создания обучающих данных	386
10.2.6	Машинное обучение для аннотаций объектов	388
10.3	Качество аннотаций для семантической сегментации	389
10.3.1	Базовая истина для аннотации семантической сегментации	390
10.3.2	Соглашение для семантической сегментации	391
10.3.3	Субъективность аннотаций семантической сегментации	391
10.3.4	Агрегирование семантической сегментации для создания обучающих данных	392
10.3.5	Машинное обучение для агрегирования задач семантической сегментации при создании обучающих данных	393
10.4	Качество аннотации для маркировки последовательности	394

10.4.1	Базовая истина для маркировки последовательности	396
10.4.2	Базовая истина для маркировки последовательностей в реально непрерывных данных.....	397
10.4.3	Согласие по маркировке последовательностей.....	398
10.4.4	Машинное обучение и перенос обучения для маркировки последовательностей	398
10.4.5	Данные на основе правил, поиска и синтетических данных для маркировки последовательностей	401
10.5	Качество аннотаций для генерирования языковых материалов.....	401
10.5.1	Базовая истина для генерации языка	402
10.5.2	Согласие и агрегирование для генерации языка	403
10.5.3	Машинное обучение и обучение переноса для генерации языка	403
10.5.4	Синтетические данные для генерации языка.....	404
10.6	Качественное аннотирование для других задач машинного обучения.....	405
10.6.1	Аннотирование для поиска информации	405
10.6.2	Аннотирование для многоплановых задач.....	408
10.6.3	Аннотирование для видео	409
10.6.4	Аннотирование аудиоданных	410
10.7	Дополнительная литература по качеству аннотирования для различных задач машинного обучения	411
10.7.1	Дополнительная литература по компьютерному зрению.....	411
10.7.2	Дополнительная литература по аннотированию для обработки естественного языка	412
10.7.3	Дополнительная литература по аннотированию для информационного поиска.....	413
	Резюме.....	413

Часть IV ВЗАИМОДЕЙСТВИЕ ЧЕЛОВЕКА И КОМПЬЮТЕРА ПРИ МАШИННОМ ОБУЧЕНИИ.....415

11	Интерфейсы для аннотирования данных.....	417
11.1	Основные принципы взаимодействия человека и компьютера	418
11.1.1	Знакомство с доступностью, обратной связью и самостоятельностью	418
11.1.2	Проектирование интерфейсов для аннотирования.....	420
11.1.3	Сведение к минимуму движения глаз и прокрутки	421
11.1.4	Клавиатурные сочетания и устройства ввода	424
11.2	Эффективное нарушение правил.....	426
11.2.1	Прокрутка для пакетного аннотирования.....	426
11.2.2	Ножные педали	427
11.2.3	Голосовой ввод	427
11.3	Прайминг в интерфейсах аннотирования.....	428
11.3.1	Прайминг повторов.....	428
11.3.2	Где прайминг вреден.....	429
11.3.3	Где прайминг полезен	430
11.4	Сочетание интеллекта человека и машины	430
11.4.1	Обратная связь с аннотатором	431
11.4.2	Максимальная объективность за счет стороннего мнения.....	432

11.4.3	Преобразование непрерывных проблем в проблемы ранжирования	433
11.5	Интеллектуальные интерфейсы для максимальной отдачи человеческого интеллекта	435
11.5.1	Интеллектуальные интерфейсы для семантической сегментации	437
11.5.2	Интеллектуальные интерфейсы для распознавания объектов	440
11.5.3	Интеллектуальные интерфейсы для генерации языка	442
11.5.4	Интеллектуальные интерфейсы для маркировки последовательностей	445
11.6	Машинное обучение для содействия работе человека	447
11.6.1	Восприятие повышения эффективности	447
11.6.2	Активное обучение для повышения эффективности	448
11.6.3	Ошибки лучше их отсутствия для максимальной завершенности	449
11.6.4	Держите интерфейсы аннотирования отдельно от повседневных рабочих интерфейсов	450
11.7	Дополнительная литература	451
	Резюме	451

12 Продукты машинного обучения с участием человека

12.1	Определение продуктов для приложений машинного обучения с участием человека	454
12.1.1	Начните с решаемой вами задачи	454
12.1.2	Проектирование систем для решения задачи	455
12.1.3	Соединение Python и HTML	457
12.2	Пример 1: исследовательский анализ данных по заголовкам новостей	458
12.2.1	Предпосылки	459
12.2.2	Разработка и воплощение	460
12.2.3	Потенциальные расширения	461
12.3	Пример 2: сбор данных о событиях в области безопасности пищевых продуктов	462
12.3.1	Предпосылки	463
12.3.2	Разработка и реализация	464
12.3.3	Потенциальные расширения	465
12.4	Пример 3: идентификация велосипедов на изображениях	466
12.4.1	Предпосылки	466
12.4.2	Разработка и реализация	467
12.4.3	Потенциальные расширения	468
12.5	Дополнительная литература по созданию продуктов машинного обучения с участием человека	469
	Резюме	469
	Приложение. Краткое пособие по машинному обучению	470
	Предметный указатель	488

Предисловие

Сегодня, когда машинное обучение широко применяется во многих отраслях экономики, системы искусственного интеллекта ежедневно взаимодействуют с человеком и его социальным окружением. Многие уже заметили некоторые из последствий такого взаимодействия для пользователей. Машинное обучение может либо улучшать жизнь людей, как, например, в случае с технологиями распознавания речи и понимания естественного языка голосовым ассистентом, либо раздражать или даже активно вредить им, и примеров тому множество: от раздражающе назойливых рекомендаций продуктов до систем проверки резюме с систематически предвзятым отношением к женщинам или недостаточно представленным этническим группам. Вместо размышлений об искусственном интеллекте, действующем в отрыве от человека, в этом веке назрела острая необходимость изучения искусственного интеллекта с упором на взаимодействие с человеком – то есть создания технологий ИИ, которые эффективно сотрудничают и взаимодействуют с людьми, а также расширяют их возможности.

Эта книга нацелена не на внимание со стороны конечных пользователей, а на изучение взаимодействия людей и машинного обучения в процессе создания и эксплуатации систем машинного обучения. Для специалистов в области практического использования систем машинного обучения не секрет тот факт, что получение нужных данных с правильными аннотациями во много раз ценнее, чем использование более совершенного алгоритма машинного обучения. Получение, отбор и аннотирование данных требуют приложения больших усилий со стороны человека. Ручная маркировка данных может быть дорогой и ненадежной, и в этой книге уделено много времени этой проблеме. Одно из возможных направлений решения проблемы – сокращение объема данных для маркировки, но с возможностью обучения высококачественных систем с помощью методов активного обучения. Другое направление – использование машинного обучения и методов взаимодействия человека и компьютера для повышения скорости и точности аннотирования человеком. На этом деле не заканчивается: большинство крупных развернутых систем также предполагают различные виды проверки и обновления данных человеком. И в этом случае машинное обучение может быть направлено либо на повышение эффективности труда человека, либо на преодоление трудностей, с которыми людям приходится сталкиваться.

Роберт Монарх является высококвалифицированным проводником в этом путешествии. В своей работе – как до, так и во время получения докторской степени – Роберт уделял основное внимание практической деятельности

и внимательному отношению к людям. Он был одним из первопроходцев в применении технологий обработки естественного языка (Natural Language Processing, NLP) для анализа сообщений о ликвидации последствий стихийных бедствий с опорой на свои собственные знания, полученные в ходе оказания помощи в нескольких кризисных ситуациях. Он начинал с методов обработки критических данных человеком, а затем искал наилучшие способы использования NLP для автоматизации отдельных процессов. Я рад, что многие из этих методов сегодня используются организациями по ликвидации последствий стихийных бедствий, и могу поделиться ими с широкой аудиторией в данной книге.

В то время как область обработки данных в машинном обучении часто воспринимается в основном как работа по управлению людьми, эта книга свидетельствует о ее высокой технической составляющей. Алгоритмы выборки данных и контроля качества аннотирования нередко близки по своей сложности к алгоритмам построения последующей модели, потребляющей обучающие данные, а в некоторых случаях в процессе аннотирования применяются методы машинного обучения и обучения переноса. Существует реальная потребность в большем количестве информационных ресурсов по процессу аннотирования, и эта книга уже начала оказывать влияние даже в процессе ее написания. По мере публикации отдельных глав их читали специалисты по анализу данных в крупных организациях в таких областях, как сельское хозяйство, развлечения и путешествия. Это свидетельствует как о широком распространении машинного обучения, так и о большой потребности в книгах по работе с данными. В этой книге кодифицированы многие из лучших современных практик и алгоритмов, но поскольку долгое время область обработки данных оставалась без внимания, я надеюсь, что предстоит сделать еще больше научных открытий в области машинного обучения с фокусом на данных и что наличие такого первичного руководства будет способствовать дальнейшему прогрессу.

– КРИСТОФЕР Д. МЭННИНГ

*Кристофер Д. Мэннинг (Christopher D. Manning),
профессор информатики и лингвистики в Стэнфордском университете,
директор Стэнфордской лаборатории искусственного интеллекта
и содиректор Стэнфордского института искусственного интеллекта,
ориентированного на человека*

Введение

Я передаю все авторские доходы от этой книги на развитие инициатив по созданию лучших наборов данных, особенно для языков с ограниченными ресурсами, а также для здравоохранения и ликвидации последствий стихийных бедствий. Когда я начинал писать эту книгу, примеры наборов данных о реагировании на стихийные бедствия были редкими и специфичными для моего двойного профиля – в качестве научного сотрудника по машинному обучению и специалиста по реагированию на стихийные бедствия. После начала пандемии COVID-19 глобальная картина изменилась, и теперь многие понимают всю важность примеров использования данных для ликвидации последствий стихийных бедствий. Пандемия выявила множество пробелов в наших навыках машинного обучения, особенно в том, что касается доступа к актуальной медицинской информации и борьбы с кампаниями по дезинформации. Когда поисковые системы не смогли вывести на поверхность самую актуальную информацию о здравоохранении, а платформы социальных сетей не смогли выявить широко распространенную дезинформацию, все мы на собственном опыте прочувствовали недостатки приложений, которые не смогли достаточно быстро адаптироваться к изменяющимся данным.

Эта книга не ограничивается рассмотрением вопросов ликвидации последствий стихийных бедствий. Наблюдения и методы, которыми я здесь поделился, также основаны на моем опыте создания наборов данных для автономных транспортных средств, музыкальных рекомендаций, онлайн-коммерции, устройств с голосовым управлением, перевода и широкого спектра других практических приложений. Мне было приятно узнать о многих новых приложениях во время написания книги. От специалистов по обработке данных, которые читали черновики глав, я узнал о практических примерах внедрения в организациях, которые исторически не были связаны с машинным обучением: сельскохозяйственная компания устанавливает умные камеры на тракторах, развлекательная компания адаптирует распознавание лиц для персонажей мультфильмов, экологическая компания прогнозирует углеродные выбросы, а компания по производству одежды персонализирует модные рекомендации. Когда я выступал с приглашенными докладами о книге в этих лабораториях по изучению данных, я уверен, что узнал больше, чем рассказал сам!

Все эти примеры использования имели две общие черты: специалистам по работе с данными требовалось создать лучшие данные для обучения и оценки своих моделей машинного обучения, а информации о том, как создавать такие данные, практически не было. Я рад поделиться стратегиями и методами, которые позволят помочь системам с сочетанием человеческого и машинного интеллекта практически в любом приложении машинного обучения.

Об этой книге

Это та книга, о существовании которой я мечтал во время знакомства с машинным обучением, потому что в ней рассматривается самая важная проблема искусственного интеллекта: как люди и машины должны работать вместе для решения проблем? Большинство моделей машинного обучения строятся на человеческих примерах, но большинство текстов и курсов по машинному обучению сосредоточены только на алгоритмах. Часто можно получить самые качественные результаты с хорошими данными и простыми алгоритмами, но редко можно получить качественные результаты с лучшим алгоритмом, построенным на плохих данных. Поэтому если вам нужно углубиться в одну из областей машинного обучения, можно с уверенностью утверждать, что данные важны в первую очередь.

Кому стоит ознакомиться с этой книгой

Эта книга предназначена в первую очередь для специалистов по работе с данными, разработчиков программного обеспечения и студентов, которые только недавно начали работать с машинным обучением (или недавно начали работать с данными). Вам необходимо иметь некоторый опыт работы с такими категориями, как контролируемое и неконтролируемое машинное обучение, обучение и тестирование моделей машинного обучения, а также с такими библиотеками, как PyTorch и TensorFlow. Но для начала чтения этой книги не обязательно быть экспертом в любой из этих областей.

По мере приобретения опыта эта книга будет оставаться полезным кратким справочником по различным методикам. Эта книга является первой, где собраны наиболее распространенные стратегии аннотирования, активного обучения и смежных задач, таких как проектирование интерфейса для аннотирования.

Как организована эта книга: план действий

Эта книга состоит из четырех частей: введение, глубокое погружение в активное обучение, глубокое погружение в аннотирование и заключительная часть, которая объединяет все вместе со стратегиями проектирования интерфейсов для человека и тремя примерами реализации.

В первой части книги представлены структурные элементы для создания учебных и оценочных данных: аннотирование, активное обучение и концепции взаимодействия человека и компьютера, которые помогают людям и машинам наиболее эффективно объединить свой интеллект. К концу второй главы вы построите приложение машинного обучения с участием человека для маркировки заголовков новостей, завершив цикл от аннотирования новых данных до переобучения модели, а затем используя новую модель для принятия решения о том, какие данные следует аннотировать в следующий раз.

Вторая часть посвящена активному обучению – набору методов для выборки наиболее важных данных для анализа человеком. В главе 3 рассматриваются наиболее распространенные методы выявления неопределенности модели, а в главе 4 – сложная проблема определения того, где ваша модель может быть уверенной, но ошибочной из-за недостаточной выборки или отсутствия репрезентативных данных. В главе 5 представлены способы объединения различных стратегий в комплексную систему активного обучения, а в главе 6 рассказывается о применении методов активного обучения к различным видам задач машинного обучения.

Третья часть посвящена аннотированию – нередко недооцениваемой проблеме получения точных и репрезентативных меток для обучающих и оценочных данных. Глава 7 рассказывает о том, как найти и управлять нужными людьми для аннотирования данных. Глава 8 посвящена основам контроля качества аннотирования, в ней представлены наиболее распространенные способы расчета точности и согласия. В главе 9 рассматриваются современные стратегии контроля качества аннотирования, включая аннотирование для субъективных задач и широкий спектр методов полуавтоматического аннотирования с помощью систем на основе правил, поиска, обучения переноса, частично контролируемого обучения, самоконтролируемого обучения и создания синтетических данных. Глава 10 рассказывает о методах управления процессом аннотирования для различных видов задач машинного обучения.

Четвертая часть завершается глубоким погружением в изучение интерфейсов для эффективного аннотирования в главе 11 и тремя примерами приложений машинного обучения с участием человека в главе 12.

На протяжении всей книги мы постоянно возвращаемся к примерам из различных задач машинного обучения: маркировке изображений и документов, непрерывным данным, распознаванию объектов,

семантической сегментации, маркировке последовательностей, языковой генерации и информационному поиску. На внутренней стороне обложки приведены краткие ссылки с указанием мест, где можно найти эти задачи по всей книге.

О коде

Весь код, используемый в этой книге, является открытым исходным кодом и доступен из моего аккаунта на GitHub. Код, использованный в первых шести главах этой книги, находится на https://github.com/rmunro/pytorch_active_learning.

В некоторых главах для анализа также применяются электронные таблицы, а три примера последней главы находятся в собственных репозиториях. Более подробную информацию см. в соответствующих главах.

Дискуссионный форум liveBook

Приобретая книгу «Машинное обучение с участием человека», вы получаете бесплатный доступ к закрытому веб-форуму издательства Manning Publications, где можно оставлять комментарии о книге, задавать технические вопросы и рассчитывать на помощь от автора и других пользователей. Чтобы получить доступ к форуму, перейдите по адресу <https://livebook.manning.com/book/human-in-the-loop-machine-learning/welcome/v-11>. Вы можете узнать больше о форумах Manning и правилах поведения на сайте по адресу <https://livebook.manning.com/#!/discussion>.

Обязательства Manning перед нашими читателями заключаются в предоставлении места для содержательного диалога между отдельными читателями и между читателями и автором. Это не обязательно по какому-либо конкретному масштабу участия автора, чей вклад в форум остается добровольным (и неоплачиваемым). Мы считаем, что вы можете попробовать задать автору несколько сложных вопросов, чтобы он не потерял интерес к теме! Форум и архивы предыдущих обсуждений будут доступны на сайте издательства до тех пор, пока книга остается в печати.

Другие интернет-ресурсы

В каждой главе есть раздел «Дополнительная литература», и, за редким исключением, все перечисленные ресурсы бесплатны и доступны в интернете. Как я уже неоднократно говорил, ищите работы с высокой цитируемостью, которые ссылаются на те же статьи, на которые

ссылался я. Включение некоторых значимых работ не имело смысла, а многие другие значимые работы будут опубликованы после выхода этой книги.

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге, – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и Manning Publications очень серьезно относятся к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Об авторе

Др. РОБЕРТ МОНАРХ (ранее Роберт Манро), эксперт по вопросам сочетания человеческого и машинного интеллектов, в настоящее время живет в Сан-Франциско и работает в компании Apple. Роберт работал в Сьерра-Леоне, на Гаити, в Амазонии, Лондоне и Сиднее, в самых разных организациях – от стартапов до Организации Объединенных Наций. Он был генеральным директором и основателем компании Idibon, техническим директором компании Figure Eight и возглавлял первые службы обработки естественного языка и машинного перевода Amazon Web Services.

Часть I

Первые шаги

Большинство специалистов по data science основное время своей работы посвящают данным, а не алгоритмам. Тем не менее многие книги и курсы по машинному обучению уделяют внимание именно алгоритмам. Данная книга направлена на устранение пробела в знаниях, связанных с машинным обучением.

В первой части этой книги представлены основные элементы для создания данных об обучении и оценке: аннотирование, активное обучение, взаимодействие человек–компьютер, – помогающие людям и машинам эффективно объединять интеллект.

К концу второй главы вы создадите приложение машинного обучения для новостных заголовков на базе системы human-in-the-loop, заключающейся в непрерывном переобучении нейросети для предоставления более точной аналитики данных. Остальные главы помогут вам усовершенствовать свое первое приложение для более сложной выборки данных и аннотирования в результате взаимодействия человеческого и машинного интеллектов. Также книга рассказывает, как применять методы, которым вы научитесь, к различным типам задач машинного обучения, включая обнаружение объектов, семантическую сегментацию, маркировку последовательностей и языковое моделирование.

Введение в машинное обучение с участием человека

Эта глава охватывает:

- аннотирование немаркированных данных для тренировки, проверку достоверности и оценку данных;
- выборку наиболее важных немаркированных элементов данных (активное обучение);
- включение в аннотирование принципов взаимодействия человека и компьютера;
- внедрение обучения переноса для использования преимуществ информации в существующих моделях.

В отличие от роботов в кино, большинство современных искусственных интеллектов (ИИ) не могут обучаться самостоятельно; вместо этого они опираются на постоянную обратную связь от человека. Вероятно, 90 % сегодняшних приложений машинного обучения работают под контролем человека. Эта цифра охватывает широкий круг вариантов использования. Беспилотный автомобиль может безопасно везти вас по улице, потому что люди потратили тысячи часов, объясняя ему, когда его датчики видят пешехода, движущийся транспорт, дорожную разметку или другой соответствующий объект.

Ваш домашний девайс знает, что делать, когда вы говорите: «Увеличь громкость», – потому что люди потратили тысячи часов, рассказывая ему, как интерпретировать различные команды. А ваш сервис машинного перевода может переводить с одного языка на другой, потому что он был обучен на тысячах (а может, миллионах) текстов, переведенных людьми.

По сравнению с прошлым, наши умные устройства все меньше учатся у программистов, придерживающихся жестких правил кодинга, и все больше – на примерах и отзывах людей, которым не нужно кодировать. Эти закодированные человеком примеры – обучающие данные – используются для моделей машинного обучения и повышения их точности при выполнении поставленных задач. Однако программисты должны создать программное обеспечение, собирающее обратную связь от нетехнических пользователей, и это поднимает один из самых важных на сегодня вопросов: какие способы взаимодействия между людьми и алгоритмами машинного обучения для решения проблем являются правильными?

Аннотирование и активное обучение – краеугольные камни машинного обучения с участием человека. Они определяют, как вы собираете данные об обучении от людей, а также решают, какие данные показывать пользователям, если у вас ограничены бюджет или время для обратной связи с предоставлением подробных данных. Трансферное обучение¹ позволяет избежать холодного старта, адаптируя существующие модели машинного обучения к новой задаче, чтобы не начинать все с нуля. В этой главе мы познакомим вас с каждым из этих понятий.

1.1 Базовые принципы машинного обучения с участием человека

Машинное обучение с участием человека (human-in-the-loop) – это набор путей взаимодействия человеческого и машинного интеллектов в приложениях, использующих ИИ.

Обычно цель заключается в том, чтобы:

- повысить точность машинного обучения;
- быстрее получить точные целевые показания модели машинного обучения;
- сочетать машинный и человеческий интеллекты для максимальной точности;
- помочь людям наиболее эффективно выполнять их задачи с помощью машинного обучения.

В этой книге рассматриваются самые распространенные модели активного обучения и аннотирования, а также способы разработки лучшего интерфейса для ваших специалистов, работающих с данными, задачами и аннотированием. Книга предназначена для последовательного чтения, а примеры, приведенные здесь, усложняются постепенно.

¹ Обучение переноса (Transfer learning), также упоминаемое в профильной литературе как «обучение с переносом» или «трансферное обучение», представляет собой методику машинного обучения, позволяющую использовать модель с данными для решения определенной задачи для ее переобучения (обучения с переносом) с целью решения иной задачи.

Однако вряд ли вы одновременно будете применять все описанные тут решения. Таким образом, вы можете использовать эту книгу в качестве настольного справочника по каждой конкретной технике.

На рис. 1.1 показан процесс машинного обучения с участием человека для добавления меток к данным. Это может быть любой процесс маркировки: добавление темы в новостные сюжеты, классификация спортивных фотографий в соответствии с видом спорта, определение настроения комментария в социальных сетях, оценка степени откровенности видеоконтента и т. д. Во всех случаях вы можете либо автоматизировать некоторые процессы маркировки за счет машинного обучения, либо ускорить выполнение функции человеком. Во всех случаях применение лучших методов означает реализацию цикла, показанного на рис. 1.1: выборка нужных данных для маркировки – применение этих данных для обучения модели – использование модели для выборки дополнительных данных с последующим аннотированием.

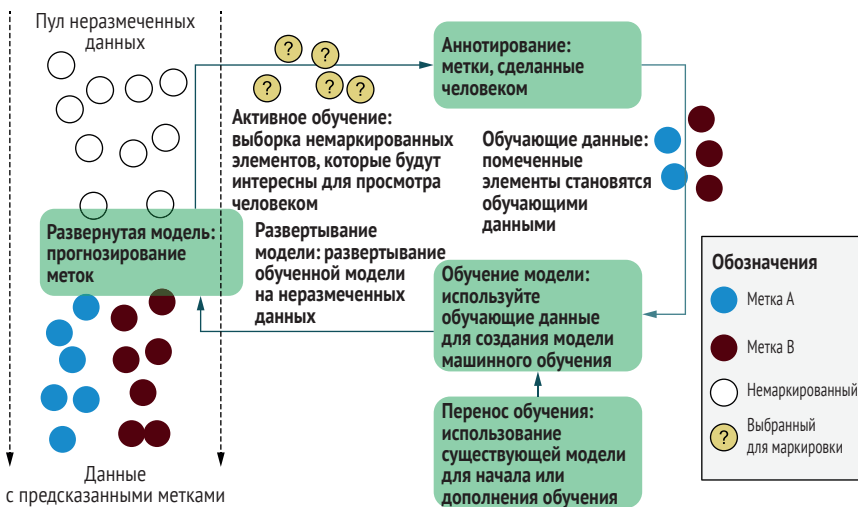


Рис. 1.1 Ментальная модель процесса human-in-the-loop для прогнозирования меток данных

В некоторых случаях вам понадобятся лишь некоторые из техник. Например, если ваша система уступает человеку, потому что модель машинного обучения является неопределенной, прочитайте главы и разделы, посвященные выборке неопределенности, качеству аннотирования и дизайну интерфейса. Этим темам посвящена большая часть данной книги, и они подходят для случаев, даже без привлечения человеческих ресурсов.

Эта книга предполагает, что вы немного знакомы с машинным обучением. Для знакомства с системами с участием человека важно глубокое понимание концепции softmax и ее ограничений. Кроме того, необходимо знать, как рассчитать точность с помощью метрик, учитывающих достоверность модели, точность с поправкой на вероятность и как из-

мерить эффективность машинного обучения с точки зрения человека (приложение содержит краткое изложение этой информации).

1.2 Введение в аннотирование

Аннотирование – это процесс маркировки необработанных данных, чтобы они могли стать тренировочными для машинного обучения. Большинство специалистов по работе с данными скажут, что они больше времени тратят на кураторство и аннотирование наборов данных, чем на построение моделей машинного обучения. Контроль качества аннотирования человеком опирается на более сложную статистику, чем ее делает большинство моделей машинного обучения, поэтому важно уделить необходимое количество времени, чтобы научиться создавать качественные обучающие данные.

1.2.1 Простые и более сложные стратегии аннотирования

Процесс аннотирования может быть простым. Если в социальных сетях вы хотите пометить сообщения о продукте как положительные, отрицательные или нейтральные, например для выводов об общих тенденциях отношения к этому продукту, вы можете создать и развернуть HTML-форму за несколько часов. Простая HTML-форма позволяет кому угодно оценить настроение каждого поста в социальных сетях, и этот рейтинг станет меткой в ваших обучающих данных.

Процесс аннотирования может быть и сложным. Например, если вы решили пометить каждый объект в видео ограничительной рамкой, простой HTML-формы недостаточно. Уже необходим графический интерфейс, помогающий аннотаторам рисовать эти рамки, и на создание хорошего пользовательского интерфейса уйдут месяцы работы инженеров.

1.2.2 Устранение пробелов в области научных знаний о данных

Стратегию машинного обучения можно оптимизировать одновременно со стратегией аннотирования данных. Они тесно взаимосвязаны, и вы получаете более высокую точность, используя комбинированный подход. Алгоритмы и аннотирования – не менее важные компоненты хорошего машинного обучения.

Все факультеты информатики предлагают курсы машинного обучения, но мало какие из них учат созданию обучающих данных. Среди пары сотен лекций по машинному обучению вы в лучшем случае найдете одну-две лекции о построении обучающих данных. Эта ситуация меняется, но пока медленно. По историческим причинам ученые-исследователи машинного обучения, как правило, не меняют наборы данных и оценивают исследования только с точки зрения алгоритмов.

В отличие от академического машинного обучения, промышленность чаще аннотирует больший объем обучающих данных с целью повышения производительности модели. Использование нескольких новых аннотаций может быть гораздо более эффективным, в отличие от попытки адаптировать существующую модель к новой области данных. Это связано с тем, что характер данных часто меняется с течением времени. Но гораздо больше научных работ сосредоточено на адаптации алгоритмов с уже имеющимися данными к новым доменам без новых обучающих данных, чем на эффективном аннотировании новых, корректных обучающих данных.

Из-за этого дисбаланса в академических кругах я часто наблюдал ошибки на практике. Предприятие нанимает дюжину докторов наук, которые знают, как создавать самые современные алгоритмы, но лишены опыта создания обучающих данных или обдумывания корректных интерфейсов для аннотирования. Недавно я наблюдал подобную ситуацию в одном из крупнейших автоконцернов. Компания наняла большое количество недавних выпускников систем машинного обучения, но не смогла внедрить технологию беспилотных авто, так как у новых сотрудников не получилось масштабировать стратегию аннотирования данных. В итоге автопроизводитель распустил всю эту команду. Я посоветовал компании, как с помощью взаимосвязанных алгоритмов и аннотирования машинного обучения перестроить свою стратегию.

1.2.3 *Качество аннотирования человеком: почему это трудно?*

Для изучающих аннотирование – это наука, тесно связанная с машинным обучением. Наиболее очевидный пример – что люди, собирающие метки, могут ошибаться, и корректировка данных требует невероятно сложной статистики. Человеческий фактор в обучающих данных может быть более или менее значимым в зависимости от сферы их применения. Если модель машинного обучения используется только для выявления общих тенденций в настроении потребителей, вероятно, 1 % погрешности в обучающих данных не столь важен. Но тот же алгоритм, приводящий в действие беспилотный автомобиль, который не распознает 1 % пешеходов, становится катастрофой.

Некоторые алгоритмы способны обрабатывать небольшой шум в обучающих данных, а каким-то из алгоритмов случайный шум даже помогает стать более точными без переобучения. Но человеческие ошибки обычно не являются случайным шумом и могут повлечь за собой неустранимую погрешность в обучающих данных. Ни один алгоритм не справится с действительно плохими данными.

Статистика определения правильной метки в случае расхождения мнений разных аннотаторов проста, если дело касается простых вычислений, таких как бинарные метки для объективных задач. Но для субъективных задач или даже для объективных с непрерывными дан-

ными не существует эвристики для определения правильной метки. Подумайте о важнейшей задаче создания обучающих данных, рисуя ограничивающую рамку вокруг каждого пешехода, которого распознает машина-беспилотник. Как быть, если у двух аннотаторов разные рамки? Какая из них правильная? Ответ не обязательно является рамкой или средним значением этих двух рамок. Фактически лучший способ объединить эти два результата – использовать машинное обучение.

Один из лучших способов получить качественное аннотирование – убедиться, что у вас есть нужные люди, которые его делают. Седьмая глава этой книги посвящена поиску, обучению и управлению лучшими аннотаторами. Пример важности сочетания правильных специалистов и правильной технологии см. ниже.

«Человеческое понимание и масштабируемое машинное обучение равны производственному ИИ», – рассказ эксперта Радхи Рамасвами Басу

Результат применения искусственного интеллекта во многом зависит от качества обучающих данных, которые вводятся. Небольшое улучшение пользовательского интерфейса, например, инструментом «волшебная палочка» для выбора областей на изображении, примененное к миллионам точек данных, в сочетании с четко определенными процессами контроля качества может заметно увеличить эффективность. Ключевой фактор – наличие высококвалифицированных специалистов. Обучение и специализация повышают качество, а при проектировании моделей важно взаимодействие опытных специалистов с экспертами узких областей знаний. Лучшие модели создаются в непрерывном взаимодействии человеческого и машинного интеллектов.

Недавно мы взяли проект, требовавший аннотирования на уровне пикселей видео роботизированного аортокоронарного шунтирования различных анатомических структур. Для наших команд аннотаторов, не являющихся экспертами в области анатомии или физиологии, мы внедрили обучение под руководством архитектора решений, являющегося квалифицированным хирургом, что помогло расширить имеющиеся навыки трехмерного пространственного мышления. Наш клиент в результате получил качественные данные обучения и оценки. Результатом для нас стали дискуссии о новых способах применения ИИ, озвученные темы, которые раньше были ограничены в своих знаниях, а теперь стал экспертом анализа медицинских изображений.

Радха Басу – основательница и генеральный директор компании iMerit. iMerit использует технологии в сфере искусственного интеллекта. Половина сотрудников компании – женщины и молодые люди из неблагополучных семей, создающие передовые технологии для ведущих мировых компаний. До этого Радха работала в HP. Позже стала главным исполнительным директором Supportsoft и основала Лабораторию экономных инноваций в Университете Санта Клары.

1.3 Введение в активное обучение: повышение скорости и снижение стоимости обучающих данных

Контролируемые модели обучения почти всегда точнее при наличии более частой маркировки данных. Активное обучение – это процесс принятия решения, какие данные использовать для аннотирования человеком. Не существует универсального алгоритма, архитектуры или набора параметров модели машинного обучения, которая была бы точной для всех вариантов и стратегий использования данных. Однако есть более успешные подходы анализа, которые стоит попробовать в первую очередь.

Большинство исследовательских работ, посвященных активному обучению, основное внимание уделяют количеству обучающих предметов, но во многих случаях важна еще и скорость. Например, при реагировании на стихийные бедствия я часто использовал машинное обучение для фильтрации и сбора информации о возникающих катастрофах. Во время стихийных бедствий любая задержка потенциально критична, поэтому оперативное создание модели важнее, чем количество в ней ярлыков.

1.3.1 Три широкие стратегии отбора активного обучения: неопределенность, разнообразие и случайность

Существует множество стратегий активного обучения, но в большинстве случаев работают три основных подхода: неопределенность, разнообразие и случайная выборка. Отправной точкой почти всегда должна быть комбинация этих трех подходов.

Случайная выборка звучит как самая простая, но на деле может оказаться самой сложной. Что является случайным, если ваши данные предварительно отфильтрованы или могут меняться с течением времени, или вы по какой-то причине знаете, что для конкретной решаемой проблемы случайная выборка не станет репрезентативной? Более подробно об этих вопросах мы поговорим в следующих разделах. Независимо от выбранной стратегии для измерения точности вашей модели и сопоставления активных стратегий обучения с базовым уровнем случайно выбранных элементов всегда требуется аннотирование некоторого объема случайных данных.

В литературе выборка случайных данных называется по-разному: разработкой или разведкой – хитрые созвучные названия, значение которых не слишком прозрачно.

Случайная выборка – это набор стратегий для выявления немаркированных элементов, пограничных для принятия решения в текущей модели машинного обучения. В бинарной классификации это элементы, вероятность принадлежности которых к меткам немного

не доходит до 50 %, поэтому модель будет неопределенной и запутанной. Скорее всего, эти элементы будут неправильно классифицированы и, вероятно, приведут к метке, отличной от прогнозируемой, смещая границу решения после их добавления к обучающим данным уже после повторного обучения модели.

Выборка разнообразия – это набор стратегий для выявления немаркированных элементов, которые недостаточно представлены или неизвестны модели машинного обучения на текущий момент. Элементы могут иметь редкие для обучающих данных функции или, например, представлять реальные демографические данные, которых пока нет в модели. Это может привести к низкой или неравномерной производительности модели, особенно при последующем изменении данных. Цель выборки разнообразия – выявить новые, необычные или недостающие для аннотирования элементы, помогающие алгоритму обучения составить более полную картину проблемной области.

Несмотря на широкое применение термина «выборка неопределенности», «выборка разнообразия» в разных областях может называться по-разному: *репрезентативной выборкой*, *стратифицированной выборкой*, *выявлением аномалий* или *обнаружением выбросов*. В некоторых случаях, таких как определение новых феноменов в астрономической базе данных или обнаружение странных сетевых активностей при проверке безопасности, целью является обнаружение выброса или аномалии, но мы можем адаптировать их в качестве стратегии для активного обучения.

Выборка неопределенности и выборка разнообразия, взятые по отдельности, имеют свои недостатки (рис. 1.2). Выборка неопределенности может сосредоточиться только на одной части границы принятия решения, а выборка разнообразия просто фокусируется на выбросах, расположенных далеко от границы. Следовательно, при нахождении набора немаркированных элементов эти стратегии используются как взаимодополняющие.

Вверху слева показана граница принятия решения алгоритмом машинного обучения в процессе выбора между элементами A и элементами B.

Вверху справа показан один возможный результат, или выборка неопределенности. Эта стратегия активного обучения эффективна при выборе немаркированных элементов, расположенных возле границы принятия решения. Высока вероятность, что они будут ошибочно предсказаны, в результате чего получают метку, что сместит границу принятия решения. Однако если вся неопределенность находится в одной проблемной области, присвоение ей меток повлияет на модель незначительно.

Внизу слева показан один из вариантов выборки разнообразия. Эта стратегия активного обучения наиболее эффективна при нахождении немаркированных элементов в очень разных частях проблемного пространства. Однако если элементы расположены далеко от границы принятия решения, то они вряд ли будут ошибочно предсказаны,

поэтому не окажут большого влияния на модель в случае присвоения им метки, совпадающей с предсказанной моделью.

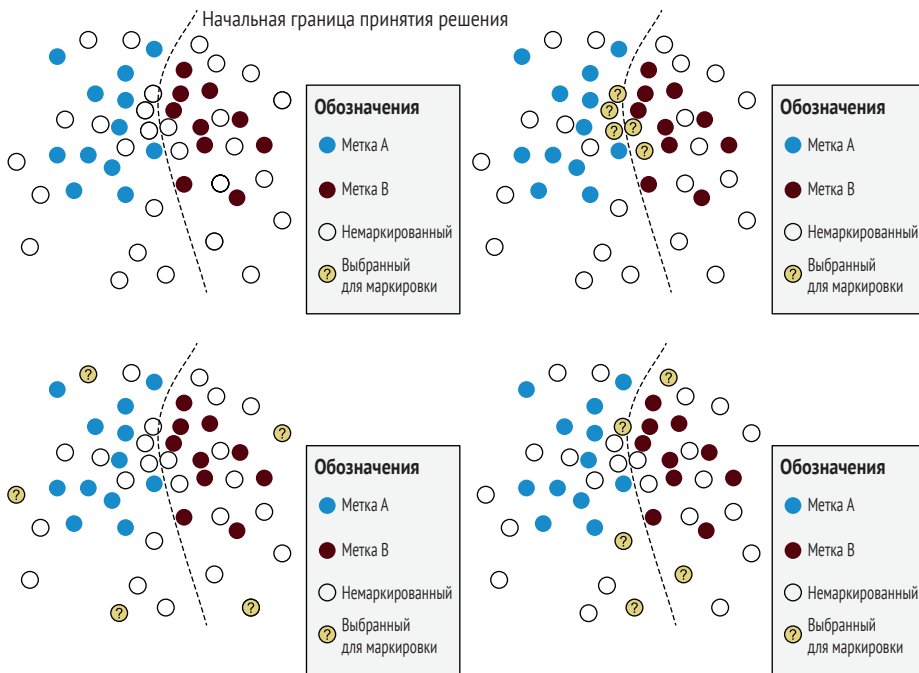


Рис. 1.2 Плюсы и минусы различных стратегий активного обучения

Внизу справа один из возможных результатов совмещения выборки неопределенности с выборкой разнообразия. При объединении стратегий выбираются элементы, по-разному удаленные от границы принятия решения. Значит, мы оптимизируем вероятность обнаружения элементов, влияющих на смещение границы принятия решения.

Важно отметить, что процесс активного обучения является повторяющимся. В процессе каждой итерации активного обучения набор элементов идентифицируется и получает от человека новую метку. Затем модель переобучается с новыми элементами, и процесс повторяется. На рис. 1.3 показаны две итерации для выборки и аннотирования новых элементов, приводящие к изменению границы.

От левого верхнего к правому нижнему: две итерации активного обучения. При каждом повторении выбираются разноудаленные от границы элементы, что смещает границу после переобучения и приводит к более точной модели машинного обучения. В идеале мы запросили метки, сделанные человеком, для минимального числа предметов в рамках нашей стратегии активного обучения. Этот запрос ускоряет время получения точной модели и снижает затраты на аннотирование человеком.

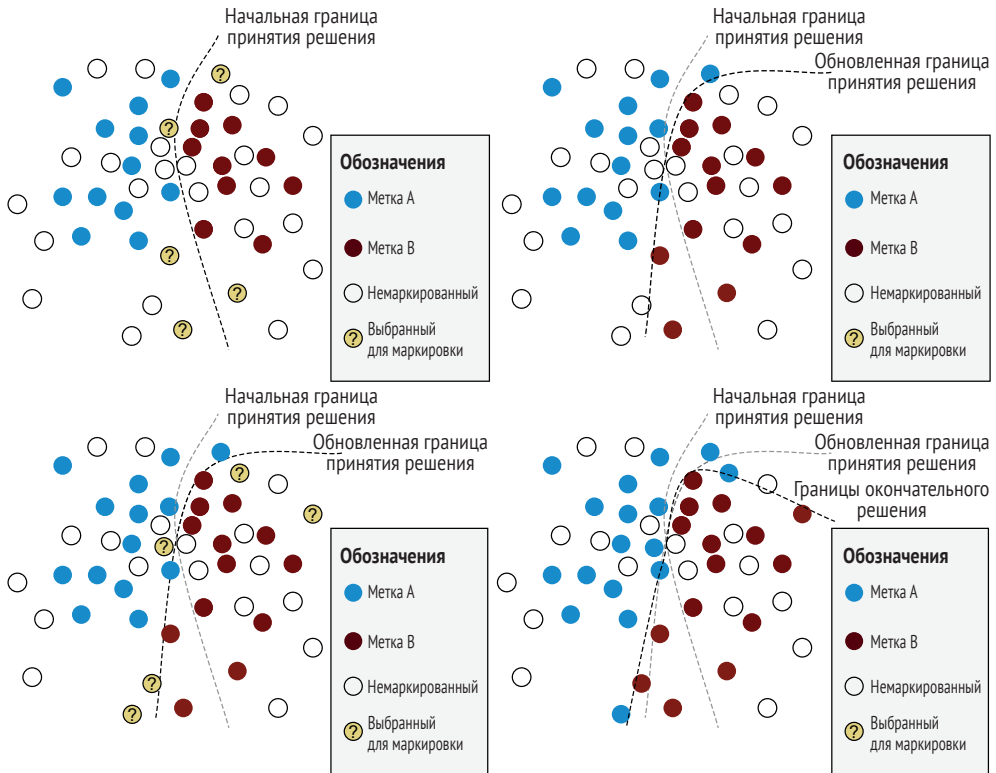


Рис. 1.3 Итеративный процесс обучения

Повторяющиеся циклы сами по себе могут быть формой выборки разнообразия. Представьте, что вы применили только выборку неопределенности и в процессе итерации использовали ее лишь для одной проблемной зоны. Вы могли бы решить всю неопределенность в этой части проблемного пространства, поэтому следующая итерация будет сосредоточена где-то в другой зоне. При достаточном количестве итераций вам вообще может не понадобиться выборка разнообразия. Каждая итерация выборки неопределенности будет сосредоточена в новой области проблемного пространства, и в сумме этих итераций будет достаточно, чтобы получить разнообразный выбор элементов обучения.

При правильной реализации активное обучение будет иметь эту самокорректирующуюся функцию: каждая итерация будет находить новые аспекты данных, которые лучше всего подходят для комментирования человеком. Однако при наличии неоднозначной области данных каждое повторение будет возвращать вас к проблемной области. Следовательно, более разумно совмещать выборку неопределенности с выборкой разнообразия, чтобы убедиться, что вы не сосредотачиваете все ваши усилия на маркировке только одной проблемной зоны, которую ваша модель не сможет решить.

Рисунки 1.2 и 1.3 дают наглядное представление о процессе обучения. Любой, кто работал с данными большого объема или последовательности, знает, что непросто определить расстояние от границы или разнообразие. По крайней мере, этот процесс сложнее, чем простое евклидово расстояние на рис. 1.2 и 1.3. Но по-прежнему актуальна та же идея: мы стремимся как можно быстрее создать точную модель, не используя человеческие метки.

Количество повторений и количество элементов, которые необходимо пометить при каждой итерации, зависят от задачи. При работе с адаптивным переводом, созданным при взаимодействии человека с машиной, одного переведенного предложения достаточно для обучения, чтобы модель обновилась в идеале в течение нескольких секунд. Это наглядно представлено на примере пользовательского опыта: если переводчик-человек исправляет машинный прогноз для какого-то слова, но машина не успевает быстро адаптироваться, человеку может потребоваться исправлять этот перевод сотни раз. Подобная проблема распространена при переводе слов, значение которых сильно зависит от контекста. Например, вы можете захотеть перевести имя человека в новостной статье буквально, а в художественном произведении адаптируете под название местности.

С технической точки зрения, конечно, гораздо сложнее быстро адаптировать модель. Рассмотрим большие модели машинного перевода. В настоящее время на их обучение тратится неделя или более. Система программного обучения, которая может быстро адаптироваться, требует непрерывного обучения по опыту переводчика. Случаи, связанные с определением настроения в социальных сетях по комментариям, над которыми я работал, как правило, требовали адаптации к новым данным примерно раз в месяц. Хотя сегодня не так много приложений с адаптивным машинным обучением в реальном времени, наблюдается тенденция к подобному подходу.

1.3.2 Что такое случайный выбор оценочных данных?

Легко сказать, что вы всегда должны оценивать по случайной выборке выхваченных данных, но на практике трудно убедиться, что это действительно случайная выборка. Если вы предварительно отфильтровали данные, с которыми работаете, по ключевому слову, времени или другому критерию, то уже получили нерепрезентативную выборку. Точность этой выборки не обязательно будет свидетельствовать о точности данных, на основании которых будет применяться ваша модель.

Я видел, как люди используют хорошо известный набор данных ImageNet и применяют модели машинного обучения к широкому спектру данных. Канонический набор данных ImageNet имеет 1000 меток, каждая из которых описывает категорию изображения, например «баскетбол», «такси» или «плавание». В задачи ImageNet входит оценка нераспределенных данных из имеющегося набора данных, чтобы система в рамках этого набора достигала точности человека. Если ту

же модель применить к случайному набору изображений, размещенных в социальных сетях, то точность сразу падает примерно до 10 %.

В большинстве приложений машинного обучения также со временем будут меняться данные. Если вы работаете с языковыми данными, то с течением времени будут меняться темы, на которые люди разговаривают, а сами языки будут обновляться и развиваться. При работе с данными компьютерного зрения меняться будут не только типы объектов, но и качество изображений вместе с прогрессом оптики.

Если вы не можете определить полноценный случайный набор оценочных данных, стоит попытаться определить репрезентативный набор данных для оценки. При определении репрезентативного набора данных вы признаете, что случайная выборка действительно невозможна или не имеет смысла для вашего набора данных. Основываясь на области применения данных, вы сами решаете, какие данные являются репрезентативными. Возможно, вы захотите выбрать точки данных для каждой интересующей вас метки, определенное количество из каждого периода времени или определенное число из выходных данных алгоритма кластеризации для обеспечения разнообразия (об этом более подробно я рассказываю в главе 4).

Вы также можете захотеть иметь несколько наборов оценочных данных, систематизированных по разным критериям. Одна из распространенных стратегий заключается в том, чтобы иметь один набор данных, взятый из того же источника, и дополнительно как минимум еще один внедоменный набор данных из другого источника для сравнения. Внедоменные наборы данных часто берутся из разных источников или из разных временных периодов. Если все обучающие данные для обработки естественного языка (NLP) получены из исторических новостных статей, то внедоменные данные могут быть взяты из последних публикаций в социальных сетях. Для большинства практических приложений вы должны использовать набор данных вне домена, так как это лучший индикатор универсального решения проблемы с помощью модели, а не просто переобучение для конкретного набора данных. Для активного обучения это может быть сложно, потому что как только вы начнете маркировать данные, они перестанут находиться вне домена. Я рекомендую сохранить набор данных вне домена, там, где не применяется активное обучение. В этом случае вы можете оценить эффективность примененной стратегии обучения: решает ли она проблему на универсальном уровне или подстраивается исключительно под те задачи, с которыми сталкивается.

1.3.3 Когда использовать активное обучение?

Вы должны использовать активное обучение, когда можете аннотировать только небольшую часть данных и когда случайная выборка не показывает всего разнообразия данных. Так как масштаб данных является важным фактором во многих случаях применения, эта рекомендация подходит для большинства сценариев в реальной жизни.

Хороший пример – количество данных, представленных на видео. Наложение ограничительной рамки вокруг каждого объекта на каждом кадре видео требует значительного времени. Представьте беспилотный автомобиль на видео с улицы, на котором, кроме него, находятся 20 небезразличных вам объектов: других автомобилей, пешеходов, знаков и т. д. При 30 кадрах в секунду потребуется $30 \text{ кадров} * 60 \text{ секунд} * 20 \text{ объектов} = 36\,000$ рамок для одноминутного видео. Самым быстрым специалистам-аннотаторам потребуется не менее 12 часов для анализа одной минуты данных.

Если посмотреть на цифры, понятно, что это трудно выполнить. Только в США люди находятся за рулем в среднем один час в день, что дает $95\,104\,400\,000$ часов, которые люди тратят на вождение в год. Скоро каждая машина будет спереди оснащена видеокамерой для помощи водителю, следовательно, для аннотирования данных вождения за год только в США потребуется $60\,000\,000\,000$ (60 трлн) часов. Сегодня на Земле не хватит людей для комментирования видео с американскими водителями, чтобы сделать вождение безопаснее, даже если весь остальной мир будет целый день только этим и заниматься.

Значит, любой специалист data science в компании по производству машин-автопилотов должен принять решение о процессе аннотации: является ли каждый N -й кадр в видео нормальным? Можно ли выбрать видео, чтобы не аннотировать их все? Есть ли способы разработать интерфейс аннотирования для ускорения процесса?

Если объем данных для аннотирования превышает бюджет или время, требуемое для проверки каждой точки данных человеком, в первую очередь необходимо использовать машинное обучение. Если бюджета и времени для ручного аннотирования всех точек данных достаточно, вероятно, вам не нужно автоматизировать задачу.

Также существуют варианты использования стратегии обучения с участием человека, где не требуется активное обучение. Это случаи, когда, например, законы требуют человеческой аннотации каждой точки данных: аудит по решению суда каждого сообщения, отправленного внутри компании для выявления возможного мошенничества. Хотя людям все равно придется просматривать каждую точку данных, активное обучение может помочь им быстрее обнаружить случаи мошенничества и определить лучший интерфейс для использования. Также активное обучение может заметить ошибки человеческого аннотирования, что используется уже сейчас.

Кроме того, есть узкие варианты применения, где почти наверняка не пригодится активное обучение. Для отслеживания работы оборудования постоянного освещения на заводе будет легко сделать модель компьютерного зрения, определяющую, включен ли свет, по лампочке или выключателю. Так как оборудование, освещение и камера с течением времени не меняются, для получения данных единожды построенной модели не требуется активное обучение. Подобных случаев, когда не требуется дополнительных данных для обучения, на моем опыте в промышленности я встречал менее 1 %.

Если ваша базовая модель уже точна для вашего бизнес-варианта применения или стоимость дополнительных обучающих данных превышает любую прибыль от более точной модели, итерации активного обучения также стоит остановить.

1.4 Машинное обучение и взаимодействие человек–компьютер

Десятилетиями множество умных людей пытались и не смогли ускорить и улучшить профессиональный перевод с помощью машинного перевода. На первый взгляд, возможность объединения человеческого и машинного переводов кажется очевидной. Как только дело доходит до необходимости исправить одну-две ошибки в машинном переводе, переводчику проще набрать все предложение заново. Использование машинного перевода предложения в качестве эталона при переводе человеком мало влияет на скорость. Если переводчик не проявит должной осторожности, он в итоге рискует закрепить машинные ошибки, тем самым снизив точность перевода.

Подходящее решение проблемы было найдено не путем уточнения алгоритмов машинного перевода, но в пользовательском интерфейсе. Вместо того чтобы заставлять переводчиков перепечатывать предложения целиком, современные системы перевода предоставляют им возможность предиктивного ввода текста, который давно является привычным явлением в телефонах и (все чаще) используется в электронной почте и программах для составления документов. Переводчики набирают перевод в обычном режиме, нажимая **Enter** или **Tab** для принятия следующего слова в предсказанном переводе, и тем самым увеличивают общую скорость при каждом правильном машинном предсказании. Так что крупнейший технический прорыв случился во взаимодействии человека и компьютера, а не в алгоритме машинного обучения.

Взаимодействие человека и компьютера – это уже сформировавшаяся область компьютерной науки, которая в последнее время стала особо важной для машинного обучения. Разрабатывая пользовательские интерфейсы для создания обучающих данных, вы используете область знаний на стыке науки о мышлении (когнитивистики), гуманитарных наук, психологии, алгоритмов пользовательского взаимодействия и ряда других.

1.4.1 Пользовательские интерфейсы: как вы создаете обучающие данные?

Зачастую для сбора обучающих данных достаточно простой веб-формы. Принципы взаимодействия человека и компьютера, лежащие в основе взаимодействия с веб-формами, довольно просты: люди

привыкли к веб-формам, потому что видят их каждый день. Они интуитивно понятны, потому что множество умных людей работали над разработкой и совершенствованием HTML-форм. Вы пользуетесь этими условностями: люди знают, как работает простая HTML-форма, и вам не нужно их обучать. С другой стороны, отказ от этих условностей может запутать людей, поэтому вы ограничены ожидаемым поведением. Возможно, у вас есть идеи для ускорения выполнения какой-то задачи с помощью динамического текста, но такое условие может больше запутать людей, нежели помочь.

Самый простой интерфейс – бинарные отклики, также является лучшим для контроля качества. Если вы можете упростить или разбить ваш проект аннотирования на бинарные задачи, вам будет гораздо проще разработать интуитивно понятный интерфейс и реализовать функции контроля качества аннотирования, о которых говорится в главах 8–11.

Когда вы имеете дело с более сложными интерфейсами, условия также становятся более сложными. Представьте, что вы просите людей обвести многоугольниками определенные объекты на изображении, что является обычным примером для компаний, занимающихся автономным транспортом. Каких действий ожидает аннотатор? Ожидает ли он линию от руки, прямые линии, использование кистей, интеллектуальное выделение по цвету/области или другие инструменты выделения? Если люди привыкли работать с изображениями в таких программах, как Adobe Photoshop, они могут ожидать такой же функциональности при аннотировании изображений. Как вы исходите из ожидания людей в отношении ограничений веб-форм, так же вы ограничены их ожиданиями в отношении выбора и редактирования изображений. К сожалению, эти ожидания могут потребовать сотен часов кодирования, если вы планируете использовать полнофункциональные интерфейсы.

Для тех, кто выполняет повторяющиеся задачи, такие как создание обучающих данных, перемещение мыши неэффективно и по возможности должно быть исключено. Если весь процесс аннотирования может происходить на клавиатуре, включая само аннотирование, ввод форм и навигацию, ритм работы аннотаторов значительно улучшится. Если все же приходится использовать мышь, компенсацией более медленного ввода должны стать более содержательные аннотации.

Для некоторых задач аннотирования требуются специализированные устройства ввода. Так, при транскрибировании речи в текст часто используют ножные педали для перемещения по временной шкале аудиозаписи. Этот процесс позволяет оставить руки на клавиатуре. Навигация по записи с помощью ног гораздо более эффективна, чем навигация по записи с помощью мыши.

Но даже для таких исключений, как транскрипция, клавиатура по-прежнему остается главной. Большинство задач аннотирования не были популярны так долго, как транскрибирование, поэтому для них

не были разработаны специализированные устройства ввода. Для большинства задач использование клавиатуры на ноутбуке или ПК быстрее, чем использование экрана планшета или телефона. Нелегко набирать текст на плоской поверхности, не отрывая глаз от вводимых данных, поэтому если только задача не является простым процессом двоичного выбора или чем-то подобным, телефоны и планшеты не подходят для аннотирования большого объема данных.

1.4.2 Прайминг: что может повлиять на человеческое восприятие?

Для получения достоверных данных для обучения необходимо учитывать сосредоточенность внимания человека-аннотатора, концентрацию его внимания, а также контекстуальные эффекты, которые могут заставить его совершить ошибку или иным образом изменить свое поведение. Рассмотрим отличный пример из лингвистических исследований. В исследовании под названием «Мягкие игрушки и восприятие речи» (<https://doi.org/10.1515/ling.2010.027>) людей просили различить австралийский и новозеландский акценты. Исследователи поместили мягкую игрушку птицы киви или кенгуру (знаковых животных для этих стран) на полку в комнате, где участники проводили исследование. Люди, проводившие исследование, не говорили участникам о мягкой игрушке; игрушка просто находилась на заднем плане. Невероятно, но люди интерпретировали акцент как более новозеландский, когда присутствовала птица киви, и более австралийский, когда присутствовал кенгуру. С учетом этого факта легко представить, что в случае создания модели машинного обучения для определения акцента (возможно, вы работаете над устройством «умного дома» для работы с максимально возможным количеством акцентов) вам необходимо учитывать контекст при сборе обучающих данных.

Явление, при котором контекст или последовательность событий может повлиять на восприятие человека, называют *прайминг* (priming). Наиболее важным типом при создании обучающих данных является *прайминг повторения*, который возникает в случае, когда последовательность задач может повлиять на восприятие человека. Например, если аннотатор маркирует сообщения в социальных сетях на предмет настроения и сталкивается с 99 сообщениями с негативным настроением подряд, он, скорее всего, допустит ошибку, маркировав сотое сообщение как негативное, в то время как оно является позитивным. Пост может быть по своей сути двусмысленным (например, сарказм) или простой ошибкой, вызванной ослаблением внимания аннотатора во время повторяющейся работы. В главе 11 я рассказываю о типах прайминга, которые необходимо контролировать.

1.4.3 Плюсы и минусы создания меток путем оценки прогнозов машинного обучения

Один из способов совместить машинное обучение и обеспечить качество аннотаций – это использование простой формы бинарного ввода, чтобы люди оценивали предсказание модели и подтверждали или отвергали это предсказание. Такая техника может быть хорошим способом превратить более сложную задачу в задачу бинарного аннотирования. Вы можете спросить кого-нибудь, корректна ли ограничительная рамка вокруг объекта – это простой бинарный вопрос, не требующий сложного интерфейса редактирования/выбора. По аналогии, проще спросить аннотатора об уместности какого-то слова в тексте, чем предоставить интерфейс для эффективного аннотирования фраз, произвольно расположенных в тексте.

И все же, поступая таким образом, вы рискуете сосредоточиться на локализованной неопределенности модели и упустить важные части проблемного пространства. Вы можете упростить интерфейс и оценку точности аннотации, поручив людям оценивать предсказания моделей машинного обучения, но вам все равно потребуется стратегия разнообразия для выборки, даже если эта стратегия сводится к обеспечению случайного выбора элементов.

1.4.4 Основные принципы проектирования интерфейсов аннотации

Вот некоторые основные принципы проектирования интерфейсов аннотаций на основании того, что я рассказал выше. Я буду более подробно рассматривать эти принципы на протяжении всей книги:

- по возможности представляйте свои вопросы в виде бинарного выбора;
- убедитесь в разнообразии ожидаемых ответов, чтобы избежать «праймминга»;
- используйте существующие правила взаимодействия;
- разрешайте ответы с клавиатуры.

1.5 Машинное обучение в помощь человеку или машинное обучение с участием человека

Машинное обучение с участием человека может преследовать две различные цели: сделать приложение машинного обучения более точным за счет вмешательства человека и улучшить работу человека с помощью машинного обучения. Иногда обе цели объединяются, хорошим примером этого служит машинный перевод. Работу переводчика можно ускорить предложением слов или фраз машинного пере-

вода, которые человек может принять или отвергнуть, подобно тому, как смартфон предсказывает следующее слово при наборе текста. Этот процесс является задачей машинного обучения, помогающего человеку обрабатывать информацию. Я также работал с клиентами, использовавшими машинный перевод из-за дороговизны человеческого перевода. Поскольку содержание данных перевода человека и машины одинаково, система машинного перевода со временем становится более точной на основе данных, переведенных человеком. Эти системы достигают обеих целей – повышения эффективности работы людей и точности машинного перевода.

Поисковые системы – еще один прекрасный пример машинного обучения с участием человека. Люди часто забывают, что поисковые системы – это одна из форм ИИ, несмотря на их повсеместное распространение для общего поиска и для конкретных случаев использования, таких как электронная коммерция и навигация (онлайн-карты). Например, когда вы ищете интернет-страницу и нажимаете на четвертую ссылку, а не на самую верхнюю, вы, вероятно, обучаете поисковый движок (информационно-поисковую систему) тому, что четвертая ссылка может быть лучшим ответом на ваш поисковый запрос. Существует распространенное заблуждение, что поисковые системы обучаются только по отзывам конечных пользователей. На самом деле все крупные поисковые системы используют тысячи аннотаторов для оценки и настройки своих поисковых движков. Оценка релевантности поиска является единственным крупнейшим случаем использования аннотации человека в машинном обучении. Несмотря на растущую популярность использования компьютерного зрения, например в автономных транспортных средствах, и речи, например в домашних устройствах и смартфонах, релевантность поиска по-прежнему остается самым крупным примером использования профессионального аннотирования человеком.

В большинстве задач машинного обучения с участием человека – как бы они ни выглядели на первый взгляд – есть элементы как машинного обучения в помощь человеку, так и машинного обучения с участием человека, поэтому необходимо проектировать и то, и другое.

1.6 Перенос обучения для запуска ваших моделей

В большинстве случаев вам не нужно создавать обучающие данные с самого начала. Зачастую существующие наборы данных уже близки к вашим требованиям. Например, при создании модели анализа настроений для отзывов о фильмах можно взять набор данных анализа настроений для отзывов о товарах, начать с него и затем адаптировать к вашим условиям использования. Этот процесс – использование

модели из одного случая применения и адаптация ее к другому – известен как *перенос обучения* (transfer learning).

В последнее время наблюдается значительный рост популярности адаптации обычных предварительно обученных моделей к новым специфическим случаям использования. Другими словами, люди создают модели *специально* для различных случаев применения переноса обучения. Такие модели часто называют *предварительно обученными* (pretrained) моделями.

Традиционный перенос обучения подразумевает передачу результатов одного процесса в другой. Примером при обработке естественного языка (natural language processing, NLP) может быть

Общий разметчик (теггер) частей речи > Синтаксический парсер > Теггер для смыслового анализа.

Сегодня перенос обучения обычно означает *переобучение части нейронной модели для адаптации к новой задаче (предварительно обученные модели) или использование параметров одной нейронной модели в качестве исходных данных для другой.*

На рис. 1.4 показан пример переноса обучения. Модель может быть обучена на одном наборе меток, а затем переобучена на другом наборе меток путем сохранения архитектуры и сохранения части модели, в данном случае – с переобучением только последнего слоя.

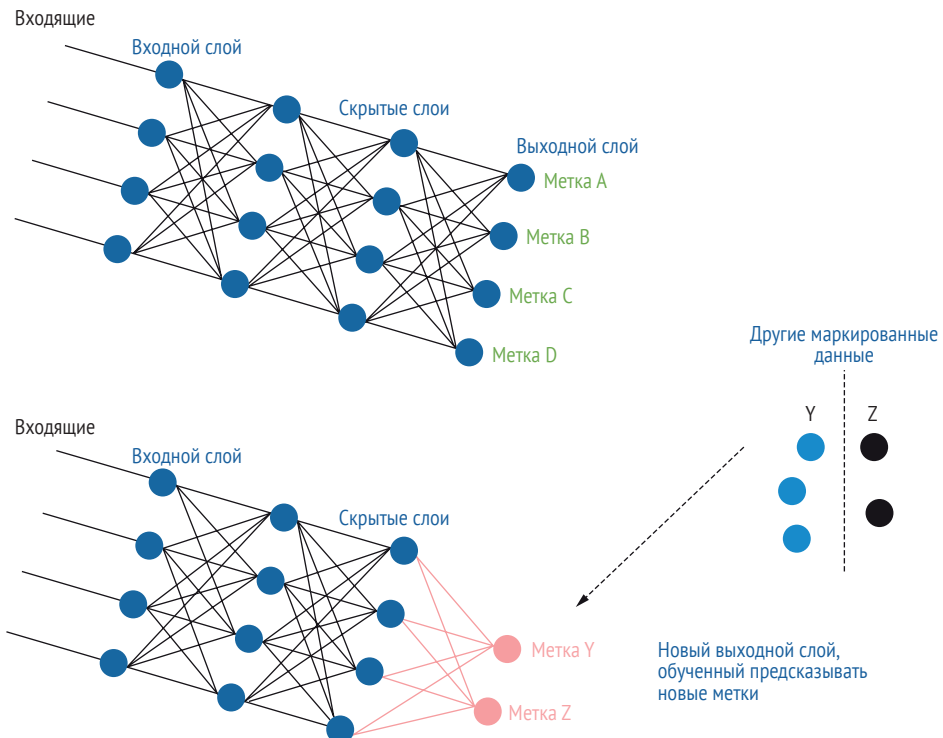


Рис. 1.4 Пример переноса обучения

Модель построена для предсказания меток A, B, C или D. Благодаря переобучению последнего слоя и использованию значительно меньшего числа помеченных человеком предметов, нежели в случае обучения с нуля, модель смогла предсказать метки Y и Z.

1.6.1 Перенос обучения в компьютерном зрении

В последнее время наибольший прогресс переноса обучения наблюдается в области компьютерного зрения. Популярной стратегией является исходное использование набора данных ImageNet и построение на основе миллионов примеров модели для классификации 1000 меток: спорт, птицы, искусственные объекты и т. д.

Чтобы научиться классифицировать различные виды спорта, животных и объектов, модель машинного обучения изучает типы текстур и контуров, необходимых для различения 1000 типов предметов на изображениях. Многие из этих текстур и контуров являются более универсальными, чем признаки 1000 меток, и могут быть использованы в других ситуациях. Поскольку все текстуры и контуры изучаются в промежуточных слоях сети, вы можете переобучить только последний слой на новом наборе меток. Для каждой новой метки могут понадобиться не миллионы, а всего несколько сотен или тысяч примеров, поскольку вы уже используете миллионы изображений для текстур и контуров. ImageNet добилась большого успеха в переобучении последнего слоя на новые метки с небольшим количеством данных, в том числе с такими объектами, как клетки в биологии и географические объекты на спутниковых снимках.

Также допустимо переобучение нескольких слоев вместо последнего и добавление в исходную модель дополнительных слоев. Обучение с переносом может использоваться с различными архитектурами и параметрами для адаптации модели к новому сценарию применения, но с той же целью уменьшения числа человеческих меток, необходимых для построения точной модели на новых данных.

Компьютерное зрение пока что ограничено рамками маркировки изображений. Для таких задач, как обнаружение объектов на изображении, трудно создать системы переноса обучения с возможностью адаптации к одному типу объектов. Проблема в том, что объекты обнаруживаются как совокупности контуров и текстур, а не как целые объекты. Однако многие работают над решением этой проблемы, и прорыв, без сомнения, будет достигнут.

1.6.2 Перенос обучения при обработке естественного языка

Серьезный спрос на использование предварительно обученных моделей в NLP появился даже позже, чем в компьютерном зрении. Перенос обучения при обработке естественного языка обрел популярность только в последние два-три года. Вот почему это одна из самых современных технологий, рассматриваемых в данном тексте, но она также может быстро устареть.

Адаптация, подобная ImageNet, не работает для языковых данных. Перенос обучения для одного набора данных анализа смыслов на другой набор дает прирост точности только примерно на 2–3 %. Модели, предсказывающие метки на уровне документа, не обладают охватом человеческой речи в той степени, в какой эквивалентные модели компьютерного зрения передают текстуры и контуры. Но вы можете узнать интересные свойства слов, рассматривая контексты, в которых они регулярно встречаются. Например, такие слова, как *доктор* и *хирург*, могут встречаться в похожих контекстах. Предположим, что вы нашли 10 000 контекстов, в которых встречается любое английское слово в наборе слов до и после него. Вы можете увидеть вероятность встречи слова *doctor* в каждом из этих 10 000 контекстов. Некоторые из этих контекстов будут связаны с медициной, поэтому у слова *doctor* будет высокий балл в данных контекстах. Но большинство из 10 000 контекстов не будут связаны с медициной, поэтому в этих контекстах у слова *doctor* будет низкий балл. Вы можете рассматривать эти 10 000 оценок как вектор длиной 10 000. Слово *хирург*, скорее всего, будет иметь вектор, похожий на вектор слова *доктор*, потому что оно часто встречается в одном и том же контексте.

Концепция понимания слова по его контексту стара и лежит в основе функциональных теорий лингвистики:

Вы узнаете слово по компании, которой оно придерживается
(Ферт, Дж. Р. 1957:11).

Строго говоря, нам нужно спуститься ниже слова, чтобы добраться до самой важной информации. Английский язык является исключением в том смысле, что слова, как правило, являются хорошими атомарными единицами для машинного обучения. Английский язык позволяет создавать сложные слова, такие как *un-do-ing*; очевидно, почему мы хотим интерпретировать отдельные части (морфемы), но в английском это делается гораздо реже, чем других языках. То, что в английском языке выражается с помощью порядка слов, например субъект–глагол–объект, гораздо чаще выражается с помощью аффиксов (добавлений к слову), которые в английском ограничиваются такими вещами, как настоящее и прошедшее время, различия между единственным и множественным числами. Поэтому для задач машинного обучения, не ориентированных на привилегированный язык вроде английского, который является исключением, нам необходимо моделировать подслово.

Ферт оценил бы этот факт. Он основал первое в Англии отделение лингвистики при Лондонском университете СОАС, где я проработал два года, помогая записывать и сохранять исчезающие языки. Во время этой работы мне стало ясно, что все многообразие языков требует от нас более глубоких характеристик, чем просто слова. Методы машинного обучения с участием человека необходимы, если мы хотим адаптировать возможности машинного обучения к как можно большему числу из 7000 языков мира.

Недавние достижения в переносе обучения стали возможны благодаря принципу понимания слов (или сегментов слов) в контексте. Мы можем бесплатно получить миллионы меток для наших моделей, если будем предсказывать слово по его контексту:

My __ is cute. He __ play-ing (Мой __ милый. Он __ игра-ет).

Никаких человеческих меток не требуется. Мы можем удалить некоторый процент слов из необработанного текста, а затем превратить оставшийся текст в задачу машинного обучения. Как вы можете догадаться, первым пропущенным словом в англоязычном примере может быть *dog* (пес), *puppy* (щенок) или *kitten* (котенок), а вторым пропущенным словом, скорее всего, будет *is* или *was*. Как и в случае со словами *хирург* и *доктор*, мы можем предсказывать слова по контексту.

В отличие от примера ранее, где перенос обучения с одного типа смысла на другой потерпел неудачу, эти виды предварительно обученных моделей оказались весьма успешны. При незначительной настройке модели, предсказывающей слово в контексте, можно создавать самые современные системы с небольшим количеством человеческих меток для таких языковых задач, как ответы на вопросы, анализ настроений и текстовые импликации. В отличие от компьютерного зрения, перенос обучения быстро набирает популярность для решения сложных задач обработки естественного языка, таких как автоматическое реферирование и перевод.

Предварительно обученные модели не являются сложными. Самые сложные из них сегодня обучены предсказывать слова в контексте, порядок слов в предложении и порядок предложений. На основе такой базовой модели с тремя типами предсказаний по характерным данным мы можем построить практически любой сценарий применения NLP. Поскольку порядок слов и порядок предложений являются неотъемлемыми свойствами документов, предварительно обученные модели не нуждаются в человеческих метках. Они по-прежнему строятся по принципу контролируемых задач машинного обучения, но обучающие данные генерируются без каких-либо затрат. Модели может быть поручено предсказать каждое десятое слово, которое было удалено из данных, и предсказать порядок следования определенных предложений в исходных документах, обеспечив мощный стартовый импульс до того момента, когда решение вашей задачи потребует разметки человеком.

И все же возможности предварительно обученных моделей ограничены количеством доступного неразмеченного текста. Гораздо больше немаркированного текста, чем на других языках, доступно на английском языке, даже с учетом общей распространенности различных языков. Существуют также культурные различия. Так, пример *My dog is cute* может часто встречаться в онлайн-текстах, которые являются основным источником данных для предварительно обученных моделей. Но не во всех культурах собака считается домашним животным. Когда я изучал язык маце (Matsés), ненадолго посетив Амазонию, вы-

яснилось, что у них популярными домашними животными были обезьяны. Фраза на английском языке *My monkey is cute* редко встречается в интернете, а эквивалент *Chuna bēdambo ikek* из языка маце вообще не встречается. Векторы слов и контекстуальные модели в предварительно обученных системах позволяют выражать несколько значений одним словом, поэтому они могут отражать и *собаку*, и *обезьяну* в рассматриваемом контексте, но они все равно пристрастны к данным, на которых они обучены, при этом контекст *обезьяны* вряд ли встречается в больших объемах на любом языке. Стоит учитывать, что предварительно обученные системы будут склонны усиливать культурные различия.

В любом случае предварительно обученные модели требуют дополнительной разметки человеком для достижения точных результатов своей работы, поэтому перенос обучения не меняет нашу общую архитектуру машинного обучения с участием человека. Однако оно может дать нам значительную фору в разметке и повлиять на выбор стратегии активного обучения, которую мы используем для выборки дополнительных элементов данных для аннотации человеком, и даже на интерфейс, с помощью которого человек предоставляет это аннотирование.

Перенос обучения лежит и в основе некоторых передовых стратегий активного обучения, обсуждаемых в главе 5, и передовых стратегий аннотирования и дополнения данных, рассмотренных в главе 9.

1.7 Чего ожидать от этого текста

Для осмысления сочетания частей этого текста друг с другом полезно представить темы в виде квадранта знаний. Он представлен на рис. 1.5, охватывает все темы этой книги и выражает их в терминах известного и неизвестного для ваших моделей машинного обучения.

Четыре квадранта:

- *известные знания* (Known knowns) – то, что ваша модель машинного обучения может уверенно и точно делать сегодня. Этот квадрант представляет собой вашу модель в ее нынешнем состоянии;
- *известные неизвестные* (Known unknowns) – то, что ваша модель машинного обучения не может уверенно делать сегодня. К этим элементам можно применить выборку неопределенности;
- *неизвестные известные* (Unknown knowns) – знания в предварительно обученных моделях, которые могут быть адаптированы к вашей задаче. Перенос обучения позволяет использовать эти знания;
- *неизвестные неизвестные* (Unknown unknowns) – пробелы в вашей модели машинного обучения. К этим элементам можно применить выборку разнообразия.

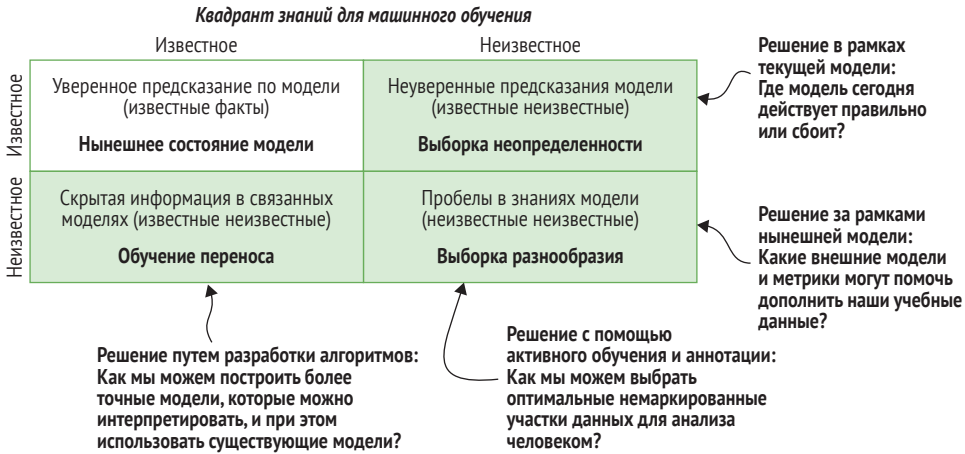


Рис. 1.5 Квадрант знаний машинного обучения

Столбцы и строки также имеют значение: строки отражают знания о вашей модели в ее нынешнем состоянии, а столбцы – тип необходимых решений:

- верхняя строка отражает знания вашей модели;
- нижняя строка отражает знания за пределами вашей модели;
- левая колонка может быть решена правильными алгоритмами;
- правая колонка может быть решена с помощью человеческого взаимодействия.

В этом тексте рассматривается широкий спектр технологий, поэтому данный рисунок будет полезно держать под рукой, чтобы знать, куда что вписывается.

В конце первых нескольких глав книги есть памятки для быстрого знакомства с основными рассмотренными темами. Вы можете держать эти памятки под рукой при чтении последующих глав.

Резюме

- Более широкая архитектура машинного обучения с участием человека является итерационным процессом, объединяющим человеческие и машинные компоненты. Понимание этих компонентов объясняет объединение частей данной книги воедино.
- Вы можете использовать некоторые базовые методы аннотирования, чтобы приступить к созданию обучающих данных. Понимание этих методов обеспечивает сбор достоверных и эффективных аннотаций.
- Две наиболее распространенные стратегии активного обучения – это выборка неопределенности и выборка разнообразия. Понимание основных принципов каждого типа поможет вам выработать

стратегию правильного сочетания подходов для решения ваших конкретных задач.

- Взаимодействие человека и компьютера предоставляет вам концепцию для разработки компонентов пользовательского восприятия систем машинного обучения с участием человека.
- Перенос обучения дает возможность адаптировать модели, обученные для одной задачи, к другой и строить более точные модели с меньшим количеством аннотаций.