

Мы посвящаем эту книгу памяти Дэвида Книла (1961–2014). Большая часть того, что все мы трое знаем о байесовском моделировании восприятия и действия, – его заслуга. Он был заботливым и терпеливым наставником и отличным учителем и сделал изучение этой темы комфортным и приятным. Область байесовского моделирования восприятия и действия не была бы такой, какой она есть, без него, и эта книга, вероятно, никогда бы не была написана

Содержание

От издательства	14
Благодарности	15
Сокращения	18
Введение	19
Глава 1. Неопределенность и вывод	24
1.1. Цель восприятия	25
1.2. Гипотезы и их вероятности	26
1.3. Сенсорный шум и неоднозначность восприятия.....	31
1.4. Байесовский вывод в зрительном восприятии	33
1.5. Байесовский вывод в слуховом восприятии	38
1.6. Исторический обзор: восприятие как бессознательное умозаключение.....	45
1.7. Заключение.....	46
1.8. Рекомендуемая литература	47
1.9. Задачи	48
Глава 2. Применение правила Байеса	52
2.1. Этапы байесовского моделирования	52
2.2. Альтернативная форма правила Байеса	57
2.3. Площадное представление	57
2.4. Ошибка прокурора.....	59
2.5. Смена априорного убеждения: пример багажной карусели	61
2.6. Плоское априорное распределение: пример гештальт-восприятия	65
2.7. Оптимальность, эволюция и мотивы байесовского моделирования.....	68
2.8. Заключение	69
2.9. Рекомендуемая литература	70
2.10. Задачи	70
Глава 3. Байесовский вывод в условиях зашумленных измерений	78
3.1. Этапы байесовского моделирования	79
3.2. Этап 1: порождающая модель	80

3.2.1. Измерение: абстрактное сенсорное представление	81
3.2.2. Графовая модель	82
3.2.3. Распределение стимулов	83
3.2.4. Распределение измерений	84
3.2.5. Совместное распределение	86
3.3. Этап 2: вывод	86
3.3.1. Априорное распределение	87
3.3.2. Функция правдоподобия	88
3.3.3. Апостериорное распределение	90
3.3.4. Апостериорное среднее	94
3.3.5. Ширина апостериорного распределения	96
3.3.6. Оценка апостериорного среднего	96
3.3.7. Оценка MAP	97
3.4. Неопределенность и уверенность	97
3.4.1. Неопределенность	97
3.4.2. Байесовская уверенность	99
3.5. Несоответствие модели в выводе	101
3.5.1. Априорное несоответствие	101
3.5.2. Неудовлетворительные априорные распределения	102
3.6. Гетероскедастичность	103
3.7. Амплитудные переменные	103
3.8. Применение байесовских моделей	105
3.9. Восприятие	106
3.10. Заключение	107
3.11. Рекомендуемая литература	108
3.12. Задачи	108
Глава 4. Распределение отклика	114
4.1. Унаследованная изменчивость	114
4.2. Распределение отклика	116
4.3. Распределение убеждений или распределение отклика?	118
4.4. Оценка максимального правдоподобия	120
4.5. Смещение и среднеквадратичная ошибка	121
4.5.1. Перспектива «обратного смещения»	123
4.5.2. Все расходы оплачены	125
4.6. Другие оценки	127
4.7. Шум принятия решения и шум отклика	128
4.8. Заблуждения	128
4.9. Размышления о байесовских моделях	131
4.10. Заключение	132
4.11. Рекомендуемая литература	132
4.12. Задачи	132
Глава 5. Комбинация признаков и накопление свидетельств	140
5.1. Что такое объединение сигналов?	141
5.2. Формулировка байесовской модели	144

5.2.1. Этап 1: порождающая модель.....	144
5.2.2. Этап 2: вывод.....	146
5.2.3. Этап 3: оценка распределения.....	149
5.3. Искусственный конфликт сигналов	149
5.3.1. Различение распределений.....	151
5.4. Обобщения: априорные распределения, множественные сигналы	151
5.5. Накопление свидетельств	153
5.6. Объединение сигналов при неоднозначности	154
5.7. Применение объединения сигналов	155
5.8. Заключение	157
5.9. Рекомендуемая литература	157
5.10. Задачи	158
Глава 6. Обучение как вывод	163
6.1. Множество форм обучения.....	163
6.2. Изучение вероятности бинарного события	165
6.2.1. Прогноз.....	168
6.2.2. Уравнения обновления	170
6.2.3. Неопределенность	170
6.2.4. Биномиальное распределение.....	170
6.2.5. Неравномерное априорное распределение.....	171
6.3. Связь между байесовским обучением и обучением с подкреплением	172
6.4. Изучение точности нормального распределения.....	174
6.4.1. Почему бы не вывести дисперсию?	177
6.4.2. Прогноз.....	177
6.5. Изучение наклона линейной зависимости	178
6.6. Изучение структуры причинно-следственной модели.....	180
6.7. Другие формы обучения.....	182
6.8. Заключение	183
6.9. Рекомендуемая литература	183
6.10. Задача.....	184
Глава 7. Различение и обнаружение	188
7.1. Примеры задач	188
7.2. Различение	189
7.2.1. Этап 1: порождающая модель	189
7.2.2. Этап 2: вывод	190
7.2.3. Гауссова модель	194
7.2.4. Решающее правило с точки зрения измерения	195
7.2.5. Несколько задач могут иметь одно и то же байесовское решающее правило.....	196
7.2.6. Этап 3: распределение отклика	196
7.3. Обнаружение	199
7.4. Уверенность в решении.....	200
7.5. Дополнительные характеристики распределения откликов.....	203
7.5.1. Рабочая характеристика приемника.....	203

7.5.2. Различимость.....	205
7.6. Связь между байесовским выводом и теорией обнаружения сигналов ...	207
7.7. Промежуточные варианты	208
7.8. Заключение.....	208
7.9. Рекомендуемая литература.....	209
7.10. Задачи.....	209
Глава 8. Бинарная классификация	214
8.1. Примеры задач бинарной классификации	214
8.2. Порождающая модель	217
8.2.1. Зеркально симметричные распределения стимулов, обусловленные классами.....	217
8.3. Маргинализация.....	218
8.3.1. Сумма двух бросков кубика.....	219
8.3.2. Непрерывные переменные	221
8.3.3. Условная маргинализация.....	221
8.3.4. Использование порождающей модели.....	222
8.4. Вывод	223
8.5. Распределение отклика.....	226
8.6. Другие распределения стимулов, обусловленных классами.....	228
8.7. «Следуйте по стрелкам».....	231
8.8. Заключение	232
8.9. Рекомендуемая литература	232
8.10. Задачи	233
Глава 9. Мешающие переменные верхнего уровня и неоднозначность.....	239
9.1. Примеры задач	239
9.2. Размер как мешающая переменная верхнего уровня в восприятии глубины сцены.....	241
9.3. Маргинализация.....	246
9.4. Цветовое зрение.....	247
9.5. Распознавание объектов	250
9.6. Заключение	252
9.7. Рекомендуемая литература.....	252
9.8. Задачи	253
Глава 10. Одинаковый или разный?	256
10.1. Примеры задач	256
10.2. Бинарные стимулы	257
10.2.1. Этап 1: порождающая модель.....	257
10.2.2. Этап 2: вывод.....	258
10.2.3. Этап 3: оценка распределения.....	260
10.3. Непрерывные стимулы	262
10.3.1. Этап 1: порождающая модель.....	262

10.3.2. Этап 2: вывод.....	264
10.3.3. Этап 3: вероятность отклика.....	265
10.3.4. Пересмотр этапа 2: вывод стимулов.....	266
10.4. Оценка сходства нескольких элементов.....	268
10.5. Организация восприятия.....	270
10.5.1. Формирование контура.....	272
10.5.2. Пересекающиеся линии.....	274
10.6. Заключение.....	277
10.7. Рекомендуемая литература.....	277
10.8. Задачи.....	278

Глава 11. Поиск..... 281

11.1. Разнообразие форм зрительного поиска.....	282
11.2. Локализация цели при маскировке.....	283
11.3. Локализация цели с помощью зашумленных измерений.....	287
11.4. Обнаружение цели при маскировке.....	290
11.5. Обнаружение цели с помощью зашумленных измерений.....	293
11.6. Применение.....	295
11.7. Заключение.....	296
11.8. Рекомендуемая литература.....	296
11.9. Задачи.....	297

Глава 12. Вывод в меняющемся мире..... 302

12.1. Отслеживание непрерывно меняющегося состояния мира.....	302
12.1.1. Этап 1: порождающая модель.....	303
12.1.2. Этап 2: вывод.....	305
12.2. Обнаружение точки изменения.....	308
12.2.1. Единственная точка изменения.....	308
12.2.2. Случайные точки изменения.....	310
12.2.3. Более реалистичное обнаружение точки изменения.....	312
12.3. Заключение.....	312
12.4. Рекомендуемая литература.....	313
12.5. Задачи.....	313

Глава 13. Сочетание вывода с полезностью..... 317

13.1. Примеры задач.....	317
13.2. Выбор между двумя действиями.....	319
13.3. Выбор между несколькими действиями.....	320
13.4. Математическое определение ожидаемой полезности.....	323
13.4.1. Бинарная классификация.....	324
13.4.2. Непрерывная оценка.....	325
13.5. Функции затрат «чистого восприятия».....	327
13.5.1. Дискретные задачи.....	327
13.5.2. Непрерывная оценка.....	328
13.5.3. Восприятие и действие.....	329

13.6. Что значит принятие оптимального решения	330
13.7. Усложненные ситуации	331
13.7.1. Функции стоимости для неопределенных результатов	331
13.7.2. Нелинейная зависимость между вознаграждением и полезностью	332
13.7.3. Искажения вероятности	334
13.7.4. Шум принятия решения	334
13.8. Применение	334
13.8.1. Зрительное различие	335
13.8.2. Целенаправленные движения	336
13.8.3. Поощряемые доверительные интервалы	336
13.8.4. Вывод функций полезности	338
13.9. Заключение	339
13.10. Рекомендуемая литература	340
13.11. Задачи	340
Глава 14. Нейронная функция правдоподобия	346
14.1. Порождающая модель активности одиночного нейрона	347
14.2. Нейронная функция правдоподобия для одного нейрона	349
14.2.1. Случай 1: колоколообразная настроечная кривая	351
14.2.2. Случай 2: монотонная настроечная кривая	352
14.3. Нейронная функция правдоподобия, основанная на популяции нейронов	354
14.4. Упрощенная модель	357
14.5. Связь между концепциями абстрактной и нейронной моделей	360
14.6. Использование нейронной функции правдоподобия для вычислений	362
14.7. Нейронная реализация байесовских вычислений	364
14.8. Применение байесовских вычислений	364
14.9. Заключение	365
14.10. Рекомендуемая литература	366
14.11. Задачи	367
Глава 15. Байесовские модели в контексте	370
15.1. Байесовское и оптимальное поведение	370
15.2. Чрезмерно сильные заявления об оптимальности	371
15.3. Почему одни модели поведения оптимальны, а другие нет	372
15.4. Байесовские модели не являются механистическими	373
15.5. Байесовский перенос	374
15.6. Вероятностные вычисления и гибридные модели	375
15.7. Сложность реального мира	376
15.8. Как стать настоящим байесовцем	378
15.9. Рекомендуемая литература	379
Приложение А. Обозначения	381

Приложение В. Основы теории вероятностей	384
V.1. Объективная и субъективная вероятности	384
V.2. Смысловое толкование вероятности	385
V.3. Дополняющее событие	386
V.4. Представление в виде диаграммы Венна	386
V.5. Случайные величины и их распределения	387
V.5.1. Сравнение дискретных и непрерывных случайных величин	388
V.5.2. Суммарная вероятность всегда равна 1	389
V.5.3. Дискретные распределения вероятности	389
V.5.4. Непрерывные распределения вероятности	390
V.5.5. Формальное определение функции плотности вероятности	392
V.5.6. Нормирование	393
V.6. Среднее значение, дисперсия и математическое ожидание	394
V.7. Нормальное распределение	395
V.7.1. Определение	395
V.7.2. Центральная предельная теорема	395
V.7.3. Произведение двух нормальных распределений	396
V.7.4. Произведение нескольких нормальных распределений	397
V.7.5. Интегральное нормальное распределение	398
V.7.6. Распределение фон Мизеса	398
V.8. Дельта-функция	399
V.9. Распределение Пуассона	400
V.10. Выборка из распределения вероятностей	400
V.11. Распределения, включающие несколько переменных	401
V.11.1. Совместная вероятность	401
V.11.2. Маргинализация	402
V.11.3. Условная вероятность	403
V.11.4. Независимость	405
V.11.5. Правило Байеса	406
V.12. Функции случайных величин	408
V.12.1. Функции одной переменной: изменение переменных	408
V.12.2. Пример с яблоками	410
V.12.3. Обязательство неведения	412
V.12.4. Математическое представление маргинализации	412
V.12.5. Функции нескольких переменных	414
V.13. Задачи	416
Приложение С. Подбор параметров и сравнение моделей	420
C.1. Что такое модель?	420
C.2. Свободные параметры	421
C.3. Вероятность параметра	421
C.4. Оценка максимального правдоподобия	422
C.5. Подбор параметров по данным из задачи оценки	423
C.5.1. Простая модель	424
C.5.2. Более удачная модель	429
C.5.3. Сравнение моделей	430

С.5.4. Тест отношения правдоподобия	431
С.5.5. Перекрестная проверка	431
С.5.6. Сопоставление методов сравнения моделей	432
С.5.7. Восстановление параметров и восстановление модели.....	433
С.5.8. Ограничения сравнения моделей.....	433
С.6. Абсолютное качество подгонки.....	434
С.7. Подбор параметров модели по данным в задаче различения.....	435
С.7.1. Простая модель	435
С.7.2. Более удачная модель?.....	437
С.8. Подбор параметров модели по данным в задаче классификации.....	437
С.9. Правильный план эксперимента для байесовского моделирования	438
С.10. Рекомендуемая литература.....	439
С.11. Задачи.....	440
Библиография.....	441
Предметный указатель.....	454

Благодарности

Мы долго шли к написанию этой книги и в долгу перед многими людьми. Впервые эта идея пришла нам в голову в июне 2009 г., когда мы вместе с Аланом Стокером и Джонатаном Пиллоу читали курс вычислительной нейробиологии в Instituto Gulbenkian de Ciência в Оэйраше, Португалия. В то время неудержимый оптимист Конрад предсказывал, что мы закончим книгу к декабрю 2009 г. И вот, спустя всего четырнадцать лет, книга у нас в руках. Однако отсрочка имела свои преимущества: на протяжении многих лет мы использовали черновики глав и идеи книги для обучения байесовскому моделированию сотен студентов бакалавриата, аспирантов и докторантов на наших курсах в Университете Макмастера, Медицинском колледже Бейлора, Северо-Западном университете, Нью-Йоркском университете, Университете Пенсильвании, а также в учебных пособиях на конференциях и летних школах. Многие из этих студентов, а также наши ассистенты, в частности Рональд ван ден Берг, Анна Кучирайтер, Люси Лай, Дженнифер Лаура Ли, Джули Ли, Хорхе Менендес, Сашанк Писупати, Анне-Лене Сакс, Шан Шен, Бэй Сяо и Хёрмет Йилтиз и члены лаборатории (невозможно перечислить их всех), внесли многочисленные исправления и предложили различные задачи и их решения. Мы благодарим Нувана де Сильва за проверку всех задач в предыдущей версии книги. Мы благодарим читателей черновиков рукописи, в частности Луиджи Ачерби, Роберта Джейкобса, Майкла Лэнди, Зили Лю и Хавьера Трэвера, за предоставление глубоких и полезных отзывов о содержании и изложении; мы также благодарим Роберта и Зили за то, что они были двумя нашими самыми стойкими сторонниками на протяжении многих лет. Мы глубоко благодарны Бреннану Кляйну, постдоку из Северо-восточного университета, за профессиональные иллюстрации, сопровождающие книгу, а также за множество забавных рисунков и за то, что мы стали более ответственными авторами. Эта книга никогда не была бы закончена без помощи 1,3,7-триметилксантина¹. Мы благодарны Роберту Прайору из MIT Press, которому, судя по фамилии², самой судьбой было велено стать нашим редактором. Он неоднократно устанавливал для нас жесткие сроки и каждый раз терпеливо соглашался их продлевать, он сделал возможной бесплатную онлайн-версию. Наконец, мы хотели бы поблагодарить наши семьи, которые оказывали незаслуженно безграничную поддержку все эти годы.

¹ Кофеин. – *Прим. перев.*

² *Prior* (англ.) в байесовской статистике означает априорное распределение, о котором говорится в этой книге. – *Прим. перев.*

Четыре этапа байесовского моделирования

Примеры из глав 3–4

Этап 1: порождающая модель

- Нарисуйте диаграмму**, где каждый узел является переменной, а каждая стрелка – зависимостью. Наблюдения/измерения располагаются внизу.
- Для каждой переменной **напишите уравнение ее вероятностного распределения**. Для каждого наблюдения примите модель шума. Для других переменных получите распределение из экспериментального плана или из естественной статистики. Если есть входящие стрелки, то распределение является условным.

Воздействие



Этап 2: байесовский вывод (правило принятия решения)

- Вычислите апостериорное распределение состояния системы с учетом наблюдений**. Оптимальный наблюдатель делает это, используя распределения в генеративной модели. В качестве альтернативы наблюдатель может предположить другие распределения (например, ошибочных убеждений). Маргинализируйте (усредните) переменные, отличные от наблюдений и целевого состояния системы.
- Определите целевую переменную апостериорного распределения**. Задайте функцию полезности, затем максимизируйте ожидаемую полезность при апостериорной вероятности. (Альтернатива: выборка из апостериорного набора.) Результат: решающее правило (отображение наблюдений в решение). Когда полезность – это точность, целью является максимизация апостериорной вероятности (максимальное апостериорное решающее правило).

$$\mathcal{L}(s; x) = p(x|s)$$

$$p(s|x) \propto \mathcal{L}(s; x)p(s)$$

$$p(s|x) = \mathcal{N}\left(s; \frac{I_s \mu + Ix}{I_s + I}, \frac{1}{I_s + I}\right)$$

$$\hat{s} = \frac{I_s \mu + Ix}{I_s + I}$$

$$I_s \equiv \frac{1}{\sigma_s^2} \text{ и } I \equiv \frac{1}{\sigma^2}$$

Этап 3: вероятность отклика

Для каждого уникального испытания в эксперименте **вычислите вероятность того, что наблюдатель примет каждое возможное решение с учетом воздействий в этом испытании**. Для этого используйте распределение наблюдений с учетом воздействий (из этапа 1) и правила принятия решений (из этапа 2).

- Хороший метод: выборочные наблюдения согласно этапу 1; для каждого применить решающее правило; свести ответы в таблицу.
- Лучший метод: численное интегрирование по наблюдениям.
- Самый лучший (когда это возможно): аналитическое интегрирование по наблюдениям.

Необязательно: добавьте шум или провалы отклика.

$$p(\hat{s}|s) = \int p(\hat{s}|x)p(x|s)dx$$

$$= \mathcal{N}\left(\hat{s}; \frac{I_s \mu + Is}{I_s + I}, \frac{I}{(I_s + I)^2}\right)$$

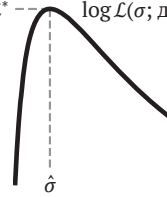
Этап 4: подгонка модели и сравнение моделей

- (a) Вычислите **логарифмическое правдоподобие параметра**, логарифмическую вероятность ответов субъекта во всех испытаниях для гипотетической комбинации параметров.
- (b) **Максимизируйте логарифмическое правдоподобие параметра**. Результат: оценки параметров и максимальное логарифмическое правдоподобие. Проверьте восстановление параметров и сводной статистики с использованием синтетических данных. Используйте более одного алгоритма.
- (c) **Добейтесь соответствия сводной статистике** путем повторного запуска обученной (подобранной) модели.
- (d) **Сформулируйте альтернативные модели** (например, измените этап 2). **Сравните максимальное логарифмическое правдоподобие по моделям**. Исправьте количество параметров (например, используя информационный критерий Акаике). Протестируйте восстановление модели, используя синтетические данные.
- (e) **Проверьте результаты сравнения моделей**, используя сводную статистику.
- (f) Необязательно: **оцените абсолютное соответствие модели**.

$$\log \mathcal{L}(\sigma; \text{данные}) =$$

$$\sum_{i=1}^{n_{\text{trials}}} \log p(\hat{s}_i | s_i, \sigma)$$

$$\log \mathcal{L}^* \quad \text{---} \quad \log \mathcal{L}(\sigma; \text{данные})$$



Сокращения

AIC	Akaike information criterion, информационный критерий Акаике
ANN	artificial neural network, искусственная нейронная сеть
BIC	Bayesian information criterion, байесовский информационный критерий
CCSD	class-conditioned stimulus distribution, обусловленное классом распределение воздействий
CDF	cumulative distribution function, интегральная функция распределения
HMM	hidden Markov model, скрытая марковская модель
LLR	log likelihood ratio, логарифмическое отношение правдоподобия
LPR	log posterior ratio, логарифмическое отношение апостериорных вероятностей
LR	likelihood ratio, отношение правдоподобия
MAP	maximum a posteriori [estimate], максимум апостериорной вероятности [оценка]
MLE	maximum-likelihood estimate, оценка максимального правдоподобия
MSE	mean squared error, среднеквадратическая погрешность
PDF	probability density function, плотность распределения вероятности
PME	posterior mean estimate, оценка апостериорного среднего
PPV	positive predictive value, положительная прогностическая ценность
PSE	point of subjective equality, точка субъективного равенства
RMSE	root-mean-square error, среднеквадратическая ошибка (квадратный корень)
ROC	receiver operating characteristic, рабочая характеристика приемника

Введение

Эта книга представляет собой введение в байесовские модели перцептивного принятия решений и действий. В таких моделях человеческий разум ведет себя как компетентный специалист по данным (или диагностирующий врач, или исследователь места преступления – выберите свою метафору), извлекая смысл из шумных и неоднозначных сенсорных наблюдений. В последние десятилетия байесовский подход к восприятию и действию становится все более популярным, а разработанные модели все шире применяются на практике. Этот подход нам особенно интересен, потому что он строго и успешно объясняет многие экспериментальные данные.

Мы решили написать эту книгу, потому что по сей день не было издано ни одного руководства, которое учило бы с нуля, как рассматривать и строить байесовские модели. Это не означает, что читателям недоступны превосходные материалы по байесовскому моделированию. Тем не менее фундаментальные исследовательские работы, как правило, сосредоточены на узких вопросах и не дают общей картины, в то время как обзорные статьи, как правило, посвящены перечислению достигнутых результатов и не очень полезны начинающему разработчику байесовских моделей. В нашем собственном исследовании мы опирались на академические издания, такие как [45, 94, 181], но они не образуют полную дидактическую последовательность. Наконец, [171] – отличное введение в правило Байеса, но оно не посвящено построению моделей для использования в исследовательской практике.

Эта книга не требует каких-либо предварительных знаний по теории вероятностей, но подразумевает понимание основ математики. Читатели с базовыми знаниями в области исчисления сочтут книгу вполне доступной; но и без знания математики можно понять значительную часть контента. Если это ваш первый опыт математического моделирования – примите наши поздравления! Использование языка математики позволяет избежать двусмысленности, а математические модели дают количественные прогнозы. Мы хорошо понимаем, что математика вызывает у некоторых читателей тревогу, но уверены, что при достаточном усердии материал этой книги будет доступен любому студенту университета, изучающему естественные науки.

Мы рекомендуем читателям уделить время выполнению промежуточных упражнений, встречающихся непосредственно в главах, а также задач в конце каждой главы, представляющих собой смесь базовых идей, математики и моделирования. Однако мы считаем, что полноценного понимания книги невозможно достичь, пока вы не научитесь связывать математические понятия с реальностью.

Восприятие как умозаключение

Фундаментальная проблема восприятия заключается в том, что состояние мира по большей части недоступно напрямую. Глаза, уши, кожа и другие органы чувств регистрируют физические сигналы и преобразуют их в электрические импульсы, направляющиеся к мозгу, – что-то вроде нейронной азбуки Морзе. Глаза регистрируют рисунок света, но сами по себе не идентифицируют зрительную сцену. Кожа ощущает давление и вибрацию, но не идентифицирует внешний объект, вызывающий эти раздражители. Уши обнаруживают звуковые волны, но не указывают на их значение. Таким образом, возникает фундаментальная проблема интерпретации сенсорных наблюдений.

Работа мозга состоит в том, чтобы выяснить основные причины, породившие сенсорные наблюдения, и отреагировать на них соответствующим поведением. Эта работа трудна, потому что сенсорное наблюдение обычно допускает множество возможных интерпретаций. Сенсорное наблюдение может быть низкого качества; например, объекты могут быть плохо освещены, находиться далеко или двигаться быстро. Но даже качественная сенсорная информация не гарантирует наличие единственно возможной интерпретации. Двумерное изображение на сетчатке совместимо с бесконечным числом трехмерных объектов. Хорошо замаскированное животное может быть неотличимо от фона. Слова «торт» и «корт» обычно производят одинаковую слуховую реакцию. Напряжение в мышцах не дает однозначную информацию о позе тела. Возможны множественные интерпретации одних и тех же сенсорных наблюдений.

Получение выводов о состоянии окружающего мира на основе некачественных или неполных наблюдений – это форма умозаключения¹ (*вывод*, *inference*). В частности, *вероятностный вывод*, или *байесовский вывод*, выполняется путем присвоения вероятностей возможным состояниям мира. Теория, изложенная в этой книге, состоит в том, что перцептивное принятие решений является формой байесовского вывода. С помощью байесовского вывода мозг может присваивать вероятности различным состояниям мира, что является решающим шагом при принятии решения о том, какие действия следует предпринять.

В байесовском выводе вероятности назначаются на основе *порождающей модели*² (*generative model*) – статистического описания того, как возникают наблюдения. Это описание включает в себя *исходные убеждения* (*prior belief*), а также убеждения о статистических зависимостях между переменными. Хотя природа сенсорной информации и структура порождающей модели зависят от задачи, сами байесовские вычисления всегда следуют одним и тем же правилам исчисления вероятностей. Таким образом, байесовский подход объединяет огромный диапазон совершенно разрозненного поведения в рамках одной целостной структуры.

¹ Далее для краткости мы будем использовать общепринятый термин «вывод». – Прим. перев.

² Следует отличать от *генеративной модели* в области глубокого обучения. – Прим. перев.

Порождающая модель воплощает предположения, сделанные наблюдателем, и поэтому может принимать различные формы. Особое значение имеет случай, когда порождающая модель является *правильным* статистическим описанием наблюдений. В этом случае байесовский вывод является *оптимальным* в том смысле, что результирующие вероятности, присвоенные состояниям системы, являются компонентом стратегии максимизации *эффективности* (performance) – независимо от того, измеряется ли эффективность с точки зрения точности или иным образом. Другими словами, среди всех стратегий, которые наблюдатель может использовать для задачи восприятия, байесовский вывод, основанный на правильной порождающей модели, является наилучшим вариантом из возможных.

Оптимальный вывод не значит безошибочный. Это всего лишь означает, что сделан наилучший вывод из информации, доступной наблюдателю. Эффективность оптимальной стратегии будет зависеть от информативности сенсорного наблюдения. Иллюзии восприятия часто можно объяснить с точки зрения байесовского вывода. Мы продемонстрируем это на нескольких примерах.

Классификация байесовских моделей

Ученые и философы давно искали способы охарактеризовать логическую структуру моделей. Один особенно полезный подход заключается в классификации моделей по трем типам – описательные, процессные и нормативные.

Описательная модель (descriptive model, иногда называется *моделью «Что»*) представляет собой математическое описание поведенческих переменных (таких как точность или время реакции) в терминах наблюдаемых переменных (таких как интенсивность стимула или количественная характеристика личности). Эти описания могут принимать форму регрессии, обобщенной линейной модели или модели машинного обучения. Утверждение, поддерживаемое подгонкой описательных моделей к данным, заключается в том, что существует модель, которая может соответствовать данным с определенной ошибкой.

Процессная модель (process model, иногда называется *моделью «Как»*) является более амбициозной: она пытается разделить формирование вывода на обобщаемые и психологически или физиологически значимые составляющие процессы, определяя, как наблюдатель/агент принимает решение на основе доступной информации. Примерами таких процессов могут быть «добавление гауссова шума измерения к переменной воздействия» или «наблюдатель сопоставляет переменную решения с решением, применяя критерий». Цель процессной модели – не просто согласовать данные, но и зафиксировать внутренние представления и конкретный поток информации. В данном случае утверждение относительно подогнанной к данным модели заключается в том, что если процесс соответствует гипотезе, то результатом будет измеренное поведение.

Нормативная модель (normative model, иногда называется *моделью «Почему»*) амбициозна по-другому. В нормативной модели мы спрашиваем, поче-

му поведение такое, какое оно есть. Точнее говоря, мы спрашиваем, почему в конкретной экологической нише полезны определенные виды поведения. Например, нормативная модель может (явно или неявно) предполагать, что в нашей жизни важна максимальная точность. Такая модель может затем делать предположения о мире, которые рассматриваются как неизменные (например, что в зрении присутствует фиксированное количество шума). Затем модель может найти оптимальное решение проблемы, которое можно сравнить с реальным поведением. Утверждение относительно этой подогнанной к данным модели заключается в том, что если мир или экологическая ниша соответствует гипотезе, то измеренное поведение будет полезным.

Байесовские модели обычно используются в качестве нормативных моделей: мы указываем цель, которую агент хочет оптимизировать, предположения, которые агент имеет о своем мире, и типы наблюдений, которые доступны агенту. Затем байесовский подход позволяет нам обнаружить поведение, которое оптимизирует цель агента. Для байесовской модели с определенными наблюдениями, априорными знаниями и целями существует одно четко определенное оптимальное действие, и в этой книге объясняется, как его вычислить. В этом смысле байесовская модель является прототипом нормативной модели.

Байесовские модели также часто используются в качестве процессных моделей. С одной стороны, байесовские модели, очевидно, являются процессными – они делят путь от воздействия к реакции на два этапа, принимая точку зрения наблюдателя. На этапе «кодирования», или «представления», они определяют характер данных, с которыми работает наблюдатель и которые могут отличаться от воздействия – мы не воспринимаем состояния мира напрямую, а получаем от него сенсорные наблюдения. На втором этапе происходит принятие решения, когда данные, доступные наблюдателю, преобразуются в действия. Эта двухэтапная последовательность составляет процессную модель. С другой стороны, однако, процессная природа байесовских моделей менее ясна. Дело в том, что байесовская процедура на этапе принятия решения выполняется путем вычисления функций правдоподобия и апостериорного распределения и создания действия на основе последнего. Исследователи расходятся во мнениях относительно того, являются ли эти шаги и связанные с ними конструкции психологически значимыми или это просто способ вывода нормативного правила принятия решений. В зависимости от позиции этап принятия решения в байесовской модели может считаться моделью процесса, а может и не считаться. Мы обсудим это далее в главе 15.

Байесовские модели также можно расположить на условной поведенческо-нейронной оси. Поскольку соображения оптимальности относятся к функционированию организма в целом, байесовские модели мозга в первую очередь являются поведенческими моделями. Если байесовская модель успешно описывает поведение в конкретной задаче, мы можем надеяться ограничить наше понимание основных нейронных процессов и разработать нейронные модели или модели уровня реализации. Мы коснемся этого в главе 14. Данный подход принципиально отличается от подхода «снизу вверх», при котором начинают с моделирования цепей нейронов вплоть до биологических

деталей, а затем пытаются построить модели, комбинируя несколько цепей.

Последним релевантным критерием является условная ось восприятия-познания. Хотя эта книга посвящена восприятию и действию, байесовские модели широко используются в других областях когнитивной науки и психологии. В частности, существует богатая история байесовских моделей высокоуровневого познания [34], восходящая, по крайней мере, к работам Джонатана Эванса [48] и Джона Андерсона [11], которые внесли большой вклад в понимание когнитивного развития (например, [68, 210]). Познание более высокого уровня время от времени фигурирует в этой книге, особенно в главах 6, 12 и 13. В главе 15 мы комментируем различия между восприятием и познанием.

О чем эта книга *не* рассказывает

Байесовские модели восприятия и действия моделируют наблюдателей, которые делают выводы о состоянии мира на основе сенсорных наблюдений. В этом заключается принципиальное отличие от *байесовского статистического анализа данных*, в котором экспериментатор делает вывод о значении параметра модели на основе собранных данных. Математический аппарат в основном совпадает, но в этой книге мы сосредоточимся на том, как мозг воспринимает или принимает решения, а не на том, как анализируются данные. При этом байесовские модели, как и все модели, имеют параметры, значения которых необходимо вывести (или обучить). По этой причине мы включили в книгу приложение С, посвященное подбору и сравнению моделей; это приложение, однако, по-прежнему не охватывает байесовские методы анализа данных. Читателям, интересующимся анализом данных, мы рекомендуем [59, 103, 164].

Заявление об ограниченном цитировании

В этой книге мы опираемся на работы многих замечательных ученых, но по очевидным причинам не можем упомянуть или перечислить их все. Отсутствие цитаты или упоминания в данной книге ни в коем случае не умаляет заслуги отдельных ученых и достоинства их работ. Мы в равной степени признательны всем исследователям.

В заключение

Мы надеемся, что вам понравится книга, и будем рады вашим отзывам на веб-сайте книги www.bayesianmodeling.com. Дополнительный материал, включая решения задач, интерактивные демонстрации и дополнительное чтение, будет размещен там в будущем.

Глава 1

Неопределенность И ВЫВОД

Как мы трансформируем наши сенсорные наблюдения в убеждения о состоянии мира?

Всякий раз, когда мы что-то воспринимаем, делаем предсказание или обдумываем решение, мы рассуждаем, опираясь на вероятности, даже если не осознаем этого. Мы используем имеющуюся у нас информацию, чтобы сделать вывод или оценить что-то еще, что нас интересует. Имеющаяся у нас информация обычно неполная или зашумленная, поэтому наш вывод не является точным. Например, если мы наблюдаем блестящий пол (имеющаяся информация), это говорит о том, что он *может быть* мокрым (предмет нашего интереса). Используя доступную сенсорную информацию и любые соответствующие знания, которые у нас могут быть, мы должны определить вероятность каждой интерпретации (мокрой или сухой). Как нам удастся выносить здравые суждения в таких ситуациях?

Краткое содержание главы

Мы обрисовываем в общих чертах процесс перцептивного вывода, подчеркивая неопределенность, присущую восприятию. Используя простые примеры, мы вводим вероятности, связанные с перцептивным выводом, правдоподобием, априорным и апостериорным распределением вероятностей, сосредоточив внимание на лежащем в их основе *неопосредованном мышлении*¹. Затем мы проиллюстрируем вездесущность перцептивного вывода в повседневной жизни серией примеров, связанных со зрительным и слуховым восприятием. В этой главе мы не используем математику, исследуя каждый пример только качественно и графически. Наша цель – дать интуитивное понимание процесса перцептивного вывода, которое послужит основой для более строгой математической формулировки в следующих главах.

¹ Intuitions, мышление, не опосредованное сознанием. – Прим. перев.

1.1. Цель восприятия

Этот вопрос разрешить единственно разум обязан;
Глаз же природу вещей познавать совершенно не может,
А потому не вини его в том, в чем повинен лишь разум¹.

– Лукреций, *De rerum natura* [111]

Люди, как и другие живые существа, наделены набором тонких органов чувств, с помощью которых они обнаруживают свойства окружающей среды. Органы чувств реагируют на такие разнообразные свойства и явления окружающей среды, как свет (глаза), звук (уши), температура (кожа), текстура материала (кожа), химический состав (нос, язык) и положение тела (суставные и мышечные рецепторы, вестибулярные органы). Наши органы чувств составляют неотъемлемую часть нас самих, настолько, что мы обычно воспринимаем их присутствие как должное. Чтобы оценить роль, которую играют наши чувства, попробуйте представить себе жизнь без зрения, слуха, осязания, обоняния или вкуса.

Какими бы сложными ни были органы чувств, их активация физическими раздражителями – это только первый шаг в восприятии. В повседневной жизни нас не волнует длина волны (цвет) и интенсивность (яркость) попадающих в наши глаза световых волн или структура изменяющейся по амплитуде и времени акустической энергии, поступающей в наши уши. Мы заботимся не об этих паттернах сенсорной активации как таковых, которые мы называем *сенсорными входами*, или *наблюдениями*, а об их интерпретации. На самом деле качество нашей жизни – а часто и сама жизнь – зависит от нашей способности придумывать правильные интерпретации. Отражает ли этот рисунок света лицо друга? Является ли эта акустическая волна звуком ветра, воем собаки или голосом нашего спутника? Короче говоря, наш интерес заключается не в сенсорной информации как таковой, а в тех знаниях, которые она предоставляет о соответствующих состояниях мира² [1].

Сделать интерпретативный переход от *ощущения* (активация органов чувств) к *восприятию* (вывод о состоянии мира) – сложная задача. Вообще говоря, эта книга о том, как мозг может оптимально выполнять эту задачу. Большой и быстро растущий объем экспериментальных и теоретических работ показывает, что восприятие – это, по крайней мере неявно, процесс вероятностного вывода, в котором организм пытается сделать вывод о наиболее вероятном состоянии мира, используя сенсорные данные и все имеющиеся в его распоряжении соответствующие знания. Как гласит цитата Лукреция, восприятие иногда идет не тем путем, но ошибки часто можно рассматривать как побочные продукты разумной стратегии вывода.

¹ Перевод: Ф. А. Петровский (1936).

² Некоторые исследователи восприятия называют состояние мира *дистальным стимулом* (distal stimulus), а наблюдение – *проксимальным стимулом* (proximal stimulus).