

Содержание

| | |
|--|----|
| От издательства | 14 |
| Предисловие | 15 |
| Благодарности | 16 |
| 1 Введение | 17 |
| 1.1. Принятие решений | 17 |
| 1.2. Области применения | 18 |
| 1.2.1. Предотвращение столкновения самолетов | 19 |
| 1.2.2. Автоматизированное вождение | 19 |
| 1.2.3. Скрининг рака молочной железы | 19 |
| 1.2.4. Доля инвестиций и распределение портфеля | 20 |
| 1.2.5. Распределенное наблюдение за лесными пожарами | 20 |
| 1.2.6. Исследование Марса | 21 |
| 1.3. Методы создания агентов | 21 |
| 1.3.1. Явное программирование | 22 |
| 1.3.2. Обучение с учителем | 22 |
| 1.3.3. Оптимизация | 22 |
| 1.3.4. Планирование | 22 |
| 1.3.5. Обучение с подкреплением | 23 |
| 1.4. История автоматизации принятия решений | 23 |
| 1.4.1. Экономика | 24 |
| 1.4.2. Психология | 25 |
| 1.4.3. Нейробиология | 25 |
| 1.4.4. Информатика | 26 |
| 1.4.5. Инженерия | 26 |
| 1.4.6. Математика | 27 |
| 1.4.7. Исследование операций | 28 |
| 1.5. Воздействие на общество | 28 |

| | |
|---|----|
| 1.6. Краткий обзор содержания книги | 30 |
| 1.6.1. Вероятностное рассуждение | 30 |
| 1.6.2. Многостадийные задачи..... | 30 |
| 1.6.3. Неопределенность модели | 31 |
| 1.6.4. Неопределенность состояния..... | 31 |
| 1.6.5. Мультиагентные системы..... | 32 |

Часть I. Вероятностные рассуждения.....33

2 Формальное представление неопределенности34

| | |
|---|----|
| 2.1. Степени доверия и вероятности | 34 |
| 2.2. Распределения вероятностей..... | 35 |
| 2.2.1. Дискретные распределения вероятностей..... | 35 |
| 2.2.2. Непрерывные распределения вероятностей | 36 |
| 2.3. Совместные распределения | 41 |
| 2.3.1. Дискретные совместные распределения..... | 41 |
| 2.3.2. Непрерывное совместное распределение..... | 44 |
| 2.4. Условные распределения | 47 |
| 2.4.1. Дискретные модели условных распределений | 48 |
| 2.4.2. Условные модели Гаусса..... | 49 |
| 2.4.3. Линейные модели Гаусса | 49 |
| 2.4.4. Условные линейные модели Гаусса..... | 50 |
| 2.4.5. Сигмовидные модели | 50 |
| 2.4.6. Детерминированные переменные | 51 |
| 2.5. Байесовские сети..... | 51 |
| 2.6. Условная независимость..... | 54 |
| 2.7. Заключение | 57 |
| 2.8. Упражнения | 57 |

3 Вероятностный вывод62

| | |
|---|----|
| 3.1. Вывод в байесовских сетях | 62 |
| 3.2. Вывод в наивных байесовских моделях | 67 |
| 3.3. Исключение переменной суммированием-перемножением..... | 70 |
| 3.4. Распространение доверия | 72 |
| 3.5. Вычислительная сложность..... | 72 |
| 3.6. Прямая выборка | 73 |
| 3.7. Выборка, взвешенная по правдоподобию | 76 |
| 3.8. Выборка Гиббса | 79 |
| 3.9. Вывод в гауссовых моделях..... | 81 |
| 3.10. Заключение..... | 83 |
| 3.11. Упражнения | 84 |

4 Параметрическое обучение90

| | |
|--|----|
| 4.1. Обучение по критерию максимального правдоподобия..... | 90 |
|--|----|

| | |
|---|------------|
| 4.1.1. Оценки максимального правдоподобия для категориальных распределений..... | 91 |
| 4.1.2. Оценки максимального правдоподобия для распределений Гаусса | 92 |
| 4.1.3. Оценки максимального правдоподобия для байесовских сетей..... | 93 |
| 4.2. Байесовское параметрическое обучение | 96 |
| 4.2.1. Байесовское обучение для бинарных распределений..... | 97 |
| 4.2.2. Байесовское обучение для категориальных распределений | 99 |
| 4.2.3. Байесовское обучение для байесовских сетей | 100 |
| 4.3. Непараметрическое обучение..... | 101 |
| 4.4. Обучение с отсутствующими данными | 103 |
| 4.4.1. Подстановка данных..... | 104 |
| 4.4.2. Алгоритм ожидания-максимизации | 107 |
| 4.5. Заключение..... | 109 |
| 4.6. Упражнения | 110 |
| 5 Структурное обучение | 116 |
| 5.1. Оценка байесовской сети | 116 |
| 5.2. Поиск ориентированного графа | 119 |
| 5.3. Марковские классы эквивалентности | 123 |
| 5.4. Поиск частично ориентированного графа | 124 |
| 5.5. Заключение..... | 126 |
| 5.6. Упражнения | 126 |
| 6 Простые решения..... | 129 |
| 6.1. Ограничения рациональных предпочтений..... | 129 |
| 6.2. Функции полезности..... | 131 |
| 6.3. Выявление полезности | 132 |
| 6.4. Принцип максимальной ожидаемой полезности..... | 134 |
| 6.5. Сети принятия решений..... | 136 |
| 6.6. Полезность информации..... | 139 |
| 6.7. Иррациональность | 141 |
| 6.8. Заключение..... | 143 |
| 6.9. Упражнения | 143 |
| Часть II. Задачи последовательного принятия решений..... | 148 |
| 7 Методы точного решения..... | 149 |
| 7.1. Марковские процессы принятия решений..... | 149 |
| 7.2. Оценка стратегии | 153 |
| 7.3. Нахождение стратегии через функцию полезности | 156 |
| 7.4. Итерация по стратегиям | 157 |
| 7.5. Итерация по критерию | 159 |

| | |
|--|-----|
| 7.6. Асинхронная итерация по критерию..... | 162 |
| 7.7. Представление задачи в виде линейной программы | 164 |
| 7.8. Линейные системы с квадратичным вознаграждением | 166 |
| 7.9. Заключение | 170 |
| 7.10. Упражнения..... | 171 |

8 Приближенное вычисление функции

| | |
|---|-----|
| полезности | 179 |
| 8.1. Параметрические представления | 179 |
| 8.2. Аппроксимация по ближайшему соседу | 181 |
| 8.3. Ядерное сглаживание..... | 183 |
| 8.4. Линейная интерполяция | 185 |
| 8.5. Симплексная интерполяция | 188 |
| 8.6. Линейная регрессия..... | 191 |
| 8.7. Регрессия на основе нейронной сети..... | 195 |
| 8.8. Заключение..... | 196 |
| 8.9. Упражнения | 196 |

9 Онлайн-планирование 201 |

| | |
|--|-----|
| 9.1. Планирование с отступающим горизонтом..... | 201 |
| 9.2. Стратегия развертывания..... | 203 |
| 9.3. Прямой поиск | 204 |
| 9.4. Метод ветвей и границ | 206 |
| 9.5. Разреженная выборка | 207 |
| 9.6. Поиск по дереву Монте-Карло | 209 |
| 9.7. Эвристический поиск..... | 218 |
| 9.8. Эвристический поиск с разметкой | 219 |
| 9.9. Планирование с открытым контуром | 224 |
| 9.9.1. Прогнозирующее управление с детерминированной моделью | 226 |
| 9.9.2. Робастное прогностическое управление | 228 |
| 9.9.3. Многовариантное прогностическое управление..... | 229 |
| 9.10. Заключение..... | 231 |
| 9.11. Упражнения | 231 |

10 Поиск стратегии 236 |

| | |
|---|-----|
| 10.1. Приблизительная оценка стратегии..... | 236 |
| 10.2. Локальный поиск | 238 |
| 10.3. Генетические алгоритмы | 241 |
| 10.4. Метод перекрестной энтропии | 242 |
| 10.5. Эволюционные стратегии | 244 |
| 10.6. Изотропные эволюционные стратегии | 248 |
| 10.7. Заключение | 250 |
| 10.8. Упражнения | 251 |

| | | |
|---|--|-----|
| 11 | Нахождение градиента стратегии | 255 |
| 11.1. | Конечная разность | 255 |
| 11.2. | Градиент регрессии | 258 |
| 11.3. | Отношение правдоподобия..... | 260 |
| 11.4. | Предстоящее вознаграждение | 263 |
| 11.5. | Вычитание базисного значения..... | 266 |
| 11.6. | Заключение..... | 270 |
| 11.7. | Упражнения..... | 270 |
| 12 | Оптимизация методом градиентного спуска по стратегиям | 273 |
| 12.1. | Обновление стратегии методом градиентного подъема | 273 |
| 12.2. | Ограниченное обновление градиента | 275 |
| 12.3. | Метод натурального градиента..... | 277 |
| 12.4. | Метод поиска в доверительной области..... | 280 |
| 12.5. | Зажатие замещенной цели | 285 |
| 12.6. | Заключение..... | 288 |
| 12.7. | Упражнения..... | 289 |
| 13 | Методы «актор–критик» | 292 |
| 13.1. | Определение актора и критика..... | 292 |
| 13.2. | Обобщенная оценка преимуществ | 294 |
| 13.3. | Градиент детерминированной стратегии | 298 |
| 13.4. | Метод «актор–критик» с поиском по дереву Монте-Карло | 301 |
| 13.5. | Заключение..... | 303 |
| 13.6. | Упражнения | 304 |
| 14 | Проверка стратегии | 306 |
| 14.1. | Оценка показателей качества стратегии..... | 306 |
| 14.2. | Моделирование редких событий | 312 |
| 14.3. | Анализ робастности системы | 315 |
| 14.4. | Анализ компромиссов | 317 |
| 14.5. | Состязательный анализ | 319 |
| 14.6. | Заключение..... | 322 |
| 14.7. | Упражнения..... | 322 |
| Часть III. Неопределенность модели | | 325 |
| 15 | Исследование среды и использование знаний | 326 |
| 15.1. | Задача однорукого бандита..... | 326 |
| 15.2. | Оценка байесовской модели | 328 |
| 15.3. | Стратегии ненаправленного исследования | 330 |
| 15.4. | Стратегии направленного исследования | 332 |

| | |
|--|------------|
| 15.5. Оптимальные стратегии исследования | 336 |
| 15.6. Исследование с несколькими состояниями | 338 |
| 15.7. Заключение | 338 |
| 15.8. Упражнения | 339 |
| 16 Методы на основе моделей | 343 |
| 16.1. Модели максимального правдоподобия | 343 |
| 16.2. Схемы обновления модели | 346 |
| 16.2.1. Полное обновление | 346 |
| 16.2.2. Рандомизированное обновление | 347 |
| 16.2.3. Приоритетный механизм обновления | 347 |
| 16.3. Исследование | 349 |
| 16.4. Байесовские методы | 352 |
| 16.5. Адаптивные по Байесу марковские процессы принятия решений | 355 |
| 16.6. Апостериорная выборка | 356 |
| 16.7. Заключение | 358 |
| 16.8. Упражнения | 359 |
| 17 Свободные методы обучения с подкреплением | 362 |
| 17.1. Инкрементное вычисление среднего значения распределения | 362 |
| 17.2. Q-обучение | 365 |
| 17.3. Алгоритм SARSA | 367 |
| 17.4. Следы приемлемости | 369 |
| 17.5. Формирование вознаграждения | 371 |
| 17.6. Аппроксимация функции полезности действия | 371 |
| 17.7. Воспроизведение опыта | 375 |
| 17.8. Заключение | 378 |
| 17.9. Упражнения | 378 |
| 18. Имитационное обучение | 383 |
| 18.1. Поведенческое копирование | 383 |
| 18.2. Агрегация наборов данных | 386 |
| 18.3. Итеративное обучение путем стохастического смешивания | 389 |
| 18.4. Обратное обучение с подкреплением с максимальной разницей | 392 |
| 18.5. Обратное обучение с подкреплением с максимальной энтропией | 396 |
| 18.6. Генеративно-сопоставительное имитационное обучение | 399 |
| 18.7. Заключение | 400 |
| 18.8. Упражнения | 400 |
| Часть IV. Неопределенность состояния | 405 |
| 19 Убеждения | 406 |
| 19.1. Начальные убеждения | 406 |
| 19.2. Фильтр дискретных состояний | 407 |

| | |
|---|-----|
| 19.3. Фильтр Калмана | 412 |
| 19.4. Расширенный фильтр Калмана..... | 414 |
| 19.5. Сигма-точечный фильтр Калмана | 415 |
| 19.6. Парциальный фильтр..... | 418 |
| 19.7. Внесение частиц | 422 |
| 19.8. Заключение..... | 425 |
| 19.9. Упражнения | 426 |

20 Точное планирование с использованием убеждений-состояний.....

| | |
|-------------------------------------|-----|
| 20.1. MDP убеждений-состояний | 436 |
| 20.2. Условные планы..... | 437 |
| 20.3. Альфа-векторы | 441 |
| 20.4. Сокращение | 444 |
| 20.5. Итерация по полезности..... | 447 |
| 20.6. Линейные стратегии | 449 |
| 20.7. Заключение | 451 |
| 20.8. Упражнения | 451 |

21 Офлайн-планирование с использованием убеждений-состояний.....

| | |
|---|-----|
| 21.1. Аппроксимация полностью наблюдаемой полезности..... | 455 |
| 21.2. Метод быстрой инфограницы | 458 |
| 21.3. Методы быстрой оценки снизу | 459 |
| 21.4. Точечная итерация по полезности..... | 461 |
| 21.5. Рандомизированная точечная итерация по полезности | 464 |
| 21.6. Пилообразная оценка сверху | 465 |
| 21.7. Выбор точек в наборе убеждений..... | 469 |
| 21.8. Пилообразный эвристический поиск | 472 |
| 21.9. Триангулированные функции полезности..... | 474 |
| 21.10. Заключение..... | 477 |
| 21.11. Упражнения | 478 |

22 Онлайн-планирование с использованием убеждений-состояний.....

| | |
|---|-----|
| 22.1. Предпросмотр с развертываниями..... | 483 |
| 22.2. Прямой поиск | 483 |
| 22.3. Метод ветвей и границ | 486 |
| 22.4. Разреженная выборка | 486 |
| 22.5. Поиск по дереву Монте-Карло | 487 |
| 22.6. Поиск по детерминированному разреженному дереву | 490 |
| 22.7. Эвристический поиск на основе разности границ..... | 494 |
| 22.8. Заключение..... | 496 |

| | |
|------------------------|-----|
| 22.9. Упражнения | 497 |
|------------------------|-----|

23 Понятие контроллера..... 500

| | |
|--|-----|
| 23.1. Контроллеры..... | 500 |
| 23.2. Итерация по стратегиям..... | 504 |
| 23.3. Нелинейное программирование..... | 509 |
| 23.4. Градиентный подъем..... | 512 |
| 23.5. Заключение..... | 518 |
| 23.6. Упражнения | 519 |

Часть V. Многоагентные системы..... 521

24 Логический вывод в многоагентных системах..... 522

| | |
|---|-----|
| 24.1. Простые игры | 522 |
| 24.2. Модели откликов..... | 525 |
| 24.2.1. Наилучший отклик..... | 526 |
| 24.2.2. Отклик softmax..... | 526 |
| 24.3. Равновесие доминирующей стратегии | 527 |
| 24.4. Равновесие Нэша..... | 528 |
| 24.5. Согласованное равновесие | 530 |
| 24.6. Итеративный поиск лучшего отклика..... | 533 |
| 24.7. Иерархическая форма модели softmax..... | 534 |
| 24.8. Фиктивная игра..... | 536 |
| 24.9. Градиентный подъем..... | 539 |
| 24.10. Заключение..... | 542 |
| 24.11. Упражнения | 542 |

25 Последовательные задачи..... 548

| | |
|--|-----|
| 25.1. Марковские игры..... | 548 |
| 25.2. Модели отклика..... | 550 |
| 25.2.1. Наилучший отклик..... | 551 |
| 25.2.2. Стратегия отклика softmax..... | 551 |
| 25.3. Равновесие Нэша..... | 552 |
| 25.4. Фиктивная марковская игра..... | 553 |
| 25.5. Градиентный подъем..... | 557 |
| 25.6. Q-обучение Нэша | 559 |
| 25.7. Заключение..... | 561 |
| 25.8. Упражнения | 561 |

26 Неопределенность состояния..... 564

| | |
|---|-----|
| 26.1. Частично наблюдаемые марковские игры..... | 564 |
| 26.2. Оценка стратегии..... | 566 |
| 26.2.1. Оценка условных планов..... | 566 |
| 26.2.2. Оценка стохастических контроллеров..... | 568 |

| | |
|---|------------|
| 26.3. Равновесие Нэша..... | 569 |
| 26.4. Динамическое программирование..... | 571 |
| 26.5. Заключение..... | 574 |
| 26.6. Упражнения | 575 |
| 27 Совместные действия агентов..... | 577 |
| 27.1. Децентрализованные частично наблюдаемые марковские процессы принятия решений | 577 |
| 27.2. Подклассы | 578 |
| 27.3. Динамическое программирование | 582 |
| 27.4. Итерация по наилучшим откликам | 582 |
| 27.5. Эвристический поиск..... | 584 |
| 27.6. Нелинейное программирование | 587 |
| 27.7. Заключение | 588 |
| 27.8. Упражнения..... | 589 |
| A Основные математические понятия..... | 592 |
| B Распределения вероятностей | 604 |
| C Вычислительная сложность | 606 |
| D Представление функций в форме нейронных сетей..... | 610 |
| E Алгоритмы поиска | 628 |
| F Задачи принятия решений..... | 637 |
| G Язык программирования Julia..... | 655 |
| Предметный указатель..... | 677 |

От издательства

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.

Предисловие

Эта книга представляет собой развернутый обзор, посвященный алгоритмам принятия решений в условиях неопределенности. Мы постарались охватить максимально широкий спектр тем, связанных с принятием решений, и познакомить читателей с формулировками основных математических задач и алгоритмами их решения. В книге приведены рисунки, примеры и упражнения, которые помогают раскрыть идеи, лежащие в основе различных подходов.

Книга предназначена для студентов старших курсов и аспирантов, а также специалистов в области искусственного интеллекта и систем принятия решений. Чтение этой книги требует определенного уровня математической подготовки и предполагает предварительное знакомство с многомерным исчислением, линейной алгеброй и теорией вероятностей. Некоторые вводные материалы представлены в приложениях. К дисциплинам, в которых книга будет особенно полезна, относятся математика, статистика, информатика, аэрокосмическая промышленность, электротехника и операционные процессы.

В основе этого учебника лежат алгоритмы, реализованные на языке программирования Julia. Мы обнаружили, что этот язык идеально подходит для описания алгоритмов в удобочитаемой форме. Приоритетом при разработке алгоритмических реализаций была простота понимания, а не эффективность кода. Для использования на практике могут понадобиться более оптимальные реализации. Разрешается бесплатно использовать любые фрагменты кода из этой книги с обязательным указанием ссылки на источник кода.

Микель Кохендерфер

Тим Уилер

Кайл Рэй

Стэнфорд, Калифорния,

28 февраля 2022 г.

Благодарности

Этот учебник написан на основе университетского курса «Принятие решений в условиях неопределенности», преподаваемого в Стэнфорде. Мы благодарны студентам и ассистентам преподавателей, которые помогли формировать курс в течение последних шести лет.

Авторы хотели бы поблагодарить многих людей, поделившихся ценными отзывами о ранних черновиках нашей рукописи, включая Дилана Асмара, Дрю Бэгнелла, Сафу Бахши, Эдварда Балабана, Джина Беттертона, Раунака Бхаттачарию, Келси Бинга, Максима Бутона, Остина Чана, Саймона Шовина, Шушмана Чоудхури, Джона Кокса, Мэттью Дейли, Викторию Дакс, Ричарда Дьюи, Деа Дрессель, Бена Дюпри, Торстейна Элиассена, Йоханнеса Фишера, Рушила Горадия, Джайеша Гупту, Арека Джамгочяна, Рохана Капре, Марка Корена, Лиамы Круза, Тора Латтимора, Бернарда Ланге, Ричи Ли, Шэна Ли, Майкла Литтмана, Роберта Мосса, Джошуа Отта, Ориану Пельтцер, Франческо Пикколи, Джефффри Сарнофф, Марка Шлихтинга, Рансалу Сенанаяке, Челси Сидран, Криса Стронга, Зака Санберга, Абия Тешоме, Александроса Цикаса, Кемаля Юра, Джоша Вольфа, Анилы Йылдыза и Цзунчжана Чжана. Мы также хотели бы поблагодарить Сиднея Каца, Кунала Менду и Аяна Мукхопадхья за их вклад в обсуждение главы 1. Росс Александер разработал многие из упражнений, включенных в книгу. При подготовке этой рукописи к публикации нам было приятно работать с Элизабет Суэйзи из MIT Press.

При работе над этим учебником мы использовали различные пакеты программ с открытым исходным кодом (см. приложение G).

1 Введение

В нашей повседневной жизни многие важные задачи связаны с принятием решений в условиях неопределенности. К ним относятся, например, предотвращение столкновений самолетов, борьба с лесными пожарами и реагирование на стихийные бедствия. При разработке автоматизированных систем принятия решений или систем поддержки принятия решений важно учитывать различные источники неопределенности, тщательно соблюдая баланс между несколькими целями. Мы обсудим эти проблемы с вычислительной точки зрения, стремясь раскрыть теорию, лежащую в основе моделей принятия решений и вычислительных подходов. В этой главе представлена проблема принятия решений в условиях неопределенности, приведены некоторые примеры применения и очерчено пространство вычислительных подходов. Затем мы покажем, как различные дисциплины способствуют нашему пониманию разумного принятия решений, и определим области потенциального взаимодействия систем принятия решений и общества. Главу завершает краткий анонс оставшейся части книги.

1.1. Принятие решений

Агент – это сущность, которая действует на основе наблюдений за окружающей его средой. Агенты могут быть физическими объектами, такими как люди или роботы, или невещественными объектами, такими как системы поддержки принятия решений, которые полностью реализованы программно. Как показано на рис. 1.1, взаимодействие между агентом и окружающей средой следует циклу «наблюдение-действие», который иногда называют *петлей*.

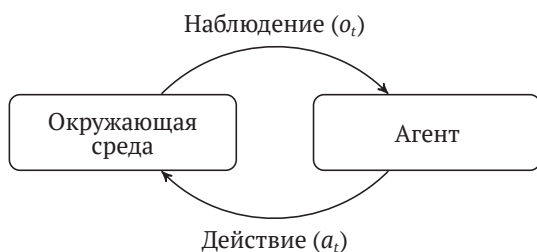


Рис. 1.1. Взаимодействие между агентом и его окружением

Агент в момент времени t получает *наблюдение* за окружающей средой, обозначенное как o_t . Наблюдения могут производиться, например, посредством биологических органов чувств, как у людей, или с помощью электронной сенсорной системы, такой как радар в системе управления воздушным движением. Наблюдения часто бывают неполными или зашумленными; например, люди могут не различить приближающийся самолет среди других отметок на экране, или радиолокационная система может пропустить обнаружение из-за электромагнитных помех. Затем агент выбирает *действие* a_t в процессе принятия решения.

Действие агента, такое как подача сигнала тревоги, может оказывать недетерминированное влияние на среду.

Мы будем говорить об агентах, которые разумно взаимодействуют для достижения своих целей на протяжении определенного времени. Учитывая прошлую последовательность наблюдений o_1, \dots, o_t и знание среды, агент должен выбрать действие, которое лучше всего способствует достижению его цели при наличии различных источников неопределенности¹, включая следующие:

- *неопределенность результата*, когда последствия наших действий не определены;
- *неопределенность модели*, когда наша модель задачи не определена;
- *неопределенность состояния*, когда истинное состояние окружающей среды остается неопределенным;
- *неопределенность взаимодействия*, когда поведение других агентов, взаимодействующих в окружающей среде, не определено.

Эта книга построена вокруг упомянутых четырех источников неопределенности. Принятие решений в условиях неопределенности занимает центральное место в области *искусственного интеллекта*², а также во многих других областях, как указано в разделе 1.4. Мы обсудим различные алгоритмы или описания вычислительных процессов для принятия решений, устойчивых к неопределенности.

1.2. Области применения

Обобщенную систему принятия решений, упомянутую в предыдущем разделе, можно применить к широкому кругу задач. В этом разделе мы рассмотрим несколько концептуальных примеров реального применения. Приложение F содержит описание других гипотетических задач, которые применяются в этой книге для наглядной демонстрации обсуждаемых алгоритмов.

¹ Здесь мы сосредоточимся на задачах дискретного времени. Задачи с непрерывным временем изучаются в области теории управления. См. D. E. Kirk, *Optimal Control Theory: An Introduction*. Prentice-Hall, 1970.

² Всестороннее введение в искусственный интеллект предоставлено в S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson, 2021.

1.2.1. Предотвращение столкновения самолетов

Чтобы помочь авиационным диспетчерам предотвратить столкновение самолетов в воздухе, мы должны разработать систему, которая способна предупреждать пилотов о потенциальных угрозах и указывать им, какие маневры следует совершить³. Система взаимодействует с транспондерами самолетов, чтобы определить их положение с определенной степенью точности. Решить, какие указания дать пилотам, очень непросто. Существует неопределенность в том, как быстро пилоты отреагируют и насколько усердно они будут выполнять указания. Кроме того, существует неопределенность в поведении других самолетов. Необходимо, чтобы наша система выдавала предупреждения достаточно рано, оставляя пилотам запас времени для маневра, чтобы избежать столкновений, но, с другой стороны, система не должна выдавать предупреждения слишком рано, так как это приведет ко множеству ненужных маневров, которые сами по себе служат источником дополнительной опасности и неопределенности. Поскольку эта система будет использоваться постоянно во всем мире, нам требуется решение, обеспечивающее исключительно высокий уровень безопасности.

1.2.2. Автоматизированное вождение

Инженеры стремятся создать автономный автомобиль, который сможет безопасно передвигаться в обычных городских условиях⁴. Автономное транспортное средство должно принимать безопасные решения, полагаясь на набор собственных датчиков для восприятия окружающей среды. Одним из таких датчиков является *лидар*, который использует отражение лазерного луча для определения расстояний до препятствий и построения объемной картины окружающей обстановки. Другой тип датчика – это камера, которая с помощью алгоритмов компьютерного зрения может обнаруживать пешеходов и другие транспортные средства. Оба этих датчика несовершенны и чувствительны к шуму и помехам. Например, припаркованный грузовик может загораживать пешехода, собравшегося перейти дорогу по пешеходному переходу. Наша система должна предсказывать намерения и ожидаемые траектории движения других транспортных средств, пешеходов и прочих участников дорожного движения на основе их наблюдаемого поведения, чтобы автомобиль безопасно добрался до пункта назначения.

1.2.3. Скрининг рака молочной железы

Во всем мире рак молочной железы является наиболее распространенным видом рака у женщин. Раннее обнаружение рака молочной железы может помочь

³ Это применение обсуждается в главе, написанной М. Кохендерфером для книги *Decision Making Under Uncertainty: Theory and Application*. MIT Press, 2015.

⁴ Аналогичное применение представлено в докладе М. Bouton, А. Nakhaei, К. Fujimura, М. J. Kochenderfer, *Safe Reinforcement Learning with Scene Decomposition for Navigating Complex Urban Environments*, in IEEE Intelligent Vehicles Symposium (IV), 2019.

спасти жизни, а маммография (рентгеновское сканирование) является наиболее эффективным доступным методом скрининга. Тем не менее маммография сопряжена с потенциальными рисками, включая ложноположительные результаты, которые могут привести к ненужному и инвазивному последующему диагностическому обследованию. Многолетние исследования ученых из разных стран позволили выработать критерии и методики скрининга населения в зависимости от возраста, позволяющие найти компромисс между пользой и рисками от обследования. Наличие системы, которая может давать рекомендации на основе личных характеристик риска и истории скрининга, может привести к существенному улучшению здоровья определенных групп населения⁵. Успех такой системы можно оценить в сравнении с подходом тотального скрининга, используя критерии общего ожидаемого количества лет жизни с поправкой на качество, количество маммограмм, долю ложноположительных результатов и риска невыявленного рака, требующего оперативного вмешательства.

1.2.4. Доля инвестиций и распределение портфеля

Предположим, мы хотим построить систему, которая рекомендует людям, какую часть личного дохода они могут потратить по итогам года и сколько должны инвестировать⁶. Инвестиционный портфель может включать в себя акции и облигации с разным уровнем риска и ожидаемой доходности. Рост личных накоплений носит стохастический характер из-за неопределенности как в зарплате, так и в инвестиционном доходе, часто увеличиваясь до тех пор, пока инвестор не достигает пенсионного возраста, а затем неуклонно снижается. Удовольствие, получаемое от потребления условной единицы богатства в течение года, обычно уменьшается с увеличением потребляемой суммы, что приводит к желанию оптимально распределить потребление и инвестирование на протяжении всей жизни человека.

1.2.5. Распределенное наблюдение за лесными пожарами

При тушении лесных пожаров одной из главных проблем является недостаточная осведомленность о текущей ситуации. Состояние пожара постоянно меняется под влиянием таких факторов, как ветер и распределение горючего материала в окружающей среде. Многие лесные пожары охватывают большие географические регионы. Один из подходов к наблюдению за лесными пожарами заключается в использовании группы дронов, оснащенных датчиками⁷.

⁵ Такая концепция предложена в статье T. Ayer, O. Alagoz, and N. K. Stout, *A POMDP Approach to Personalize Mammography Screening Decisions*, *Operations Research*, vol. 60, no. 5, pp. 1019–1034, 2012.

⁶ Схожая задача была представлена в R. C. Merton, *Optimum Consumption and Portfolio Rules in a Continuous-Time Model*, *Journal of Economic Theory*, vol. 3, no. 4, pp. 373–413, 1971.

⁷ Это применение было детально представлено в K. D. Julian, M. J. Kochenderfer, *Distributed Wildfire Surveillance with Autonomous Aircraft Using Deep Reinforcement Learning*, *AIAA Journal of Guidance, Control, and Dynamics*, vol. 42, no. 8, pp. 1768–1778, 2019.

Дальность действия датчиков отдельных дронов ограничена, но информация от группы дронов может быть объединена для получения единого снимка ситуации и принятия решений о распределении ресурсов. В идеале хотелось бы, чтобы члены группы самостоятельно определяли, как сотрудничать друг с другом для построения наилучшей информационной картины пожара. Эффективный мониторинг требует принятия решения о том, как маневрировать, чтобы охватить области, в которых новая информация может оказаться полезной, – тратить время на сканирование участков, про которые мы точно знаем, есть ли там пожар, было бы расточительно. Выявление предпочтительных областей для исследования требует рассуждений о стохастической эволюции пожара, основанных лишь на неполных знаниях о его текущем состоянии.

1.2.6. Исследование Марса

Марсоходы сделали важные открытия и значительно улучшили наше понимание Марса. Однако основным узким местом в научных исследованиях была связь между марсоходом и операторами на Земле. Для отправки информации датчиков с Марса на Землю и для отправки ответных команд с Земли на Марс может потребоваться до получаса. Кроме того, наведение антенн марсоходов и другие аспекты их работы необходимо планировать заранее, потому что окна радиообмена Земля–Марс ограничены из-за расположения орбитальных аппаратов, служащих ретрансляторами радиосигнала между планетами. Недавние исследования показали, что эффективность научно-исследовательских миссий может быть повышена в пять раз за счет достижения более высоких уровней автономности⁸. Люди-операторы по-прежнему будут обеспечивать руководство высокого уровня в соответствии с целями миссии, но марсоход получит возможность выбирать свои собственные краткосрочные цели, используя самую свежую информацию. Кроме того, желательно, чтобы марсоход адекватно реагировал на различные опасности и системные сбои без вмешательства человека.

1.3. Методы создания агентов

Существует множество методов создания агентов, принимающих решения. В зависимости от назначения агента и предметной области некоторые могут быть более подходящими, чем другие. Они отличаются обязанностями разработчика и задачами, которые подлежат автоматизации. В этой книге основное внимание уделено планированию и обучению с подкреплением, но некоторые методы будут включать в себя элементы обучения с учителем и оптимизации.

⁸ Эту концепцию представили и оценили D. Gaines, G. Doran, M. Paton, B. Rothrock, J. Russino, R. Mackey, R. Anderson, R. Francis, C. Joswig, H. Justice, K. Kolcio, G. Rabideau, S. Schaffer, J. Sawoniewicz, A. Vasavada, V. Wong, K. Yu, A.-a. Agha-mohammadi, *Self-Reliant Rovers for Increased Mission Productivity, Journal of Field Robotics*, vol. 37, no. 7, pp. 1171–1196, 2020.

1.3.1. Явное программирование

Самый прямой метод создания агента, принимающего решения, состоит в том, чтобы предусмотреть все сценарии, в которых может оказаться агент, и явно запрограммировать ответное поведение агента. Подход с явным программированием может хорошо работать для простых задач, но он возлагает на разработчика большую нагрузку и ответственность за полноту выработанной стратегии. Специально для упрощения программирования агентов разработаны различные языки и среды программирования.

1.3.2. Обучение с учителем

В некоторых случаях проще показать агенту, что делать, чем писать программу, которую будет выполнять агент. Разработчик предоставляет набор обучающих примеров, и алгоритм автоматического обучения должен обобщить эти примеры, найдя связи и закономерности в обучающих данных. Этот подход известен как *обучение с учителем* и широко применяется для задач классификации. Данный метод иногда называют *поведенческим клонированием*, когда он применяется для обучения сопоставлению наблюдений с действиями. Поведенческое клонирование работает наиболее хорошо, когда опытный разработчик действительно знает наилучший план действий для представленного набора ситуаций. Хотя существует множество различных алгоритмов обучения, в новых ситуациях они, как правило, не могут работать лучше, чем разработчики-люди, предложившие обучающие примеры.

1.3.3. Оптимизация

При использовании метода *оптимизации* разработчик задает пространство возможных стратегий принятия решений и показатель качества работы агента, который необходимо максимизировать. Оценка эффективности стратегии принятия решений обычно предусматривает запуск серии прогонов модели. Затем алгоритм оптимизации ищет в этом пространстве оптимальную стратегию. Если пространство относительно невелико, а мера качества не имеет множества локальных оптимумов, то можно использовать различные методы локального или глобального поиска. Хотя обычно предполагают, что для запуска прогонов требуется знание динамической модели, его не используют для направленного поиска.

1.3.4. Планирование

Планирование – это форма оптимизации, в которой используется модель динамики задачи, помогающая вести поиск. На сегодняшний день в литературе описано множество различных задач планирования. Большинство исследований сосредоточены на детерминированных задачах. Для некоторых задач может быть приемлемо аппроксимировать динамику детерминированной

моделью. Предположение о детерминированной модели позволяет нам использовать методы, которые легче масштабировать для многомерных задач. Для некоторых задач принципиально важно учитывать будущую неопределенность. Эта книга полностью посвящена именно таким задачам.

1.3.5. Обучение с подкреплением

Обучение с подкреплением ослабляет предположение о том, что модель известна заранее. Вместо этого стратегия принятия решений изучается «на ходу» – во время взаимодействия агента с окружающей средой. Разработчик должен только предоставить меру качества решений; оптимизацией поведения агента занимается обучающий алгоритм. Одна из интересных сложностей, возникающих при обучении с подкреплением, заключается в том, что выбор действия влияет не только на немедленный успех агента в достижении своих целей, но и на способность агента узнавать об окружающей среде и определять особенности задачи, которыми можно воспользоваться.

1.4. История автоматизации принятия решений

Идея автоматизировать процесс принятия решений уходит своими корнями в мечты ранних философов, ученых, математиков и писателей. Древние греки начали упоминать автоматизацию в мифах и исторических записях еще в 800 г. до н. э. Слово *αὐτόματο* впервые встречается в «Илиаде» Гомера, где содержится упоминание автоматических машин, включая механические тренажеры, используемые для обслуживания гостей за обедом⁹. В XVII веке философы предложили использовать логические правила для автоматического разрешения разногласий. Их идеи заложили основу для механических систем принятия решений.

Начиная с конца XVIII века изобретатели начали создавать автоматические машины для выполнения рутинной работы. В частности, ряд инноваций в текстильной промышленности привел к разработке автоматических ткацких станков, которые, в свою очередь, заложили основу для первых фабричных роботов¹⁰. В начале XIX века описания интеллектуальных машин для автоматизации труда стали появляться в научно-фантастических романах. Слово «робот» впервые появилось в пьесе чешского писателя Карела Чапека «R.U.R.» (сокращение от чеш. *Rossumoví univerzální roboti*, «Россумские универсальные роботы») о машинах, которые могут выполнять работу вместо людей. Пьеса вдохновила других писателей-фантастов включить роботов в свои произведения. В середине XX века известный писатель и профессор Айзек Азимов изложил свое видение робототехники в своей знаменитой серии «Робот».

⁹ S. Vasileiadou, D. Kalligeropoulos, N. Karcianas, *Systems, Modelling and Control in Ancient Greece: Part 1: Mythical Automata, Measurement and Control*, vol. 36, no. 3, pp. 76–80, 2003.

¹⁰ N. J. Nilsson, *The Quest for Artificial Intelligence*. Cambridge University Press, 2009.

Практические реализации автоматизированного принятия решений хуже всего справляются с неопределенностью. Еще в 1991 г. Джордж Данциг, получивший известность благодаря разработке симплексного алгоритма, заявил:

Оглядываясь назад, интересно отметить, что первоначальная задача, с которой началось мое исследование, до сих пор не решена, а именно задача планирования или составления расписания действий, особенно динамического планирования в условиях неопределенности. Успешное решение такой задачи могло бы (в конечном итоге за счет лучшего планирования) способствовать благополучию и стабильности в мире¹¹.

Хотя принятие решений в условиях неопределенности по-прежнему остается областью активных исследований, за последние несколько столетий исследователи и инженеры заметно приблизились к реализации идей древних мечтателей и философов. Современные алгоритмы принятия решений основаны на сближении концепций, возникших в самых разных дисциплинах, включая экономику, психологию, неврологию, информатику, инженерию, математику и исследование операций. Далее мы кратко рассмотрим некоторые основные достижения этих дисциплин. Своеобразное «перекрестное опыление» между дисциплинами уже не раз становилось источником передовых идей в области принятия решений. Вероятно, так будет и впредь.

1.4.1. Экономика

Экономика нуждается в моделях, имитирующих принятие решений человеком. Один из подходов к построению таких моделей опирается на теорию полезности, впервые представленную в конце XVIII века¹². *Теория полезности* предоставляет методику для моделирования и сравнения желательности различных результатов. Например, полезность можно использовать для сравнения желательности денежных количеств. В книге «Основы законодательства» (1887) Иеремия Бентам отметил убывающую полезность богатства:

1. Каждой порции богатства соответствует порция счастья.
2. Из двух людей с неравным состоянием тот, у кого больше всего богатства, имеет больше всего счастья.
3. Избыток счастья богатого не будет столь велик, как избыток его богатства¹³.

Объединив понятие полезности с понятием рационального принятия решений, экономисты середины XX века пришли к формулировке *принципа максимальной ожидаемой полезности*. Этот принцип является ключевой идеей при создании автономных агентов, принимающих решения. Из теории полезности впоследствии развилась *теория игр*, которая пытается описать поведение не-

¹¹ G. B. Dantzig, *Linear Programming, Operations Research*, vol. 50, no. 1, pp. 42–47, 2002.

¹² G. J. Stigler, *The Development of Utility Theory. I*, *Journal of Political Economy*, vol. 58, no. 4, pp. 307–327, 1950.

¹³ J. Bentham, *Theory of Legislation*. Trübner & Company, 1887.

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru