Моей спутнице жизни Фатиме,
а также Оливеру и Якубу.
Павел

Нико и Леунтье ван Рейн за то,
что научили меня тому, что важно в жизни.
Ян

Моим родителям, а также Мануэле,
Кике, Манель и Артуру.
Карлос

Аде, Элиасу, Кобе и Вирле за то,
что напомнили мне, как прекрасен мир.
Хоакин

Содержание

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. Издательство «ДМК Пресс» очень серьезно относится 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

Первое издание этой книги вышло в свет в 2009 г., т. е. к моменту написания второго издания прошло более 10 лет. Поскольку за эти годы область машинного обучения существенно продвинулась вперед, мы решили подготовить второе издание. Мы постарались включить в него наиболее важные достижения, чтобы новая версия книги содержала актуальную информацию об этой области и была полезна исследователям, аспирантам и прикладным специалистам, работающим в этой области.

Каковы основные изменения? Прежде всего если вы просто сравните количество глав двух изданий, то заметите, что их стало вдвое больше. Примерно так же увеличилось количество страниц.

Отметим, что на момент написания первого издания термина *автоматизированное машинное обучение* (automated machine learning, AutoML) даже не существовало. Очевидно, что мы должны были описать это новаторское направление в новом издании, а также прояснить его связь с метаобучением. Кроме того, автоматизация методов проектирования цепочек операций – в настоящее время называемых *конвейерами* (pipeline) или *рабочими процессами* (workflow) – находилась в зачаточном состоянии. Разумеется, мы осознавали необходимость обновить существующий материал, чтобы не отставать от этого развития.

В последние годы исследования в области AutoML и метаобучения привлекают большое внимание не только исследователей, но и многих компаний, занимающихся искусственным интеллектом, включая, например, Google и IBM. Как можно использовать метаобучение для улучшения систем AutoML – это один из важнейших вопросов, на который в настоящее время пытаются ответить многие исследователи.

Эта книга нацелена в будущее. Как это обычно случается, чем лучше исследователи разбираются в какой-то области, тем больше перед ними возникает новых вопросов. Мы позаботились о том, чтобы включить некоторые из них в соответствующие главы.

Авторами первого издания были Павел Браздил, Кристоф Жиро-Каррье, Карлос Соарес и Рикардо Вилальта. Учитывая масштабные нововведения в этой области, мы решили укрепить команду, пригласив Хоакина Ваншорена и Яна ван Рейна присоединиться к проекту. К сожалению, Кристоф и Рикардо не смогли принять участие в работе над новым изданием. Тем не менее все авторы второго издания очень благодарны за их вклад в начало проекта.

Как устроена эта книга

Эта книга состоит из трех частей. В части I (главы 2–7) рассмотрены основные концепции и архитектура систем метаобучения и AutoML, а в части II (главы 8–15) обсуждаются различные расширения. Часть III (главы 16–18) описывает способы хранения и управления метаданными (например, хранилища метаданных) и заканчивается заключительными замечаниями.

Часть I. Основные понятия и архитектура

Глава 1 начинается с объяснения основных понятий, используемых в этой книге, таких как машинное обучение, метаобучение, автоматизированное машинное обучение и др. Затем она продолжается обзором базовой архитектуры системы метаобучения и служит введением к остальной части книги. Над этой главой работали все соавторы книги.

Глава 2 посвящена методам ранжирования на основе метаданных, поскольку их относительно легко реализовать, но это не умаляет их полезности в практических приложениях. Эта глава была написана П. Браздилом и Я. ван Рейном[1]. Глава 3, написанная теми же авторами, посвящена теме оценки метаобучения и систем AutoML. В главе 4 обсуждаются различные показатели наборов данных, которые играют важную роль в качестве метапризнаков в системах метаобучения. Эта глава, как и следующая, также написана П. Браздилом и Я. ван Рейном. Главу 5 можно рассматривать как продолжение главы 2. В ней обсуждаются различные подходы к метаобучению, включая, например, попарные сравнения, которые применялись в прошлом. В главе 6 обсуждается оптимизация гиперпараметров. Она охватывает как базовые методы поиска, так и более продвинутые, применяемые в области AutoML. Эту главу написали три автора – П. Браздил, Я. ван Рейн и Х. Ваншорен. В главе 7 обсуждается вопрос автоматизации построения рабочих процессов или конвейеров, представляющих собой последовательности операций. Эта глава написана П. Браздилом, но в ней повторно использованы некоторые материалы из первого издания, подготовленного К. Жиро-Каррье.

Часть II. Передовые технологии и методы

Часть 2 (главы 8–15) продолжает темы части I, но охватывает различные расширения базовой методологии. Глава 8, написанная П. Браздилом и Я. ван Рейном, посвящена теме построения пространств конфигураций и планированию экспериментов. В двух последующих главах обсуждается конкретная тема ансамблей моделей. Глава 9, написанная Ш. Жиро-Каррье, дополняет материал этой книги. Она описывает различные способы организации набора алгоритмов базового уровня в ансамбли. Авторы второго издания не видели необходимости изменять эту главу, поэтому она сохранена в том виде, в каком появилась в первом издании.

Глава 10 продолжает тему ансамблей и показывает, как метаобучение можно использовать при построении ансамблей (ансамблевом обучении).

---

[1] Части глав 2 и 3 первого издания, написанные К. Соаресом и П. Браздилом, были повторно использованы и адаптированы для этой главы.

Эта глава была написана К. Соаресом и П. Браздилом. Последующие главы посвящены более конкретным темам. Глава 11, написанная Я. ван Рейном, описывает применение метаобучения для предоставления рекомендаций по выбору алгоритма потоковой обработки данных. Глава 12, написанная Р. Вилалтой и М. Месхи, посвящена переносу метамоделей и представляет собой вторую дополняющую главу этой книги. Это существенно обновленная версия аналогичной главы из первого издания, написанная Р. Вилалтой. Глава 13, написанная М. Хьюисманом, Я. ван Рейном и А. Плаатом, обсуждает метаобучение в глубоких нейронных сетях и представляет собой третью дополняющую главу этой книги. Глава 14 посвящена относительно новой теме автоматизации науки о данных. Эта глава была составлена П. Браздилом и содержит обзор различных идей и предложений его соавторов. Цель главы состоит в том, чтобы обсудить различные операции, обычно выполняемые в науке о данных, и рассмотреть вопрос о том, возможна ли здесь автоматизация и можно ли использовать в этом процессе метазнания. Цель главы 15, написанной П. Браздилом, также состоит в том, чтобы заглянуть в будущее и рассмотреть возможность автоматизации проектирования более сложных решений. В их число могут входить не только конвейеры операций, но и более сложные структуры управления (например, итерации) и автоматические изменения в базовом представлении.

Часть III. Организация и использование метаданных

Часть III охватывает некоторые практические вопросы и содержит последние три главы (16–18). В главе 16, написанной Х. Ваншореном и Я. ван Рейном, обсуждаются репозитории метаданных и, в частности, репозиторий, известный под названием OpenML. Этот репозиторий содержит данные о многих экспериментах по машинному обучению, проведенных в прошлом, и их соответствующие результаты. В главе 17, написанной Я. ван Рейном и Х. Ваншореном, показано, как можно изучать метаданные, чтобы получить более глубокое представление об исследованиях машинного обучения и метаобучения и, как следствие, сконструировать новые эффективные прикладные системы. Глава 18 завершает книгу краткими заключительными замечаниями о роли метазнания, а также представляет некоторые перспективные направления исследований. В основном глава была написана П. Браздилом, но содержит вклад других соавторов, в частности Я. ван Рейна и К. Соареса.

Благодарности

Мы выражаем благодарность всем тем, кто помог осуществить этот проект.

Мы признательны за грант 612.001.206 от Голландского исследовательского совета (NWO), направленный на финансирование издания этой книги.

Павел Браздил выражает благодарность Университету Порту, экономическому факультету, научно-исследовательскому институту INESC TEC и одному из его научных центров, а именно Лаборатории искусственного интеллекта и поддержки принятия решений (LIAAD), за их постоянную поддержку. Выполненная работа была частично поддержана национальными фонда-