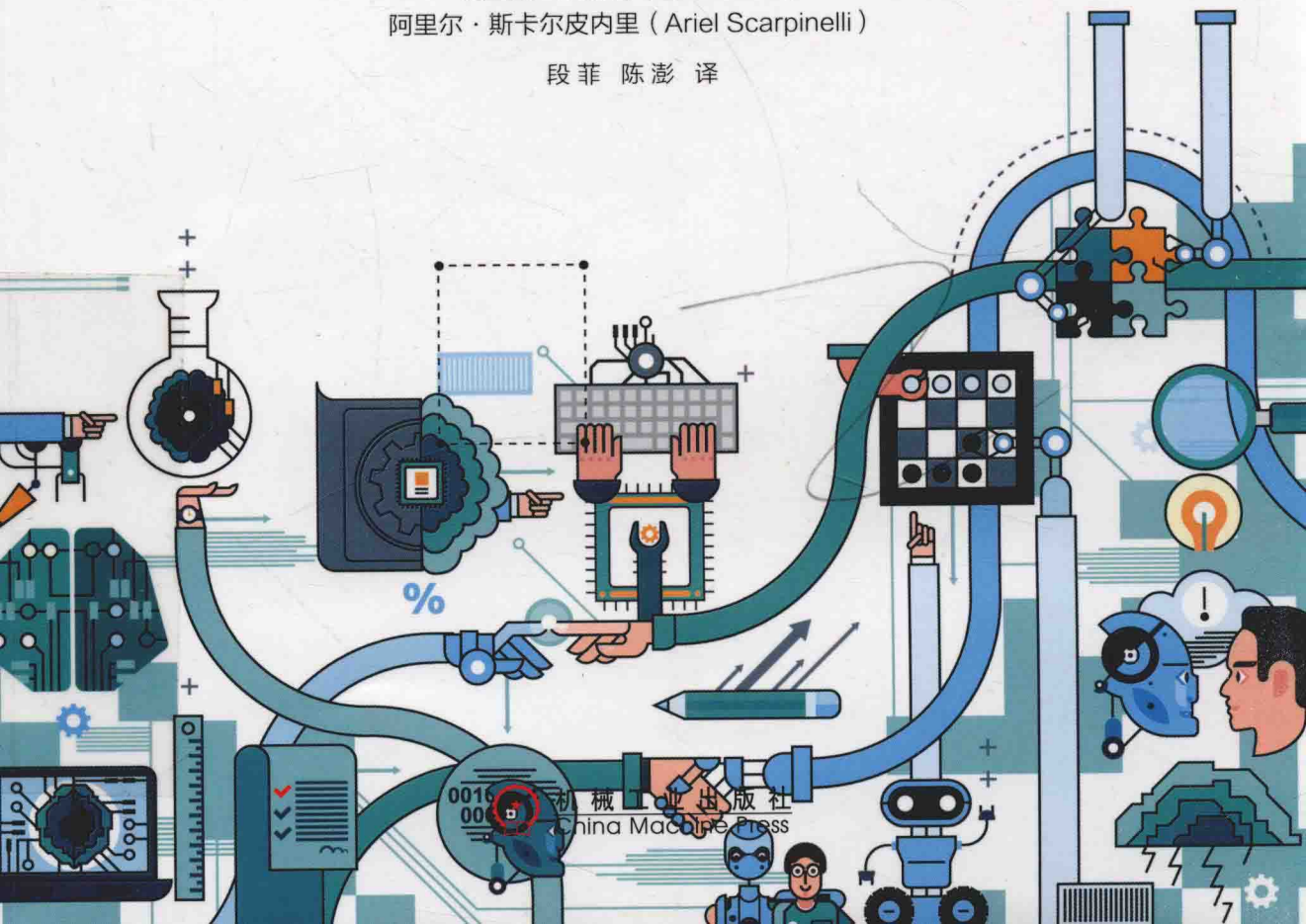


TensorFlow for Machine Intelligence

面向机器智能的 TensorFlow实践

山姆·亚伯拉罕 (Sam Abrahams)
[美] 丹尼亚尔·哈夫纳 (Danijar Hafner) 著
埃里克·厄威特 (Erik Erwit) 著
阿里尔·斯卡尔皮内里 (Ariel Scarpinelli)

段菲 陈澎 译



TensorFlow for Machine Intelligence

面向机器智能的 TensorFlow实践

山姆·亚伯拉罕 (Sam Abrahams)

[美] 丹尼亚尔·哈夫纳 (Danijar Hafner) 著

埃里克·厄威特 (Erik Erwit)

阿里尔·斯卡尔皮内里 (Ariel Scarpinelli)

段菲 陈澎 译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

面向机器智能的 TensorFlow 实践 / (美) 山姆·亚伯拉罕 (Sam Abrahams) 等著; 段菲, 陈澎译. —北京: 机械工业出版社, 2017.4

(智能系统与技术丛书)

书名原文: TensorFlow for Machine Intelligence

ISBN 978-7-111-56389-1

I. 面… II. ①山… ②段… ③陈… III. 人工智能—算法—研究 IV. TP18

中国版本图书馆 CIP 数据核字 (2017) 第 055049 号

本书版权登记号: 图字: 01-2016-5125

TensorFlow for Machine Intelligence (ISBN: 9781939902351).

Copyright © 2016 Bleeding Edge Press. All rights reserved. First published in the English language under the title “TensorFlow for Machine Intelligence” By: Sam Abrahams, Danijar Hafner, Erik Erwitt, Ariel Scarpinelli by Bleeding Edge Press, an imprint of Backstop Media.

Chinese simplified language edition published by China Machine Press.

Copyright © 2017 by China Machine Press.

本书中文简体字版由 Bleeding Edge Press 授权机械工业出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

面向机器智能的 TensorFlow 实践

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 张梦玲

责任校对: 李秋荣

印刷: 北京诚信伟业印刷有限公司

版次: 2017 年 4 月第 1 版第 1 次印刷

开本: 186mm × 240mm 1/16

印张: 14.25

书号: ISBN 978-7-111-56389-1

定价: 69.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzit@hzbook.com

版权所有·侵权必究

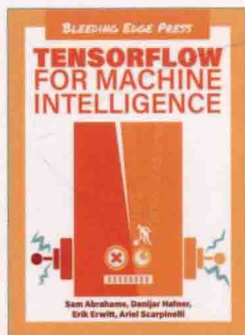
封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

内容简介

本书是一本绝佳的TensorFlow入门指南。几位作者都来自研发一线，他们用自己的宝贵经验，结合众多高质量的代码，生动讲解TensorFlow的底层原理，并从实践角度介绍如何将两种常见模型——深度卷积网络、循环神经网络应用到图像理解和自然语言处理的典型任务中。此外，还介绍了在模型部署和编程中可用的诸多实用技巧。

全书分为四部分，共9章。第一部分（第1~2章）讨论TensorFlow的设计模式以及选择TensorFlow作为深度学习库的优势和面临的挑战，并给出详细的安装指南。第二部分（第3~4章）深入介绍TensorFlow API的基础知识和机器学习基础。第三部分（第5~6章）探讨如何用TensorFlow实现高级深度模型，涉及卷积神经网络（或CNN）模型和循环神经网络（或RNN）模型。第四部分（第7~8章）探讨TensorFlow API中最新推出的特性，包括如何准备用于部署的模型、一些有用的编程模式等。第9章给出一些进一步了解TensorFlow的学习资源。



原书封面

华章IT
HZBOOKS | Information Technology



译者简介

段 菲 清华大学信号与信息处理专业博士，前三星电子中国研究院高级研究员，现为英特尔中国研究院高级研究员。研究方向是深度学习、计算机视觉、数据可视化。参与翻译过《机器学习》《机器学习实践：测试驱动的开发方法》《DirectX 10 3D游戏编程深度探索》等多本图书。

THE TRANSLATOR'S WORDS

译者序

我们正处在一个激动人心的时代，深度学习作为近年来最具突破性的技术之一，极大地推动了人工智能领域的研究进程，并迅速渗透各行各业，同时带动了一大批如家庭服务机器人、自动驾驶这样的新兴产业。不夸张地说，深度学习技术正在深刻地影响着人们的工作、生活和思维方式。

为了帮助人们更高效地从事深度学习研究和应用，并积极分享其研究成果，优秀的开源深度学习框架如雨后春笋般不断涌现。作为深度学习领域巨擘之一的 Google 也在 2015 年 11 月将自己的第二代分布式机器学习系统 TensorFlow 开源。虽然发布时间较晚，但凭借 Google 在业内广泛而巨大的影响力以及顶尖的人工智能研究水平，其代码的简洁性、部署的便利性，以及在学术研究和产品研发两端取得的良好平衡，在极短的时间内便从众多优秀的深度学习框架中脱颖而出，获得了广大开发者的强烈关注。自公开发布以来，TensorFlow 始终保持着兼收并蓄的态势，不断地从其他优秀开源框架中汲取优秀特性，在广大研究和开发人员的强力推动下，不断快速迭代并大幅提升代码的效率，平均每周的代码更新量都超过了万行，所形成的社区开创了空前的活跃度。完全可以预见，TensorFlow 将长期位列一流开源框架的行列。

虽然 TensorFlow 的优点数不胜数，但其“不足”也较为突出，那就是其接口过于复杂，对初学者的编程技能和知识水平要求偏高，学习曲线过陡。本书的问世在一定程度上缓解了这个矛盾。本书的几位作者都来自研发一线，他们用自己的宝贵经验，结合众多高质量的代码，生动讲解了 TensorFlow 的底层原理，并从实战角度介绍了如何将两种常见模型——深度卷积网络、循环神经网络应用到图像理解和自然语言处理的典型任务中。难能可贵的是，他们还介绍了在模型部署和编程中可用的诸多实用技巧。总之，本书非常适合 TensorFlow 的入门学习。

需要说明的是，这不是一本机器学习理论或深度学习的入门读物，阅读本书需要读者对经典机器学习理论和算法、深度卷积网络、循环神经网络的基本原理有初步的了解，并对 Python 编程和常用的 Python 库（如 NumPy 和 matplotlib）较为熟悉。另外，本书的代码是基于 TensorFlow 0.8 版的，译者对 0.9 版所做的接口变动以“译者注”的形式做了部

分说明，并对原书中的一些错误进行了订正。尽管 TensorFlow 1.0 版已经正式发布，接口升级相比以往任何一版都要更多，但笔者认为读者朋友大可不必过于担忧。只要清楚地掌握了 TensorFlow 的基本原理和 Python 编程，并勤于借助互联网，相信接口问题都可迎刃而解。同时，为方便广大读者的学习，原书作者和译者也会抽时间对本书中的实例代码按照 TensorFlow 最新版本进行升级，请大家关注原出版社和机械工业出版社相关主题的后续图书。

感谢机械工业出版社张梦玲编辑在本书翻译过程中给予的诸多帮助，也感谢家人对我无微不至的关心。

深入理解深度学习，从了解优秀的开源框架开始，愿读者朋友们的 TensorFlow 学习之旅一帆风顺！

段菲

2017 年 3 月 3 日

PREFACE

前言

欢迎

自 2015 年 11 月 TensorFlow 第一个开源版本发布以来，它便迅速跻身于最激动人心的机器学习库的行列，并在科研、产品和教育等领域正在得到日益广泛的应用。这个库也在不断地得到改进、充实和优化。与此同时，TensorFlow 社区正以惊人的速度发展壮大。无论你是新手还是有经验的用户，笔者都希望通过本书帮助你提升使用 TensorFlow 的能力，使你自如地充分利用这个功能强大的开源库。

本书的内容编排

第一部分：开启 TensorFlow 之旅

本书第一部分将帮助读者做好使用 TensorFlow 的准备。第 1 章为引言，对 TensorFlow 的历史脉络进行了简要的梳理，并对 TensorFlow 的设计模式以及选择 TensorFlow 作为深度学习库的优势和面临的挑战进行了讨论。

引言之后的第 2 章将介绍安装 TensorFlow 时应当考虑的因素，并给出了详细的 TensorFlow 安装指南，即如何从二进制安装包安装和从源码构建 TensorFlow。

第二部分：TensorFlow 与机器学习基础

从第 3 章开始，进入本书第二部分。在 TensorFlow 安装完毕后，第 3 章将深入介绍 TensorFlow API 的基础知识，而不会涉及过多的机器学习概念。这样做是为了将“学习 TensorFlow”和“学习如何利用 TensorFlow 从事机器学习相关工作”区分开来。第 3 章将对 TensorFlow API 中许多重要的部分进行深入剖析。此外，还将演示如何用可视化的数据流图表示模型，并将其转化为 TensorFlow 代码，以及如何利用 TensorBoard 验证数据流图是否正确建模。

介绍完 TensorFlow API 的核心概念之后，便进入第 4 章。这一章将利用 TensorFlow 实现一些简单的机器学习模型，如线性回归、对数几率回归（logistic regression）和聚类模型。

第三部分：用 TensorFlow 实现更高级的深度模型

第三部分由两章构成，每章都只关注一种更为复杂的深度学习模型。每章首先对模型进行描述，然后介绍如何用可视化的数据流图表示所要创建的模型。这两章还将讨论为什么要以特定方式构建这些模型，并对所涉及的数学难点进行讲解，之后再介绍如何利用 TensorFlow 有效地构建这些模型。

所要研究的第一个模型是卷积神经网络（CNN），对应于第 5 章。该章会介绍如何使用图像数据训练 TensorFlow 模型，并对卷积的数学原理和使用目的展开讨论，同时还将介绍如何将图像裸数据转化为一种与 TensorFlow 兼容的格式，以及如何对最终的输出进行测试。

第 6 章将探讨如何使用 TensorFlow 正确地构建循环神经网络（RNN）模型。通过各种自然语言处理（NLP）任务，读者将了解如何利用长短时记忆网络（LSTM）以及如何将预训练的词向量包含到模型中。

第四部分：其他提示、技术与特性

本书最后一部分将探讨 TensorFlow API 中最新推出的特性，内容包括如何准备用于部署的模型、一些有用的编程模式，以及其他精选主题。

其他机器学习库

TensorFlow 并非唯一可用的开源机器学习库。下面列出一份可用于深度学习的简短开源库清单：

- Caffe 专注于卷积神经网络和图像处理，使用 C++ 语言编写。
- Chainer 是另一个灵活的机器学习 Python 库，支持单机多 GPU 运算。
- CNTK 是微软公司发布的首个开源机器学习库，它拥有自己的模型定义语言，支持声明式的分布式模型构建。
- Deeplearning4j 是一个专门针对神经网络的 Java 库，它易于与 Spark、Hadoop 和其他基于 Java 的分布式软件集成，具有良好的可伸缩性。
- Nervana Neon 是一个高效的 Python 机器学习库，支持单机多 GPU 运算。
- Theano 是一个极为灵活的 Python 机器学习库，因其出众的用户友好性以及可以用异常简单的方式定义复杂模型等特点，在科研领域深受欢迎。TensorFlow 的 API 与 Theano API 最为相似。

- Torch 是一个专注于 GPU 实现的机器学习库，它是用 Lua 语言编写的，并由来自若干家大公司的研究团队提供支持。

限于篇幅，本书不打算对上述这些库的优缺点展开深入讨论，但如果有时间，非常值得深入展开。TensorFlow 的作者在进行框架设计时，便是从当中的几个库汲取了灵感。

先修知识

虽然本书主要关注 TensorFlow API，但笔者希望读者已经熟悉大量数学和编程概念，包括：

- 微积分（一元和多元）
- 矩阵代数（尤其是矩阵乘法）
- 基本的编程原理
- 机器学习的基本概念

此外，读者若能够掌握下列知识，则将从本书中获得更大的收获：

- 拥有 Python 编程及模块组织的经验
- 拥有 NumPy 库的使用经验
- 拥有 matplotlib 库的使用经验
- 掌握机器学习中更高级的概念，尤其是前馈神经网络、卷积神经网络和循环神经网络

在适宜的时候，笔者会通过一些提示信息帮助读者重新熟悉那些为充分理解相关数学和 Python 概念所必需的概念。

预期的收获

通过阅读本书，读者将掌握以下内容：

- TensorFlow 的核心 API
- TensorFlow 的工作流：数据流图的定义和数据流图的执行
- 如何在各种设备上安装 TensorFlow
- 组织代码和项目的最佳实践
- 如何用 TensorFlow 创建核心机器学习模型
- 如何用 TensorFlow 实现 RNN 和 CNN
- 如何用 TensorFlow Serving 部署代码
- 利用 TensorBoard 分析模型的基础知识

在学习完本书之后，如果读者想对 TensorFlow 获得更多了解，可参考下列资源：

- **TensorFlow 官网** 其中包含最新的文档、API 和入门材料。
- **TensorFlow Github 代码库** 在此，可对 TensorFlow 的开源实现做出贡献，并直接对源代码进行审查。
- **官方发布的用 TensorFlow 实现的机器学习模型** 可原封不动地使用这些模型，也可稍加调整以适合自己的设计目的。
- **谷歌研究院的博客 (Google Research Blog)** 提供了来自谷歌的有关 TensorFlow 的应用和更新的最新消息。
- **Kaggle** 获取公开数据集并与其他从事数据分析工作的人开展竞赛的绝佳网站。
- **Data.gov** 美国政府的门户网站，从中可找到全美国的公开数据集。

至此，“动员演讲”已经结束，现在让我们开启本书的学习之旅吧！

目 录

译者序
前言

第一部分 开启 TensorFlow 之旅

第 1 章 引言 2

1.1 无处不在的数据 2

1.2 深度学习 2

1.3 TensorFlow: 一个现代的机器学习库 3

1.4 TensorFlow: 技术概要 3

1.5 何为 TensorFlow 4

1.5.1 解读来自官网的单句描述 4

1.5.2 单句描述未体现的内容 6

1.6 何时使用 TensorFlow 7

1.7 TensorFlow 的优势 8

1.8 使用 TensorFlow 所面临的挑战 9

1.9 高歌猛进 9

第 2 章 安装 TensorFlow 10

2.1 选择安装环境 10

2.2 Jupyter Notebook 与 matplotlib 12

2.3 创建 Virtualenv 环境 12

2.4 TensorFlow 的简易安装 13

2.5 源码构建及安装实例: 在 64 位
Ubuntu Linux 上安装 GPU 版

TensorFlow 14

2.5.1 安装依赖库 14

2.5.2 安装 Bazel 15

2.5.3 安装 CUDA 软件 (仅限
NVIDIA GPU) 162.5.4 从源码构建和安装
TensorFlow 18

2.6 安装 Jupyter Notebook 20

2.7 安装 matplotlib 20

2.8 测试 TensorFlow、Jupyter Notebook
及 matplotlib 21

2.9 本章小结 23

第二部分 TensorFlow 与机器
学习基础

第 3 章 TensorFlow 基础 26

3.1 数据流图简介 26

3.1.1 数据流图基础 26

3.1.2 节点的依赖关系 29

3.2 在 TensorFlow 中定义数据流图 33

3.2.1 构建第一个 TensorFlow	5.2.4 数据格式	104
数据流图	5.2.5 深入探讨卷积核	105
3.2.2 张量思维	5.3 常见层	107
3.2.3 张量的形状	5.3.1 卷积层	108
3.2.4 TensorFlow 的 Operation	5.3.2 激活函数	108
3.2.5 TensorFlow 的 Graph 对象	5.3.3 池化层	111
3.2.6 TensorFlow Session	5.3.4 归一化	113
3.2.7 利用占位节点添加输入	5.3.5 高级层	114
3.2.8 Variable 对象	5.4 图像与 TensorFlow	116
3.3 通过名称作用域组织数据流图	5.4.1 加载图像	116
3.4 练习：综合运用各种组件	5.4.2 图像格式	117
3.4.1 构建数据流图	5.4.3 图像操作	121
3.4.2 运行数据流图	5.4.4 颜色	127
3.5 本章小结	5.5 CNN 的实现	129
第 4 章 机器学习基础	5.5.1 Stanford Dogs 数据集	129
4.1 有监督学习简介	5.5.2 将图像转为 TFRecord	
4.2 保存训练检查点	文件	130
4.3 线性回归	5.5.3 加载图像	133
4.4 对数几率回归	5.5.4 模型	134
4.5 softmax 分类	5.5.5 训练	136
4.6 多层神经网络	5.5.6 用 TensorBoard 调试	
4.7 梯度下降法与误差反向传播算法	滤波器	137
	5.6 本章小结	139
第三部分 用 TensorFlow 实现更高级的深度模型	第 6 章 循环神经网络与自然语言处理	140
第 5 章 目标识别与分类	6.1 循环神经网络简介	140
5.1 卷积神经网络	6.1.1 时序的世界	140
5.2 卷积	6.1.2 近似任意程序	141
5.2.1 输入和卷积核	6.1.3 随时间反向传播	142
5.2.2 跨度	6.1.4 序列的编码和解码	143
5.2.3 边界填充	6.1.5 实现第一个循环神经网络	145
	6.1.6 梯度消失与梯度爆炸	145
	6.1.7 长短时记忆网络	147

6.1.8 RNN 结构的变种	148
6.2 词向量嵌入	149
6.2.1 准备维基百科语料库	151
6.2.2 模型结构	155
6.2.3 噪声对比分类器	156
6.2.4 训练模型	156
6.3 序列分类	157
6.3.1 Imdb 影评数据集	158
6.3.2 使用词向量嵌入	159
6.3.3 序列标注模型	159
6.3.4 来自最后相关活性值的 softmax 层	161
6.3.5 梯度裁剪	162
6.3.6 训练模型	163
6.4 序列标注	164
6.4.1 OCR 数据集	164
6.4.2 时间步之间共享的 soft- max 层	166
6.4.3 训练模型	169
6.4.4 双向 RNN	171
6.5 预测编码	174
6.5.1 字符级语言建模	174
6.5.2 ArXiv 摘要 API	175
6.5.3 数据预处理	177
6.5.4 预测编码模型	178

6.5.5 训练模型	182
6.5.6 生成相似序列	185

6.6 本章小结	188
----------	-----

第四部分 其他提示、技术与特性

第 7 章 产品环境中模型的部署 190

7.1 搭建 TensorFlow 服务开发环境	190
7.1.1 Docker 镜像	190
7.1.2 Bazel 工作区	191
7.2 导出训练好的模型	192
7.3 定义服务器接口	195
7.4 实现推断服务器	197
7.5 客户端应用	201
7.6 产品准备	203
7.7 本章小结	203

第 8 章 辅助函数、代码结构和类 204

8.1 确保目录结构存在	204
8.2 下载函数	204
8.3 磁盘缓存修饰器	205
8.4 属性字典	206
8.5 惰性属性修饰器	207
8.6 覆盖数据流图修饰器	209

第 9 章 结语：其他资源 212

PART 1

第一部分

开启 TensorFlow 之旅

第 1 章 引言

第 2 章 安装 TensorFlow

引言

1.1 无处不在的数据

我们正实实在在地处于“信息时代”。如今，各种数据从无穷无尽的渠道不断涌入：智能手机、手表、汽车、停车计时表、家用电器。几乎任何一种能够说出名字的技术与生俱来都具备与位于云端的某个数据库进行通信的能力。在看似无限的存储能力的支持下，开发者们为数据仓库选择了一种“更多即是更好”的方法，存储着从他们的产品和客户那里收集到的以拍字节（PB）为单位计的海量数据。

与此同时，计算机的性能也在持续提升。虽然 CPU 的发展速度已经放缓，但并行处理架构取得了爆炸式的发展。一向主要服务于计算机游戏的图形处理单元（GPU）如今已被大量运用于通用计算领域，并拉开了机器学习迅猛发展的序幕。

机器学习（有时简记为 ML）试图利用通用的数学模型回答涉及数据的特定问题。机器学习成功应用于垃圾邮件检测、产品推荐（向客户）、预测商品的价格等领域已有多年的。近年来，一种特殊类型的机器学习范式在几乎所有领域都取得了无数巨大的成功，它就是深度学习。

1.2 深度学习

“深度学习”已成为用于描述使用多层神经网络的过程的标准术语，多层神经网络是一类极为灵活的可利用种类繁多的数学方法以及不同数学方法组合的模型。这类模型极为强大，但直到最近几年，人们才有能力卓有成效地利用神经网络，其背后原因主要有两点，一是获取足够数量的数据成为现实；二是得益于通用 GPU 的快速发展，多层神经网络拥有了超越其他机器学习方法所必需的计算能力^①。

深度学习的强大之处在于当决定如何最有效地利用数据时，它能够赋予模型更大的灵

① 还应补充第三点原因：更有效的训练算法。——译者注