联邦学习

Federated Learning

杨强 刘洋 程勇 康焱 陈天健 于涵 著

电子工业出版社

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内容简介

如何在保证本地训练数据不公开的前提下，实现多个数据拥有者协同训练一个共享的机器学习模型？传统的机器学习方法需要将所有的数据集中到一个地方（例如，数据中心），然后进行机器学习模型的训练。但这种基于集中数据的做法无疑会严重侵害用户隐私和数据安全。如今，世界上越来越多的人开始强烈要求科技公司必须根据用户隐私法律法规妥善地处理用户的数据，欧盟的《通用数据保护条例》是一个很好的例子。在本书中，我们将描述联邦学习（亦称联邦机器学习）如何将分布式机器学习、密码学、基于金融规则的激励机制和博弈论结合起来，以解决分散数据的使用问题。我们会介绍不同种类的面向隐私保护的机器学习解决方案以及技术背景，并描述一些典型的实际问题解决案例。我们还会进一步论述联邦学习将成为下一代机器学习的基础，可以满足技术和社会需求并促进面向安全的人工智能的开发和应用。

本书可供计算机科学、人工智能和机器学习专业的学生，以及大数据和人工智能应用程序的开发人员阅读，也可供高等院校的教员、研究机构的研究人员、法律法规制定者和政府监管部门参考。

Original English Language edition published by Morgan & Claypool Publishers.
Part of Synthesis Lectures on Artificial Intelligence and Machine Learning.
Edited by Ronald Brachman, Francesca Rossi, and Peter Stone.
Copyright © 2020 by Morgan & Claypool Publishers
All rights reserved.
The Simplified Chinese translation rights arranged through Rightol Media.（本书中文简体版权经由锐拓传媒取得，E-mail:copyright@rightol.com）

版权贸易合同登记号　图字：01-2020-0923

图书在版编目 (CIP) 数据

联邦学习 = Federated Learning / 杨强等著. — 北京：电子工业出版社，2020. 4
ISBN 978-7-121-38522-3

Ⅰ. ①联… Ⅱ. ①杨… Ⅲ. ①机器学习 Ⅳ. ① TP181

中国版本图书馆 CIP 数据核字（2020）第 029798 号

责任编辑：宋亚东
印　　刷：
装　　订：
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　　邮编：100036
开　　本：720×1000　1/16　　印张：13　　字数：229 千字
版　　次：2020 年 4 月第 1 版
印　　次：2020 年 4 月第 1 次印刷
定　　价：89.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。

电子工业出版社

Foreword
序　言

人工智能安全

21 世纪初，人工智能（Artificial Intelligence，AI）进入以深度学习为主导的大数据时代，基于大数据的机器学习既推动了 AI 的蓬勃发展，也带来一系列安全隐患。这些隐患来源于深度学习本身的学习机制，无论是在它的模型建造（训练）阶段，还是在模型推理和使用阶段。这些安全隐患如果被有意或无意地滥用，后果将十分严重。当前 AI 安全已引起人们普遍的关注，各项的治理措施也因此积极开展。AI 治理有以下几个不同的维度，即技术、法律、经济和文化等。“联邦学习”（Federated Learning）正是在这个背景下提出和发展起来的，它主要从技术维度出发，重点研究其中的隐私保护和数据安全问题。那么联邦学习是如何保护隐私和数据安全的？它包括两个过程，分别是模型训练和模型推理。在模型训练阶段，模型相关的信息可以在各方之间交换，但数据不能交换，因此各个站点上的数据将受到保护。在模型推理阶段，训练好的联邦学习模型可以置于联邦学习系统的各参与方，也可以供多方共享。这是联邦学习的具体过程，也就是它的定义。

本书是关于联邦学习的介绍，共 11 章，内容丰富。从广度上看，书中讨论了四种联邦学习的基本类型，即横向联邦学习、纵向联邦学习、联邦迁移学习和联邦强化学习，还讨论了相关的联邦学习激励机制和分布式机器学习。从深度上看，书中包括原理、算法、平台和应用实例。本书作者杨强等均来自微众银行，他们都参与了联邦智能使能器（Federated AI Technology Enabler，FATE）的联邦学习平台的开发。本书的许多思想来源于这个实践，因此具有实用性。本书可以作为计算机科学、人工智能和机器学习专业的学生，以及大数据和人工智能应用程序开发人员的入门参考书，也可供本科高年级学生或者研究生、大学的教员和研究机构的研究人员阅读。

张钹

中国科学院院士，清华大学人工智能研究院院长

电子工业出版社

Preface
前　言

本书讲述在数据间彼此孤立、同时被不同组织所拥有且并不能被轻易地聚合在一起的环境下，联合构建机器学习模型的方法。我们经常可以听到，当今是大数据（Big Data）时代，而大数据正是人工智能（Artificial Intelligence，AI）应用蓬勃发展的“燃料”。事实却是，我们面对的数据常常既是小规模，又是碎片化的。例如，我们不能随意收集由移动终端设备产生的数据，这些数据都以碎片化的形式分散存在。像医院这样的机构，由于行业的特殊性，对用户数据的掌握量往往是有限的。然而，由于用户隐私和数据安全方面的需求，使得在不同机构间以简单的方式将所有数据聚合到一处并进行处理变得越来越不可行。在这样的环境中，联邦机器学习（Federated Machine Learning），或者简称为联邦学习（Federated Learning），作为一种行之有效的解决方案引起了人们的广泛关注。联邦学习既能帮助多个参与方搭建共享的高性能模型，又符合用户隐私和数据保密性的要求。

除了保护用户隐私和数据安全，联邦学习的另一发展动机是为了最大化地利用云系统下终端设备的计算能力。如果只在设备和服务器之间传输计算结果而不是原始数据，那么通信将会变得极为高效。人造卫星能够完成绝大部分的信息收集计算，并只需使用最低限度的信道与地面计算机通信。联邦学习通过交换中间计算结果即可在多台设备和计算服务器之间进行同步。

我们可以打个比方来通俗地解释联邦学习，把机器学习模型比作羊，把数据比作羊吃的草。在传统方法中，要建立机器学习模型，需要到各个草场的供应商处收购草。这就像一家人工智能公司需要到处收集数据一样，会面对很多的挑战，例如用户隐私、各个组织的利益和法律法规的约束等。联邦学习则换了一种思路，我们可以牵着羊，到各个草场去吃草，这样羊就可以吃到每个地方的草，羊可以成长，而草不出

电子工业出版社

本地，就像联邦学习系统里的数据不出本地一样。羊吃了各家的草，可以逐渐长大，就像联邦模型在各个地方的数据集上都获得知识，变得越来越好，最后联邦模型可以供大家一起使用一样。这也是本书的封面所展示的意义。

如今，现代社会需要人们更负责任地使用人工智能，而用户隐私和数据完整性是人工智能系统的重要特征。在这一方向，从安全地更新移动电话上的输入法预测模型，到与多家医院一同改善医疗图像识别模型的性能，联邦学习已经产生了显著的积极影响。在计算机科学领域，有许多已有的研究成果为联邦学习技术奠定了基础。自从谷歌发布了一个名为 Gboard 的应用程序后，联邦学习技术在 2018 年左右开始迅速崛起。

谷歌的 Gboard 系统是一个企业对消费者（Business-to-Consumer，B2C）应用的例子。它也能够用于支持边缘计算，云系统的终端（边缘）设备可以处理许多计算任务，从而减少了通过原始数据与中央服务器通信的需要。另一个维度是企业对企业（Business-to-Business，B2B）应用。在此类应用中，多个组织联合起来搭建一个共享的机器学习模型。模型是在确保没有本地数据离开任何站点的同时构建的，而模型性能可以根据业务需求进行一定程度的定制。在本书中，我们涵盖了 B2C 模型和 B2B 模型。

为了推进联邦学习技术，需要多个学科领域的合作，包括机器学习算法、分布式机器学习、密码学与安全、隐私保护数据挖掘、博弈论与经济学原理、激励机制设计、法律与监管要求等。要同时精通如此多的学科，对一位研究者或工程师来说是一个极其艰巨的任务。目前，研究联邦学习领域的资源分散于许多研究论文和博客上，因此，我们有必要在一本书中进行全面的介绍。

本书的内容是关于联邦学习的介绍，可以作为读者入门和探究联邦学习所需阅读的第一本书。本书是为计算机科学、人工智能和机器学习专业的学生，以及大数据和人工智能应用程序的开发人员编写的。本科高年级学生或者研究生、大学的教员和研究机构的研究人员都能够发现这本书的有用之处。在课堂上，本书可以作为研究生研讨课程的教科书，也可以作为研究联邦学习的参考文献。法律法规制定者和政府监管部门也可以把这本书作为一本关于大数据和人工智能法律事务的参考书。

本书的想法来自我们在微众银行开发的一个名为联邦智能使能器（Federated AI Technology Enabler，FATE）的联邦学习平台，是第一个工业级联邦学习开源框架。

电子工业出版社

FATE 平台现已是 Linux 基金会的一部分。微众银行是一家服务于中国数亿用户的数字银行，拥有来自不同背景的商业合作伙伴，包括银行、保险公司、互联网公司、零售公司和供应链公司等。我们亲身体会到，由于数据不能轻易地共享和传输，导致合作构建由机器学习所支撑的新业务的需求正变得愈加强烈。

谷歌将联邦学习大规模地应用在其面向消费者的移动服务中。我们进一步扩大了联邦学习的适用范围，使多家企业结为伙伴关系。基于联邦学习的横向、纵向和迁移学习分类首次在我们发表于 *ACM TIST* (*ACM Transactions on Intelligent Systems and Technology*) 的研究论文中提出[1]，也于 2019 年在夏威夷由人工智能发展协会组织举办的 AAAI（Association for the Advancement of Artificial Intelligence）会议上提出。随后，在第 14 届中国计算机联盟科技前沿大会等会议上，参会者们提供了许多关于联邦学习的教程。在本书的编写过程中，我们的第一个开源联邦学习系统 FATE 诞生了[2]。此外，联邦学习的第一个 IEEE ① 国际标准正在制定中[3]。各种教程和相关的研究论文是本书的基础所在。

本书的结构安排如下。第 1 章介绍当前人工智能面临的挑战以及将联邦学习作为可行的解决方案。第 2 章提供面向隐私保护的机器学习的背景知识，包括常用的隐私保护技术和数据安全技术。第 3 章是分布式机器学习概述，包括面向扩展性的分布式机器学习和面向隐私保护的分布式机器学习，并强调了联邦学习和分布式机器学习的区别。第 4 章、第 5 章和第 6 章分别详细地介绍了横向联邦学习、纵向联邦学习和联邦迁移学习。第 7 章探讨联邦学习激励机制的设计，以便更好地激励联邦学习的参与方。第 8 章介绍联邦学习在计算机视觉、自然语言处理及推荐系统领域的研究和应用。第 9 章介绍联邦强化学习。第 10 章讨论联邦学习在各个领域的应用前景。第 11 章总结此书，并展望联邦学习的未来发展。最后，附录 A 中提供了当前最新的欧盟、美国和中国的数据保护法律和法规概况。

杨强，刘洋，程勇，康焱，陈天健，于涵

2020 年 4 月，中国 深圳

① The Institute of Electrical and Electronics Engineers。

电子工业出版社

Acknowledgments

致　谢

为完成本书的撰写，一群非常敬业的学者和工程师付出了巨大的努力。除了本书的作者，也有许多博士研究生、研究人员和研究伙伴为不同章节做出了贡献。我们衷心地感谢以下为本书的写作和编校做出贡献的人士。

- 周雨豪协助完成了本书从英文到中文翻译的初稿。
- 高大山协助撰写了第 2 章和第 3 章。
- 吴学阳协助撰写了第 3 章和第 5 章。
- 梁新乐协助撰写了第 3 章和第 9 章。
- 黄云峰协助撰写了第 5 章和第 8 章。
- 万晟协助撰写了第 6 章和第 8 章。
- 魏锡光协助撰写了第 9 章。
- 邢鹏威协助撰写了第 8 章和第 10 章。

最后，我们要感谢我们的家人对我们的理解与支持！

杨强，刘洋，程勇，康焱，陈天健，于涵

2020 年 4 月，中国 深圳

电子工业出版社

Author's Biographies
作 者 简 介

电子工业出版社

杨强

杨强教授是微众银行的首席人工智能官（CAIO）和香港科技大学（HKUST）计算机科学与工程系讲席教授。他是香港科技大学计算机科学与工程系的前任系主任，并曾担任大数据研究院的创始主任（2015-2018 年）。他的研究兴趣包括人工智能、机器学习和数据挖掘，特别是迁移学习、自动规划、联邦学习和基于案例的推理。他是多个国际协会的会士（Fellow），包括 ACM、AAAI、IEEE、IAPR 和 AAAS。他于 1982 年获得北京大学天体物理学学士学位，分别于 1987 年和 1989 年获得马里兰大学帕克分校计算机科学系硕士学位和博士学位。他曾在在滑铁卢大学（1989-1995 年）和西蒙弗雷泽大学（1995-2001 年）担任教授。他是 *ACM TIST* 和 *IEEE TBD* 的创始主编。他是国际人工智能联合会议（IJCAI）的理事长（2017-2019 年）和人工智能发展协会（AAAI）的执行委员会成员（2016-2020 年）。杨强教授曾获多个奖项，包括 2004/2005 ACM KDDCUP 冠军、ACM SIGKDD 卓越服务奖（2017）、AAAI 创新人工智能应用奖（2018, 2020）和吴文俊人工智能杰出贡献奖（2019）。他是华为诺亚方舟实验室的创始主任（2012-2014 年）和第四范式（AI 平台公司）的共同创始人。他是 *Intelligent Planning*（Springer）、*Crafting Your Research Future* (Morgan & Claypool)、*Transfer Learning*（Cambridge University Press）与 *Constraint-based Design Recovery for Software Engineering*（Springer）等著作的作者。

刘洋

刘洋是微众银行 AI 项目组的高级研究员。她的研究兴趣包括机器学习、联邦学习、迁移学习、多智能体系统、统计力学，以及这些技术的产业应用。她于 2012 年获得普林斯顿大学博士学位，2007 年获得清华大学学士学位。她拥有多项国际发明专利，研究成果曾发表于 *Nature*、IJCAI 和 *ACM TIST* 等科研刊物和会议上。她曾获 AAAI 人工智能创新应用奖、IJCAI 创新应用奖等多个奖项，并担任 IJCAI 高级程序委员会委员，NeurIPS 等多个人工智能会议研讨会联合主席，以及 *IEEE Intelligent Systems* 期刊客座编委等。

程勇

程勇是微众银行 AI 项目组的高级研究员。他曾任华为技术有限公司（深圳）高级工程师和德国贝尔实验室高级研究员，也曾在华为-香港科技大学创新实验室担任研究员。他的研究兴趣和专长主要包括联邦学习、深度学习、计算机视觉和 OCR、数学优化理论和算法、分布式和网络计算以及混合整数规划。他发表期刊和会议论文 20 余篇。他于 2006 年、2010 年、2013 年分别在浙江大学、香港科技大学、德国达姆施塔特工业大学获工学学士学位（一等荣誉）、硕士学位和博士学位。他于 2014 年获达姆施塔特工业大学最佳博士论文奖，于 2006 年获浙江大学最佳学士论文奖。他在 ICASSP'15 会议上做了关于“混合整数规划”的教程。他是 IJCAI'19 和 NIPS'19 等国际会议的程序委员会委员。

康焱

康焱是微众银行 AI 项目组的高级研究员。他的工作重点是面向隐私保护的机器学习和联邦迁移学习技术的研究和实现。他在马里兰大学巴尔的摩分校获计算机硕士和博士学位。他的博士论文研究的是以机器学习和语义网络进行异构数据集成，并获得了博士论文奖学金。在就读研究生期间，他参与了与美国国家

电子工业出版社

标准与技术研究院（NIST）和美国国家科学基金会（NSF）合作的多个项目，设计和开发语义网络集成系统。他在商业软件项目方面也有着丰富的经验。他曾在美国 Stardog Union 公司和美国塞纳公司工作了四年多的时间，从事系统设计和实现方面的工作。

陈天健

陈天健是微众银行 AI 项目组的副总经理。他现在负责构建基于联邦学习技术的银行智能生态系统。在加入微众银行之前，他是百度金融的首席架构师，同时也是百度的首席架构师。他拥有超过 12 年的大规模分布式系统设计经验，并在 Web 搜索引擎、对等网络存储、基因组学、推荐系统、数字银行和机器学习等多个应用领域中实现了技术创新。他现居于中国深圳，与其他工作伙伴一起建设和推广联邦 AI 生态系统和相关的开源项目 FATE。

于涵

于涵现任职新加坡南洋理工大学（NTU）计算机科学与工程学院助理教授、微众银行特聘顾问。在 2015—2018 年期间，他在南洋理工大学担任李光耀博士后研究员（LKY PDF）。在加入南洋理工大学之前，他曾在新加坡惠普公司担任嵌入式软件工程师。他于 2014 年获南洋理工大学计算机科学博士学位。他的研究重点是在线凸优化、人工智能伦理、联邦学习及其在众包等复杂协作系统中的应用。他在国际学术会议和期刊上发表研究论文 120 余篇，获得了多项科研奖项。

电子工业出版社

微信扫码回复：38522

- 获取博文视点学院 20 元付费内容抵扣券；
- 获取免费增值资源，如白皮书、演讲视频及 PPT、FATE 教程和应用案例等；
- 加入本书读者交流群，与本书作者互动；
- 获取精选书单推荐。

轻松注册成为博文视点社区（www.broadview.com.cn）用户，您对书中内容的修改意见可在本书页面的“提交勘误”处提交，若被采纳，将获赠博文视点社区积分。在您购买电子书时，积分可用来抵扣相应金额。

您也可以登录https://cn.fedai.org/，获取联邦学习的最新进展和更多动态。

电子工业出版社

Contents

目　录

序言 iii
前言 iv
作者简介 viii

第 1 章　引言 /1
1.1　人工智能面临的挑战 /2
1.2　联邦学习概述 /4
1.2.1　联邦学习的定义 /5
1.2.2　联邦学习的分类 /8
1.3　联邦学习的发展 /11
1.3.1　联邦学习的研究 /11
1.3.2　开源平台 /13
1.3.3　联邦学习标准化进展 /14
1.3.4　联邦人工智能生态系统 /15

第 2 章　隐私、安全及机器学习 /17
2.1　面向隐私保护的机器学习 /18
2.2　面向隐私保护的机器学习与安全机器学习 /18
2.3　威胁与安全模型 /19
2.3.1　隐私威胁模型 /19
2.3.2　攻击者和安全模型 /21

电子工业出版社

2.4 隐私保护技术 /22
2.4.1 安全多方计算 /22
2.4.2 同态加密 /27
2.4.3 差分隐私 /30

第 3 章 分布式机器学习 /35
3.1 分布式机器学习介绍 /36
3.1.1 分布式机器学习的定义 /36
3.1.2 分布式机器学习平台 /37
3.2 面向扩展性的 DML /39
3.2.1 大规模机器学习 /39
3.2.2 面向扩展性的 DML 方法 /40
3.3 面向隐私保护的 DML /43
3.3.1 隐私保护决策树 /43
3.3.2 隐私保护方法 /45
3.3.3 面向隐私保护的 DML 方案 /45
3.4 面向隐私保护的梯度下降方法 /48
3.4.1 朴素联邦学习 /49
3.4.2 隐私保护方法 /49
3.5 挑战与展望 /51

第 4 章 横向联邦学习 /53
4.1 横向联邦学习的定义 /54
4.2 横向联邦学习架构 /55
4.2.1 客户-服务器架构 /55
4.2.2 对等网络架构 /58
4.2.3 全局模型评估 /59
4.3 联邦平均算法介绍 /60
4.3.1 联邦优化 /60
4.3.2 联邦平均算法 /63

电子工业出版社

4.3.3 安全的联邦平均算法 /65
4.4 联邦平均算法的改进 /68
4.4.1 通信效率提升 /68
4.4.2 参与方选择 /69
4.5 相关工作 /69
4.6 挑战与展望 /71

第 5 章 纵向联邦学习 /73
5.1 纵向联邦学习的定义 /74
5.2 纵向联邦学习的架构 /75
5.3 纵向联邦学习算法 /77
5.3.1 安全联邦线性回归 /77
5.3.2 安全联邦提升树 /80
5.4 挑战与展望 /85

第 6 章 联邦迁移学习 /87
6.1 异构联邦学习 /88
6.2 联邦迁移学习的分类与定义 /88
6.3 联邦迁移学习框架 /90
6.3.1 加法同态加密 /93
6.3.2 联邦迁移学习的训练过程 /94
6.3.3 联邦迁移学习的预测过程 /95
6.3.4 安全性分析 /95
6.3.5 基于秘密共享的联邦迁移学习 /96
6.4 挑战与展望 /97

第 7 章 联邦学习激励机制 /99
7.1 贡献的收益 /100
7.1.1 收益分享博弈 /100
7.1.2 反向拍卖 /102

电子工业出版社

7.2 注重公平的收益分享框架 /103
7.2.1 建模贡献 /103
7.2.2 建模代价 /104
7.2.3 建模期望损失 /105
7.2.4 建模时间期望损失 /105
7.2.5 策略协调 /106
7.2.6 计算收益评估比重 /108
7.3 挑战与展望 /109

第 8 章 联邦学习与计算机视觉、自然语言处理及推荐系统 /111
8.1 联邦学习与计算机视觉 /112
8.1.1 联邦计算机视觉 /112
8.1.2 业内研究进展 /114
8.1.3 挑战与展望 /115
8.2 联邦学习与自然语言处理 /116
8.2.1 联邦自然语言处理 /116
8.2.2 业界研究进展 /118
8.2.3 挑战与展望 /118
8.3 联邦学习与推荐系统 /119
8.3.1 推荐模型 /120
8.3.2 联邦推荐系统 /121
8.3.3 业界研究进展 /123
8.3.4 挑战与展望 /123

第 9 章 联邦强化学习 /125
9.1 强化学习介绍 /126
9.1.1 策略 /127
9.1.2 奖励 /127
9.1.3 价值函数 /127

电子工业出版社

9.1.4 环境模型 /127
9.1.5 强化学习应用举例 /127
9.2 强化学习算法 /128
9.3 分布式强化学习 /130
9.3.1 异步分布式强化学习 /130
9.3.2 同步分布式强化学习 /131
9.4 联邦强化学习 /131
9.4.1 联邦强化学习背景 /131
9.4.2 横向联邦强化学习 /132
9.4.3 纵向联邦强化学习 /134
9.5 挑战与展望 /136

第 10 章 应用前景 /139
10.1 金融 /140
10.2 医疗 /141
10.3 教育 /142
10.4 城市计算和智慧城市 /144
10.5 边缘计算和物联网 /146
10.6 区块链 /147
10.7 第五代移动网路 /148

第 11 章 总结与展望 /149

附录 A 数据保护法律和法规 /151
A.1 欧盟的数据保护法规 /152
A.1.1 GDPR 中的术语 /153
A.1.2 GDPR 重点条款 /154
A.1.3 GDPR 的影响 /156
A.2 美国的数据保护法规 /157
A.3 中国的数据保护法规 /158

参考文献 /161

电子工业出版社

CHAPTER 1 引言

本章将介绍当前人工智能面临的挑战以及联邦学习可以作为一个有效的解决方案，并介绍联邦学习的分类和发展。

电子工业出版社

CHAPTER 2
隐私、安全及机器学习

本章介绍面向隐私保护的机器学习相关的背景知识，包括面向隐私保护的机器学习技术和数据分析。

电子工业出版社

CHAPTER 3
分布式机器学习

我们从第 1 章了解到，联邦学习和分布式机器学习（Distributed Machine Learning，DML）有许多共同之处，例如，二者都使用分散化的数据集和分布式的模型训练。很多研究者也把联邦学习看作分布式机器学习的一种特殊形式，例如文献[15, 46, 73, 153]，或者把联邦学习看成是分布式机器学习的下一步发展。为了让读者更深入地了解联邦学习，我们将在本章介绍 DML 的概况，包括面向扩展性的 DML 和面向隐私保护的 DML。

DML 包括许多方面，例如，训练数据的分布式存储、计算任务的分布式操作和模型服务的分布式部署等。关于 DML，有大量的研究论文、著作以及章节内容，例如文献[43, 44, 154–157]。因此，我们不打算再次就这一问题提供一个全面的介绍或综述，这里将重点放在 DML 中与联邦学习关系紧密的方面。我们会为对 DML 技术感兴趣的读者指出一些参考资料。

电子工业出版社

CHAPTER 4 横向联邦学习

本章将介绍横向联邦学习（Horizontal Federated Learning，HFL），包括横向联邦学习的定义、架构、相关研究工作和应用案例，以及面临的挑战。

电子工业出版社

CHAPTER 5
纵向联邦学习

本章将介绍如何搭建跨部门或机构的联邦学习，使得他们可以通过共有的样本集在保证安全隐私的条件下，利用他们各自不同的特征集展开联邦学习，并建立和使用模型。

电子工业出版社

CHAPTER 6
联邦迁移学习

我们分别在第 4 章和第 5 章讨论了横向联邦学习（HFL）和纵向联邦学习（VFL）。横向联邦学习要求所有参与方的样本具有相同的特征空间，而纵向联邦学习需要各参与方具有相同的样本空间和不同的特征空间。然而在实践中，我们经常会面临的情况是，各参与方间并没有足够的共同特征或样本。在这种情况下，通过迁移学习（Transfer Learning, TL）技术，我们仍然可以建立一个拥有良好性能的联邦学习模型。我们将这种结合称为联邦迁移学习（Federated Transfer Learning，FTL）。本章将提供关于联邦迁移学习的正式定义，并讨论联邦迁移学习和传统迁移学习之间的区别。之后，将介绍在文献[56]中提出的一种安全联邦迁移学习框架，并总结目前该领域面临的挑战和亟待解决的问题。

电子工业出版社

CHAPTER 7
联邦学习激励机制

在联邦学习中，如何建立激励机制使得参与方持续参与到数据联邦中是一项重要的挑战。实现这一目标的关键是制定一种奖励方法，公平公正地与参与方们分享联邦产生的利润。本章介绍联邦学习激励方法（Federated Learning Incentivizer，FLI）的各种激励分配方法[265]。它的任务目标为最大化联邦的可持续性经营，同时最小化参与方间的不公平性，动态地将给定的预算分配给联邦中的各个参与方，还可以扩展为一种能够帮助联邦抵御恶意的参与方的调节机制。

电子工业出版社

CHAPTER 8

联邦学习与计算机视觉、自然语言处理及推荐系统

本章将讨论将联邦学习技术应用于计算机视觉、自然语言处理及推荐系统，以实现面向隐私保护的人工智能应用。

电子工业出版社

CHAPTER 9
联邦强化学习

在本章中，我们介绍联邦强化学习 (Federated Reinforcement Learning)，涵盖强化学习、分布式强化学习、横向联邦强化学习和纵向联邦强化学习的基础知识，以及联邦强化学习的应用示例。

电子工业出版社

CHAPTER 10
应用前景

作为能够在不违反隐私和安全的前提下，使用分散于多方的数据来构建共享和定制化模型的一种创新的建模机制，联邦学习在诸多领域都有广阔的应用前景，如电子商务、金融、医疗、教育、城市计算、智慧城市、边缘计算、物联网、区块链以及第 5 代 (5G) 移动网络等。由于各种原因，这些领域的数据不能被直接地聚合用来训练机器学习模型。本章给出了几种已经落地或者富有潜力的、通过联邦学习技术实现的应用。

电子工业出版社

CHAPTER 11
总结与展望

人工智能和大数据的发展涉及它们对人类社会的社会责任这一重要议题。这需要我们兼顾到算法性能提升和对隐私、安全的严格要求。针对这一矛盾的挑战，联邦学习提供了一个有效的解决方案，使得我们既能利用多方数据源和模型提升系统的效率，同时也能保护用户隐私和信息的安全。本书系统地概述了联邦学习这一前沿领域的发展现状，也为这一技术未来的发展及应用描述了客观的前景。

电子工业出版社