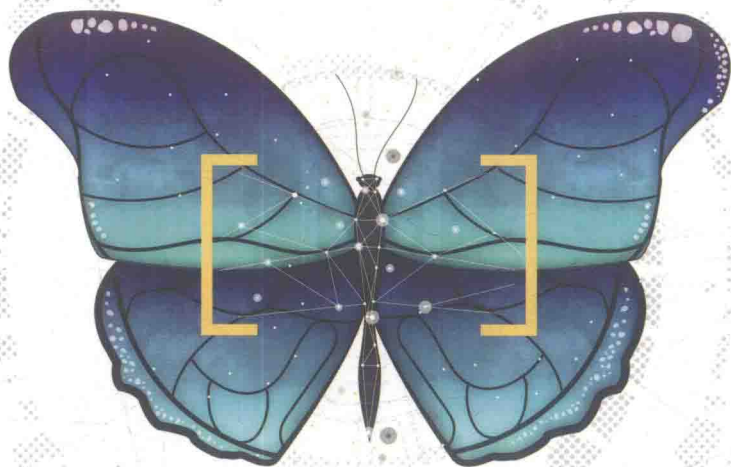


校宝在线董事长&CEO 张以弛
微软加速器(上海)CTO 俞榕
SegmentFault 合伙人&CTO 祁宁
南京理工大学信息管理系教授, 博士生导师 章成志

联合
推荐



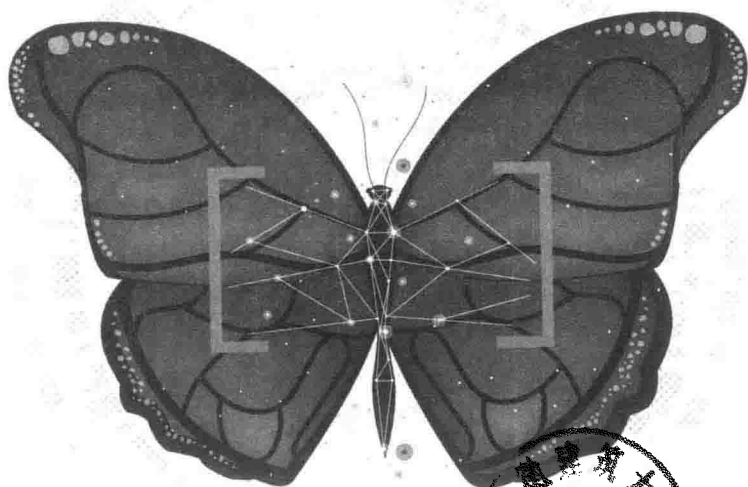
PyTorch 机器学习

从入门到实战

校宝在线 孙琳 蒋阳波 汪建成 项斌 编著



机械工业出版社
China Machine Press



PyTorch机器学习 从入门到实战

校宝在线

孙琳 蒋阳波 汪建成 项斌 编著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

PyTorch机器学习从入门到实战 / 校宝在线等编著. —北京: 机械工业出版社, 2018.10

ISBN 978-7-111-61045-8

I. ①P… II. ①校… III. ①机器学习 IV. ①TP181

中国版本图书馆CIP数据核字 (2018) 第226208号

近年来, 基于深度学习的人工智能掀起了一股热潮。本书是一本使用 PyTorch 深度学习框架的入门图书, 从深度学习原理入手, 由浅入深地阐述深度学习中的神经网络、深度神经网络、卷积神经网络、自编码器、循环神经网络等内容, 同时穿插 PyTorch 框架的知识点和基于知识点的实例, 最后综合运用 PyTorch 和深度学习知识来解决实践中的具体问题, 比如图像识别、文本分类和命令词识别等。可以说, 本书是深度学习和 PyTorch 的入门教程, 引导读者进入机遇和挑战并存的人工智能领域。

本书针对的是机器学习和人工智能的爱好者和研究者, 希望其有一定的机器学习和深度学习知识, 并有一定的 Python 编程基础。

PyTorch 机器学习从入门到实战

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: 夏非彼 迟振春

责任校对: 王 叶

印 刷: 中国电影出版社印刷厂

版 次: 2018 年 11 月第 1 版第 1 次印刷

开 本: 170mm × 242mm 1/16

印 张: 12.5

书 号: ISBN 978-7-111-61045-8

定 价: 59.00 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

序

人类自从蒙昧时代走出，经历了多次重大的技术革命，无论是以蒸汽机为代表的工业革命、以电力和内燃机为代表的第二次工业革命，还是以核能、信息技术、互联网技术为主体的第三次工业革命，都是对物质进行研究和处理的变革。蓬勃发展的 AI 时代与之有显著的区别。对于人的思维的研究和实践，AI 是真正意义上对思维进行拓荒的革命性技术变革。尽管现在的 AI 还算不上真正意义上的 AI，但是 AI 时代已经露出曙光。如今，如何占据 AI 时代的制高点已是各国政府和互联网巨头的重要课题。在商业的具体应用上，基于大数据的 AI 的优势越发凸显：在“阿里巴巴”，AI 算法基于用户购买的物品和行为进行数据挖掘，给用户画像，给用户推荐相关的商品；在“今日头条”，AI 算法根据用户阅读的文章，给用户推荐感兴趣的内容，越是用户感兴趣的领域，推荐的文章越多。

近期 AI 的爆发离不开大数据和高性能计算平台。自从互联网兴起后，数据累积的速度越来越快，一些专业领域如图像识别还出现了专门的数据集，如 ImageNet（这个数据集极大地促进了图像领域深度学习算法的进步）；同时，计算机的性能一直都呈指数级增长，特别是 GPU 领域日新月异的进步，提供了强大而高效的并行计算能力，减少了完成实验所需要的时间，大大加速了 AI 算法的进步。AI 爆发离不开算法的进步，而深度学习正是 AI 算法的主力军。那么，如何学习和使用 AI 算法呢？答案是：一个优秀的深度学习框架！PyTorch 作为一个新兴的深度学习框架，自提出之日起，便以简洁、优雅的开发语言，方便、快捷的构建模型以及易用的调试功能，赢得了广大开发者的青睐。可以说，PyTorch 是新手入门深度学习的最佳利器之一。但是目前基于 PyTorch 的深度学习书籍比较匮乏，中文资料少，体系零散，从算法入手阐述 PyTorch 框架的资料更是缺乏：要么侧重理论，对 PyTorch 的介绍深度不够；要么对 PyTorch 代码着眼太多，没

有相关原理的系统介绍，激发不了学习的兴趣。

“众里寻他千百度。蓦然回首，那人却在，灯火阑珊处。”读罢本书，我十分欣喜，这不正是一本适合读者入门和实践深度学习的书吗？此书理论和实践并重，深入浅出，循序渐进地讲述深度学习的基础理论，又结合具体问题给出 PyTorch 框架的解决方法，帮助读者一步步解决问题，体验 AI 算法的强大魅力，享受解决问题的喜悦和成就感。衷心地希望有志于 AI 学习大潮的莘莘学子，能够凭借此书开启深度学习之旅。“小荷才露尖尖角，早有蜻蜓立上头。”在 AI 时代来临之际，希望读者抓住机会，早做准备，成为 AI 潮流的真正弄潮儿。

浙江工商大学研究生院院长，教授，博士生导师

琚春华

2018年8月

前言

人工智能的发展日新月异，大学等研究机构和互联网巨头投入大量的经费和人力到这场没有硝烟的战争中，谁在这场天王山之战占据有利地位，谁就能在未来的竞争中一马当先。2016年3月，Google研发的AlphaGo与围棋世界冠军、职业九段棋手李世石进行了惊心动魄的围棋人机大战，并以4:1的比分赢得胜利。2017年3月，第二代的AlphaGo与柯洁在乌镇围棋峰会上的比赛中以3:0获胜。2017年10月，Google推出了最强版的AlphaGo——AlphaGo Zero，经过3天的自我训练就打败了第一版的AlphaGo，经过40天的自我训练打败了第二代的AlphaGo。2018年5月，Google在I/O大会上推出打电话的AI——Duplex，模仿真人的语气打电话，通过多轮对话，帮助用户完成餐馆预订和美发沙龙预约等。Google母公司董事长宣称，Duplex部分通过了图灵测试。（图灵测试被认为是考验机器是否拥有智能的测试：如果一个机器能在与人交流“沟通”的过程中不被识别出“机器身份”，那么这个机器就具有智能。）这个系统虽然离真正的人工智能尚远，但是这种人机交互技术对很多产业产生了深远的影响。这些影响深远的技术背后就是深度学习。

各大巨头尽力建立以深度学习框架为核心的AI生态系统。2017年年初，深度学习框架PyTorch横空出世。这个Facebook推出的框架是一个支持强大GPU加速的张量计算，构建基于Autograd系统的深度学习研究平台。其一面世，就以简洁优雅的接口、能够快速实现的代码和直观灵活且简单的网络结构给业界留下了深刻的印象。作为一个在2017年才诞生的深度学习框架，PyTorch相关学习文档和资料缺乏，而笔者在研究和实践的过程中进行了大量的深度学习模型构建和使用，对PyTorch简洁且灵活的编程风格深有体会，因此决定编写一本用PyTorch进行机器学习和深度学习入门的图书。

本书主要针对的是对深度学习有一定了解、希望用PyTorch进行机器学习和深度学习研究的初学者。阅读本书不需要太多的数学基础，但需要有一定的编程

基础，特别是要求有 Python 编程经验。希望读者学完本书后，能够对深度学习有大致地了解，基本掌握 PyTorch 的使用方法，知道如何根据基于 PyTorch 的神经网络模型来解决实际问题，并能够利用各种模型调参的方法进行模型优化。本书仅仅是一本入门的图书，要对深度学习进行深入研究的学习者，还要更加深入阅读相关专业书籍和学术论文。

本书从机器学习原理入手，延伸到神经网络，直至深度学习，由浅入深地阐述深度学习中的各个分支，即深度神经网络、卷积神经网络、自编码器、循环神经网络等，同时穿插 PyTorch 框架的知识点和基于知识点的实例。最后，综合运用 PyTorch 和深度学习理论来解决实践中的具体问题，比如文本分类和关键词识别等。可以说，本书是深度学习和 PyTorch 的入门教程，引导读者进入机遇和挑战共存的人工智能领域。

本书的代码开源在 GitHub 上，具体地址是 <https://github.com/xiaobaoonline/pytorch-in-action>。代码以章节划分文件夹，每个函数的作用和细节在代码中均有注释，以便帮助理解。本书的代码在 PyTorch0.3 上运行，由于 Python2 即将过时，因此本书代码只支持 Python3。大部分代码既支持 CPU 又支持 GPU，但第 8 章有部分代码只支持 GPU，读者在运行代码的过程中要注意相关提示。

由于编者水平有限，书中难免出现不太准确的地方，恳请读者批评指正。大家可以在 <https://github.com/xiaobaoonline/pytorch-in-action/issues> 处提出意见和反馈，让我们在机器学习之路上共同进步。

在本书写作的过程中，得到不少人的鼓励和支持。首先要感谢校宝在线（杭州）科技股份有限公司（证券代码：870705）上下的鼎力支持，特别是公司董事长兼 CEO 张以弛先生的大力支持，让我们在工作之余有足够的时间投入本书写作中。然后，感谢家人的鼓励和支持，没有他们，这本书的写作将不可能完成。除此之外，在写作和编码的过程中，还参考了很多书籍和资料，在此表示感谢。

编者

2018 年 7 月

目录

序

前言

第 1 章 深度学习介绍.....	1
1.1 人工智能、机器学习与深度学习.....	2
1.2 深度学习工具介绍.....	5
1.3 PyTorch 介绍.....	7
1.4 你能从本书中学到什么.....	9
第 2 章 PyTorch 安装和快速上手.....	11
2.1 PyTorch 安装.....	12
2.1.1 Anaconda 安装.....	12
2.1.2 PyTorch 安装.....	19
2.2 Jupyter Notebook 使用.....	19
2.3 NumPy 基础知识.....	22
2.3.1 基本概念.....	23
2.3.2 创建数组.....	24
2.3.3 基本运算.....	26
2.3.4 索引、切片和迭代.....	27
2.3.5 数组赋值.....	32
2.3.6 更改数组的形状.....	33
2.3.7 组合、拆分数组.....	34
2.3.8 广播.....	35

2.4	PyTorch 基础知识.....	37
2.4.1	Tensor 简介.....	37
2.4.2	Variable 简介.....	38
2.4.3	CUDA 简介.....	38
2.4.4	模型的保存与加载.....	39
2.4.5	第一个 PyTorch 程序.....	40
第 3 章	神经网络.....	43
3.1	神经元与神经网络.....	44
3.2	激活函数.....	46
3.2.1	Sigmoid.....	47
3.2.2	Tanh.....	48
3.2.3	Hard Tanh.....	49
3.2.4	ReLU.....	50
3.2.5	ReLU 的扩展.....	51
3.2.6	Softmax.....	54
3.2.7	LogSoftmax.....	55
3.3	前向算法.....	55
3.4	损失函数.....	57
3.4.1	损失函数的概念.....	57
3.4.2	回归问题.....	57
3.4.3	分类问题.....	58
3.4.4	PyTorch 中常用的损失函数.....	59
3.5	反向传播算法.....	62
3.6	数据的准备.....	65
3.7	PyTorch 实例：单层神经网络实现.....	66
第 4 章	深度神经网络及训练.....	70
4.1	深度神经网络.....	72
4.1.1	神经网络为何难以训练.....	72
4.1.2	改进策略.....	74
4.2	梯度下降.....	75
4.2.1	随机梯度下降.....	75
4.2.2	Mini-Batch 梯度下降.....	75

4.3	优化器	77
4.3.1	SGD	77
4.3.2	Momentum	77
4.3.3	AdaGrad	78
4.3.4	RMSProp	79
4.3.5	Adam	80
4.3.6	选择正确的优化算法	81
4.3.7	优化器的使用实例	82
4.4	正则化	85
4.4.1	参数规范惩罚	85
4.4.2	Batch Normalization	86
4.4.3	Dropout	87
4.5	PyTorch 实例：深度神经网络实现	89
第 5 章 卷积神经网络		93
5.1	计算机视觉	95
5.1.1	人类视觉和计算机视觉	95
5.1.2	特征提取	95
5.1.3	数据集	97
5.2	卷积神经网络	100
5.2.1	卷积层	102
5.2.2	池化层	104
5.2.3	经典卷积神经网络	105
5.3	MNIST 数据集上卷积神经网络的实现	110
第 6 章 嵌入与表征学习		114
6.1	PCA	115
6.1.1	PCA 原理	115
6.1.2	PCA 的 PyTorch 实现	116
6.2	自编码器	117
6.2.1	自编码器原理	118
6.2.2	PyTorch 实例：自编码器实现	118
6.2.3	PyTorch 实例：基于自编码器的图形去噪	122
6.3	词嵌入	125

6.3.1	词嵌入原理	125
6.3.2	PyTorch 实例：基于词向量的语言模型实现	128
第 7 章	序列预测模型	132
7.1	序列数据处理	133
7.2	循环神经网络	134
7.3	LSTM 和 GRU	138
7.4	LSTM 在自然语言处理中的应用	142
7.4.1	词性标注	142
7.4.2	情感分析	144
7.5	序列到序列网络	145
7.5.1	序列到序列网络原理	145
7.5.2	注意力机制	146
7.6	PyTorch 实例：基于 GRU 和 Attention 的机器翻译	147
7.6.1	公共模块	147
7.6.2	数据处理	147
7.6.3	模型定义	151
7.6.4	训练模块定义	155
7.6.5	训练和模型保存	161
7.6.6	评估过程	162
第 8 章	PyTorch 项目实战	165
8.1	图像识别和迁移学习——猫狗大战	166
8.1.1	迁移学习介绍	166
8.1.2	计算机视觉工具包	166
8.1.3	猫狗大战的 PyTorch 实现	167
8.2	文本分类	172
8.2.1	文本分类的介绍	173
8.2.2	计算机文本工具包	174
8.2.3	基于 CNN 的文本分类的 PyTorch 实现	174
8.3	语音识别系统介绍	182
8.3.1	语音识别介绍	182
8.3.2	命令词识别的 PyTorch 实现	183

第 1 章

◀ 深度学习介绍 ▶

围棋号称人类最复杂的棋类运动，但近两年来，在 AlphaGo 的冲击下，已经溃不成军。继 2016 年 AlphaGo 以 4:1 击败韩国李世石，2017 年 AlphaGo Master 以 3:0 零封柯洁后，最新的 Alpha Zero 在没有棋谱的情况下，进行 3 天的自我训练后，就击败了 AlphaGo；经过 40 天训练后，击败了 AlphaGo Master。在 AlphaGo 背后隐藏的知识就是近来发展如火如荼的深度学习。深度学习不仅在围棋领域大放异彩，在图像识别、语音识别、自然语言处理等领域也全面开花。人们对深度学习充满了渴望，向往到该领域学习和发展。几乎所有的计算机课程里面都会包含人工智能和深度学习。在各种媒体、学术和其他会议中，人工智能和深度学习也都是热门的话题。

开宗明义，本书第 1 章将简单扼要地介绍说明什么是人工智能及其历史的沿革，讲述人工智能研究的流派，以及最新发展起来的深度学习，阐明人工智能、机器学习和深度学习的联系和区别。接着介绍一个优秀的深度学习框架——PyTorch，阐述 PyTorch 这一年多来的发展历程及支持的厂商和研究机构。PyTorch 的 API 封装合理，具有几乎原生的 Python 使用体验，支持动态网络的构建，支持扩展等。正是由于 PyTorch 的优越特性，本书也通过 PyTorch 框架来进行深度学习的学习。随后介绍主流深度学习工具 TensorFlow、Theano、Keras、Caffe/Caffe2、MXNet 和 CNTK 以及它们的优缺点。最后，介绍本书的结构，方便学习者快速有效地使用本书，无论是初学者还是经验丰富的研究人员或者从业者，都能够从本书学到所需的知识，进入人工智能这个充满魔力和激情的领域。

1.1 人工智能、机器学习与深度学习

维基百科词条指出，人工智能（Artificial Intelligence），也称为机器智能，是指由人制造出来的机器所表现出的智能。通常人工智能是指通过普通计算机程序的手段实现的人类智能技术。广义上的人工智能，是让机器拥有完全的甚至超越人类的智能（General AI 或者 Strong AI）。计算机科学当中的研究更多聚焦在弱人工智能（Narrow AI 或者 Weak AI）上：人工智能是研究如何能让计算机模拟人类的智能，来实现特定的依赖人类智能才能完成的任务（例如学习、语言、识别）。

最早的人工智能探索可以追溯到 1818 年 Mary Shelly 对于复制人体的想象。计算机科学家先驱（例如 Alan Turing）早在 1950 年就完整提出了计算机智能的概念，并且提出了如何评估计算机是否拥有智能的测试——图灵测试，如图 1-1 所示。尽管有争议（例如 Chinese Room Test），这项测试今天依然被很多研究人员当作测试人工智能的一项重要标准。人工智能这个名字的正式提出来自于 1956 年的 Darmouth 会议。从 20 世纪 50 年代到 90 年代，人工智能在跌跌撞撞中前进，期间经历了两次人工智能寒冬。在 2000 年后，随着互联网时代的发展，人工智能领域拥有了期盼已久的大数据，也有了足够快的硬件处理能力的支持。在 2010 年深度学习出现之后，给人工智能领域带来了一场革命，大大加快了人工智能领域的研究。不过，这个领域的发展才刚刚起步，人工智能的未知世界，可能比计算机科学的其他领域（例如硬件、安全、系统）的总和还要多。

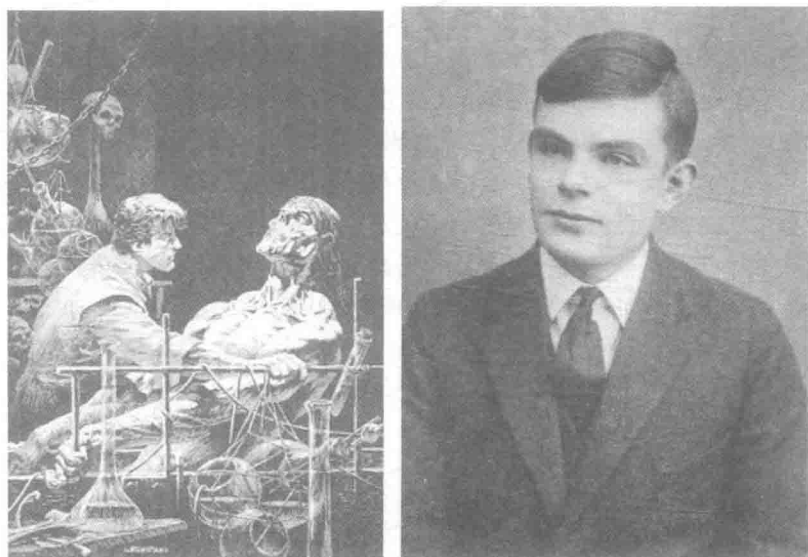


图 1-1 Mary Shelly 关于复制人体的想象及人工智能之父

人工智能的研究，分为多种学派，例如符号主义和连接主义等。每种学派所有的方法完全不同，甚至尝试解决的问题也不一样。符号主义源自于数理逻辑，目前依然是主流学派之一（例如知识图谱）。机器学习则是连接主义的产物，机器学习中最早的模型是源自于人脑的仿生学，我们书中会讲到的感知机（Perceptron）就是对于人脑单一神经元的模拟。基于统计学的方法，例如决策树、支撑向量机、逻辑回归等也被归于此类算法当中。从广义上讲，机器学习是从数据中自动学习算法，并且使用学习到的算法去进行预测。机器学习所学习到的算法，是自动从数据中学习的，随着数据改变也会有本质上的改变。这与传统计算机科学中面对数据提前编程有本质区别。

深度学习是使用深度神经网络来实现机器学习的方法。如图 1-2 所示，我们可以看到人工智能、机器学习和深度学习这三者的关系。深度神经网络跟人脑的模型有相似性，人脑中每个神经元之间都是可以连接的，而深度神经网络中，神经元往往被分为了很多层，从而有了深度。拿图像识别（如图 1-3 所示）来举例子，图像可以被裁剪成很多小块，然后输入到神经网络的第一层，接着第一层再向后面的层传导，每层做不同的任务，然后最后一层完成预测。每个神经元对于输入都有权重，这些权重被用来计算最后的输出。例如图像识别的例子，有些神经元被用来识别颜色，有些被用来识别形状，最后一层的神经元对于所有权重进行总结，最终做出预测。

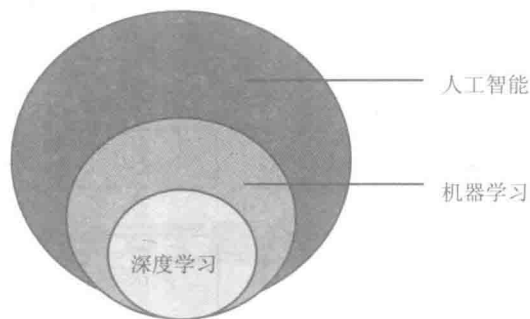


图 1-2 人工智能、机器学习和深度学习的关系



图 1-3 识别猫是深度神经网络最早的成功之一

深度学习在所有人工智能领域当中都取得了成功。以 ImageNet 图像分类任务举例，深度学习自 2012 年以来早已远远超过了传统的机器学习算法，在 2014 年之后甚至超过了人类的正确率，如图 1-4 所示。（蓝色为非深度学习，紫色为深度学习，红色为人类。）



图 1-4 ImageNet 图像分类错误率

为什么深度学习能获得成功呢？主要是由于以下三个因素的影响。

第一个因素是深度学习神经网络增加了学习能力。例如，传统的规则系统，100%依赖于人工规则。而机器学习模型，往往要依赖于少量人工设计的特征，深度学习模型则几乎不依赖任何人工定义的特征和规则，如图 1-5 所示。



图 1-5 规则、机器学习、深度学习比较

1. Theano

Theano 是加拿大蒙特利尔大学 LISA 实验室在 2008 年提出的深度学习框架，是一个 Python 库，可用于定义、优化和计算数学表达式，特别是多维数组。Theano 结合了计算机代数系统（Computer Algebra System）和优化编译器，能够为多种数学运算符生成定制的 C 语言代码，同时，它还支持 GPU 加速，为早期的深度学习研究立下了汗马功劳。Theano 在深度学习框架中是祖师级的存在，但它诞生于研究机构，学术气息浓厚，工程设计存在缺陷。在 2017 年 9 月，在 Theano 10.0 发布之际，LISA 实验室负责人，深度学习三巨头之一的 Yoshua Bengio 宣布 Theano 停止开发。尽管 Theano 已经完成了使命，但它为深度学习的早期研究提供了极大的帮助，同时也为后来的深度学习框架开发奠定了方向：以计算图为框架的核心，采用 GPU 加速计算。

2. TensorFlow

2015 年 11 月，Google 推出了机器学习开源工具 TensorFlow。TensorFlow 是由 Google Brain 团队开发的，主要用于机器学习和深度神经网络研究。同时，它也是一个基础系统，能够应用于其他领域。TensorFlow 使用 C++ 作为开发语言，使用计算流图的形式进行计算。图中节点表示数学运算，而图中的线条表示 Tensor 之间的交互。TensorFlow 对开发不是很友好，但是方便部署，不仅可以在 CPU 和 GPU 上运行，还可以在台式机、服务器、移动设备上运行。

在 Google 的强大号召力下，加上支持各种语言和硬件，TensorFlow 是目前流行的深度学习框架，在工业上应用广泛，有强大的开发者社区。但是 Tensorflow 的系统设计过于复杂，对于初学者来讲，学习曲线有点陡峭。同时，Tensorflow 作为静态图框架，不太方便直接调试，打印中间结果必须借助 Session 运行才能生效，或者学习额外的 tfdbg 工具。

3. Caffe/Caffe2

Caffe 的全称是 Convolutional Architecture for Fast Feature Embedding，是加州大学伯克利分校的贾扬清开发的，目前由伯克利视觉中心维护。这是一个清晰、高效的深度学习框架，核心语言是 C++，支持命令行、Python 和 MATLAB 接口，在 CPU、GPU 均可运行。Caffe2 沿袭了大量的 Caffe 设计，并解决了 Caffe 在使用和部署上发现的问题。Caffe2 能够提供速度和便携性，其 Python 库和 C++ API 使用户在 Linux、Windows、iOS、Android，甚至 Raspberry 和 Nvidia Tegra 上进行原型设计、训练和部署。由于 Caffe2 对全平台的支持，适合工业部署。

4. MXNet

MXNet 是由李沐和陈天奇等人开发的深度学习库。2016 年，成为亚马逊云计算的官方深度学习平台。MXNet 支持 C++、Python、R、Scala、Julia、MATLAB