

深度学习

核心技术与实践

猿辅导研究团队 著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

深度学习

深度学习



深度学习

核心技术与应用

——

清华大学出版社

深度学习 核心技术与实践

猿辅导研究团队 著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书主要介绍深度学习的核心算法，以及在计算机视觉、语音识别、自然语言处理中的相关应用。本书的作者们都是业界一线的深度学习的从业者，所以书中所写内容和业界联系紧密，所涵盖的深度学习相关知识点比较全面。本书主要讲解原理，较少贴代码。

本书适合深度学习从业人士或者相关研究生作为参考资料，也可以作为入门教程来大致了解深度学习的相关前沿技术。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目 (C I P) 数据

深度学习核心技术与实践 / 猿辅导研究团队著. —北京: 电子工业出版社, 2018.2
(博文视点 AI 系列)

ISBN 978-7-121-32905-0

I. ①深…II. ①猿…III. ①机器学习 IV. ①TP181

中国版本图书馆 CIP 数据核字 (2017) 第 258324 号

策划编辑: 张春雨

责任编辑: 葛 娜

印 刷: 三河市良远印务有限公司

装 订: 三河市良远印务有限公司

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: 787×980 1/16 印张: 33 字数: 718 千字

版 次: 2018 年 2 月第 1 版

印 次: 2018 年 5 月第 3 次印刷

定 价: 119.00 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 zlt@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。

前言

本书的大部分作者在深度学习流行之前有幸从事机器学习相关工作多年。在我们内部，一直认同一个段子：有多少人工就有多少智能。

- 在深度学习流行之前的传统机器学习年代，我们认为“人工”更多强调的是特征工程之难，需要机器学习从业者不断分析数据，挖掘新的特征。
- 在深度学习流行的这几年，我们认为这句话依然成立，只是“人工”更多地强调人工标注，因为深度学习需要大量的标注数据。当然，也有人反驳说不需要标注，用户的使用历史天然就是标注。实际上，这可以理解作为一种众筹标注。
- 在深度学习发展的未来，我们希望这句话不再成立，期待无监督模型取得更长足的进步，使得“人工”智能变为真正的智能。

在追求智能的路上，我们虽然是创业公司，但一直坚持机器学习相关课程的学习和 Paper Reading，陆续学习了传统的机器学习相关算法，也探索了深度学习的相关原理，并不断应用到实践中。

受益于当今学术开放开源的氛围，深度学习的最新算法甚至代码实践大家都能在第一时间进行学习。所以在创业公司的早期深度学习实践中，最重要的并不是算法理论方面的创新，而是结合产品需求如何进行深度学习技术的落地。这就要求团队不仅需要对业务非常熟悉，也需要对深度学习相关算法了如指掌，同时还需要有人可以真正用代码将算法落地。很幸运，我们的团队具备这样的能力，所以在深度学习的实践中较少走弯路。随着多年的积累，团队在深度学习方面开始有不少自己的创新，也对理论有了整体的认识。从 2016 年下半年开始，团队部分成员利用周末等业余时间撰写了这本书，算是对团队过去所学深度学习知识的一个总结。本书的撰写都是大家牺牲周末时间完成的，且在撰写过程中，碰到多次项目进度非常紧急的情况，周末时间也被项目占用，但大家还是克服困难，完成了书稿，非常感谢这些作者的配合！此外，猿辅导研究团队的大部分成员参与了审稿相关工作，在此一并表示感谢！

当然，本书撰写较仓促，作者人数也较多，错误和不足在所难免，烦请读者及时反馈，我们将及时纠正。

在这个过程中，有了一点点微不足道的积累。希望通过本书，对过去学过的知识做一些总结归纳，同时分享出来让更多的深度学习爱好者一起受益。

写作分工

朱珊珊编写了第1章的1.2.1节主要部分、1.3节，第2章的绝大部分内容，第13章。

邓澍军编写了前言，第1章的1.1节、1.2.2节至1.2.4节，第2章的2.2.2节、2.2.6节至2.2.8节，第3章，第6章的6.1节、6.2节，第7章，第8章的8.1节，第9章，第10章的10.6节、10.7节，第11、17、18、21章，第25章的25.3节。

陈孟阳编写了第4、10章。

孙萌编写了第5、22章。

冯超编写了第6章的6.3节至6.10节，第8章的8.3节、8.5节至8.7节，第27、28章。

曹月恬编写了第8章的8.2节、8.4节，第24章。

杨晓庆编写了第12、26章。

夏龙编写了第14、15章，第16章的16.1节、16.2节、16.5节、16.6节。

吴凡编写了第16章的16.3节、16.4节。

赵薇编写了第19章。

陈冬晓编写了第20章。

赵玲玲编写了第23章。

王锐坚编写了第25章。

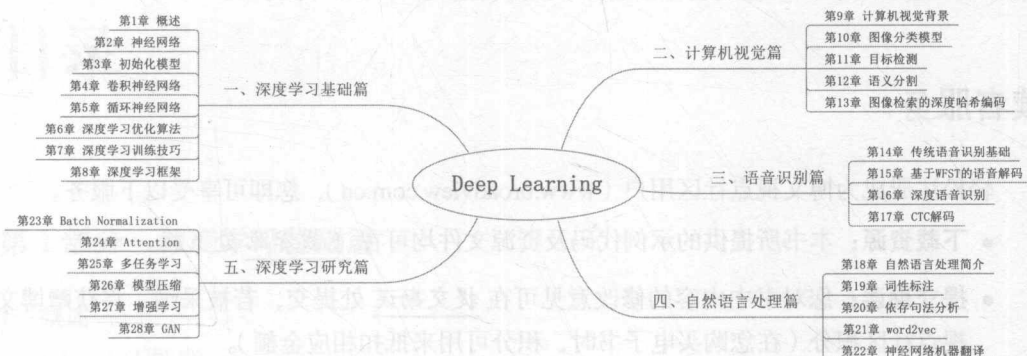
本书特点

本书首先介绍了深度学习的一些基本原理，然后介绍了计算机视觉、语音识别、自然语言处理的相关应用，最后介绍了一些较前沿的研究方向。

本书具有如下特点：

- 计算机视觉、语音识别、自然语言处理这三方面的介绍内容绝大部分是作者团队有过相关实践和研究的方向，和业界联系紧密。
- 所涵盖的深度学习相关知识点比较全面。
- 干货：主要讲解原理，较少贴代码。

本书的篇章脉络如下：



本书读者

本书适合深度学习从业人士或者相关研究生作为参考资料，也可以作为入门教程大致了解深度学习的相关前沿技术。

关于团队

猿辅导研究团队成立于2014年，是创业公司中较早从事深度学习的团队。该团队陆续将深度学习应用于如下领域：

- 拍照印刷体 OCR (Optical Character Recognition, 光学字符识别)：从0开始打造拍照搜题 APP 小猿搜题（目前累计安装量达1.6亿次）。
- 拍照手写体 OCR：包括斑马速算产品中的屏幕手写笔迹的在线手写识别、拍照手写图片的离线手写识别、与公务员考试相关的申论手写识别等。
- 语音识别：包括古诗词背诵、高考听说自动判卷、英语口语打分等项目。
- 自然语言处理：主要应用于英语作文自动批改、自动判卷、短文本对话等项目。

关于公司

猿辅导公司是中国领先的移动在线教育机构，拥有中国最多的中学生移动用户，以及国内最大的中学生练习行为数据库，旗下有猿题库、小猿搜题、猿辅导三款移动教育 APP。

2017年6月猿辅导获得由华平投资集团领投、腾讯跟投的1.2亿美元E轮融资，估值超过10亿美元，成为国内K-12在线教育领域首个独角兽公司。

读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），您即可享受以下服务。

- **下载资源：**本书所提供的示例代码及资源文件均可在 [下载资源](#) 处下载。
- **提交勘误：**您对书中内容的修改意见可在 [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）。
- **与作者交流：**在页面下方 [读者评论](#) 处留下您的疑问或观点，与作者和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/32905>



目录

第 1 部分 深度学习基础篇	1
1 概述	2
1.1 人工智能	3
1.1.1 人工智能的分类	3
1.1.2 人工智能发展史	3
1.2 机器学习	7
1.2.1 机器学习的由来	7
1.2.2 机器学习发展史	9
1.2.3 机器学习方法分类	10
1.2.4 机器学习中的基本概念	11
1.3 神经网络	12
1.3.1 神经网络发展史	13
参考文献	16
2 神经网络	17
2.1 在神经科学中对生物神经元的研究	17
2.1.1 神经元激活机制	17
2.1.2 神经元的特点	18
2.2 神经元模型	19
2.2.1 线性神经元	19
2.2.2 线性阈值神经元	19
2.2.3 Sigmoid 神经元	21
2.2.4 Tanh 神经元	22
2.2.5 ReLU	22
2.2.6 Maxout	24
2.2.7 Softmax	24

2.2.8	小结	25
2.3	感知机	27
2.3.1	感知机的提出	27
2.3.2	感知机的困境	28
2.4	DNN	29
2.4.1	输入层、输出层及隐层	30
2.4.2	目标函数的选取	30
2.4.3	前向传播	32
2.4.4	后向传播	33
2.4.5	参数更新	35
2.4.6	神经网络的训练步骤	36
	参考文献	36
3	初始化模型	38
3.1	受限玻尔兹曼机	38
3.1.1	能量模型	39
3.1.2	带隐藏单元的能量模型	40
3.1.3	受限玻尔兹曼机基本原理	41
3.1.4	二值 RBM	43
3.1.5	对比散度	45
3.2	自动编码器	47
3.2.1	稀疏自动编码器	48
3.2.2	降噪自动编码器	48
3.2.3	栈式自动编码器	49
3.3	深度信念网络	50
	参考文献	52
4	卷积神经网络	53
4.1	卷积算子	53
4.2	卷积的特征	56
4.3	卷积网络典型结构	59
4.3.1	基本网络结构	59
4.3.2	构成卷积神经网络的层	59
4.3.3	网络结构模式	60

4.4	卷积网络的层	61
4.4.1	卷积层	61
4.4.2	池化层	66
	参考文献	67
5	循环神经网络	68
5.1	循环神经网络简介	68
5.2	RNN、LSTM 和 GRU	69
5.3	双向 RNN	75
5.4	RNN 语言模型的简单实现	76
	参考文献	79
6	深度学习优化算法	80
6.1	SGD	80
6.2	Momentum	81
6.3	NAG	82
6.4	Adagrad	84
6.5	RMSProp	85
6.6	Adadelta	86
6.7	Adam	87
6.8	AdaMax	89
6.9	Nadam	89
6.10	关于优化算法的使用	91
	参考文献	91
7	深度学习训练技巧	93
7.1	数据预处理	93
7.2	权重初始化	94
7.3	正则化	95
7.3.1	提前终止	95
7.3.2	数据增强	95
7.3.3	L2/L1 参数正则化	97
7.3.4	集成	99
7.3.5	Dropout	100

参考文献	101
8 深度学习框架	102
8.1 Theano	102
8.1.1 Theano	102
8.1.2 安装	103
8.1.3 计算图	103
8.2 Torch	104
8.2.1 概述	104
8.2.2 安装	105
8.2.3 核心结构	106
8.2.4 小试牛刀	109
8.3 PyTorch	112
8.3.1 概述	112
8.3.2 安装	112
8.3.3 核心结构	113
8.3.4 小试牛刀	113
8.4 Caffe	116
8.4.1 概述	116
8.4.2 安装	117
8.4.3 核心组件	118
8.4.4 小试牛刀	124
8.5 TensorFlow	124
8.5.1 概述	124
8.5.2 安装	124
8.5.3 核心结构	125
8.5.4 小试牛刀	126
8.6 MXNet	130
8.6.1 概述	130
8.6.2 安装	130
8.6.3 核心结构	130
8.6.4 小试牛刀	132
8.7 Keras	134

8.7.1	概述	134
8.7.2	安装	135
8.7.3	模块介绍	135
8.7.4	小试牛刀	135
	参考文献	138
第 2 部分 计算机视觉篇		139
9	计算机视觉背景	140
9.1	传统计算机视觉	140
9.2	基于深度学习的计算机视觉	144
9.3	参考文献	145
10	图像分类模型	146
10.1	LeNet-5	146
10.2	AlexNet	148
10.3	VGGNet	153
10.3.1	网络结构	154
10.3.2	配置	156
10.3.3	讨论	156
10.3.4	几组实验	157
10.4	GoogLeNet	158
10.4.1	NIN	160
10.4.2	GoogLeNet 的动机	160
10.4.3	网络结构细节	161
10.4.4	训练方法	163
10.4.5	后续改进版本	164
10.5	ResNet	164
10.5.1	基本思想	164
10.5.2	网络结构	166
10.6	DenseNet	168
10.7	DPN	169
	参考文献	169

11 目标检测	172
11.1 相关研究	174
11.1.1 选择性搜索	174
11.1.2 OverFeat	176
11.2 基于区域提名的方法	178
11.2.1 R-CNN	178
11.2.2 SPP-net	180
11.2.3 Fast R-CNN	181
11.2.4 Faster R-CNN	183
11.2.5 R-FCN	184
11.3 端到端的方法	185
11.3.1 YOLO	185
11.3.2 SSD	186
11.4 小结	187
参考文献	189
12 语义分割	191
12.1 全卷积网络	192
12.1.1 FCN	192
12.1.2 DeconvNet	194
12.1.3 SegNet	196
12.1.4 DilatedConvNet	197
12.2 CRF/MRF 的使用	198
12.2.1 DeepLab	198
12.2.2 CRFasRNN	200
12.2.3 DPN	202
12.3 实例分割	204
12.3.1 Mask R-CNN	204
参考文献	205
13 图像检索的深度哈希编码	207
13.1 传统哈希编码方法	207
13.2 CNNH	208
13.3 DSH	209

13.4 小结	211
参考文献	211
第 3 部分 语音识别篇	213
14 传统语音识别基础	214
14.1 语音识别简介	214
14.2 HMM 简介	215
14.2.1 HMM 是特殊的混合模型	217
14.2.2 转移概率矩阵	218
14.2.3 发射概率	219
14.2.4 Baum-Welch 算法	219
14.2.5 后验概率	223
14.2.6 前向-后向算法	223
14.3 HMM 梯度求解	226
14.3.1 梯度算法 1	227
14.3.2 梯度算法 2	229
14.3.3 梯度求解的重要性	233
14.4 孤立词识别	233
14.4.1 特征提取	233
14.4.2 孤立词建模	234
14.4.3 GMM-HMM	236
14.5 连续语音识别	239
14.6 Viterbi 解码	242
14.7 三音素状态聚类	244
14.8 判别式训练	247
参考文献	253
15 基于 WFST 的语音解码	255
15.1 有限状态机	256
15.2 WFST 及半环定义	256
15.2.1 WFST	256
15.2.2 半环 (Semiring)	257

15.3	自动机操作	259
15.3.1	自动机基本操作	260
15.3.2	转换器基本操作	261
15.3.3	优化操作	264
15.4	基于 WFST 的语音识别系统	276
15.4.1	声学模型 WFST	278
15.4.2	三音素 WFST	280
15.4.3	发音字典 WFST	280
15.4.4	语言模型 WFST	281
15.4.5	WFST 组合和优化	283
15.4.6	组合和优化实验	284
15.4.7	WFST 解码	285
	参考文献	286
16	深度语音识别	287
16.1	CD-DNN-HMM	287
16.2	TDNN	291
16.3	CTC	294
16.4	EESSEN	298
16.5	Deep Speech	300
16.6	Chain	309
	参考文献	312
17	CTC 解码	314
17.1	序列标注	314
17.2	序列标注任务的解决办法	315
17.2.1	序列分类	315
17.2.2	分割分类	316
17.2.3	时序分类	317
17.3	隐马模型	317
17.4	CTC 基本定义	318
17.5	CTC 前向算法	320
17.6	CTC 后向算法	323
17.7	CTC 目标函数	324

17.8 CTC 解码基本原理	326
17.8.1 最大概率路径解码	326
17.8.2 前缀搜索解码	327
17.8.3 约束解码	328
参考文献	332
第 4 部分 自然语言处理篇	333
18 自然语言处理简介	334
18.1 NLP 的难点	334
18.2 NLP 的研究范围	335
19 词性标注	337
19.1 传统词性标注模型	337
19.2 基于神经网络的词性标注模型	339
19.3 基于 Bi-LSTM 的神经网络词性标注模型	341
参考文献	343
20 依存句法分析	344
20.1 背景	345
20.2 SyntaxNet 技术要点	347
20.2.1 Transition-based 系统	348
20.2.2 “模板化”技术	352
20.2.3 Beam Search	354
参考文献	356
21 word2vec	357
21.1 背景	358
21.1.1 词向量	358
21.1.2 统计语言模型	358
21.1.3 神经网络语言模型	361
21.1.4 Log-linear 模型	363
21.1.5 Log-bilinear 模型	364
21.1.6 层次化 Log-bilinear 模型	364