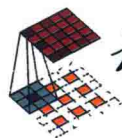


深度学习
轻松学



核心算法与视觉实践

冯超 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

轻松学

核心算法与视觉实践



冯超 (著)

电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书介绍了深度学习基本算法和视觉领域的应用实例。书中以轻松直白的语言，生动详细地介绍了深层模型相关的基础知识，并深入剖析了算法的原理与本质。同时，书中还配有大量案例与源码，帮助读者切实体会深度学习的核心思想和精妙之处。除此之外，书中还介绍了深度学习在视觉领域的应用，从原理层面揭示其思路思想，帮助读者在此领域中夯实技术基础。

本书十分适合对深度学习感兴趣，希望对深层模型有较深入了解的读者阅读。

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 (CIP) 数据

深度学习轻松学：核心算法与视觉实践 / 冯超著. —北京：电子工业出版社，2017.7

ISBN 978-7-121-31713-2

I. ①深…II. ①冯…III. ①人工智能①机器学习 IV. ①TP18

中国版本图书馆 CIP 数据核字 (2017) 第 121113 号

策划编辑：郑柳洁

责任编辑：郑柳洁

印 刷：北京季蜂印刷有限公司

装 订：北京季蜂印刷有限公司

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：720×1000 1/16 印张：21.75 字数：520 千字

版 次：2017 年 7 月第 1 版

印 次：2017 年 7 月第 1 次印刷

定 价：79.00 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 zltz@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

从你拿起这本书的那一刻开始，我就和你一起踏上了这段有关深度学习的冒险之旅。本书中有简单直白的叙述，也有复杂冗长的证明；有调皮的调侃，也有深刻的沉思。所有的一切，都是为了帮助你更轻松地对深度学习有更多了解、更多感悟。

本书的前身是我在知乎上的《无痛的机器学习》专栏（<https://zhuanlan.zhihu.com/hsmyy>）。作为在路上前进的一枚“小白”，我一直有写读书笔记的习惯。早期，我会把笔记写在阅读的文献上，在这些篇章的关键处做标记、写好注释。时间一长，当我想再读这些曾经阅读过的文献时，那些注释竟然变得莫名其妙起来。当初难以理解的概念又回到难以理解的状态，于是我不得不再次阅读原文，重新获得其中的感悟，这样周而复始，一次次在这个循环中打转。

后来，我选择彻底甩掉曾经阅读过的文献——既然它给我带来了那么多痛苦，为什么每次回顾时还要与它相对？于是我开始尝试将文献的所有内容重新组织成一篇短小的文章，文章中列出文献中的关键内容。这种文章简明扼要、直击重点，看着很清爽，可以帮助我快速回顾阅读的内容，又不用再回到原文，所以这种方法很快就替代了以前的那种方法。

再后来，我发现了新的问题。虽然我摆脱了晦涩的论文内容，但摆脱不了自己重述的文字。尤其是这些浓缩后的“精华”文字，理解起来并不比原始文献强多少，而且由于缺少很多细节，一旦遇到理解困难的状况，我不得不回到原文搜寻答案，于是这样的痛苦又得经历一次。

这时，我想起了父亲曾对我的教导：“书要先读厚，然后再读薄”。所谓的读厚，就是给阅读的书增加更多的内容，例如注释、自己的理解等。当自己能完全掌握书中的内容时，再将书中的细枝末节慢慢砍掉，将主体思想保留在心中，这就是读薄的过程。这样就完成了对一本书的深入阅读，其中读薄的过程和《倚天屠龙记》中张三丰教授张无忌太极拳剑的过程很相似。如果站在信息论的角度看，所谓的“重意不重形”，可以看作保证低频信息的正确性，依概率分布生成高频信息的一种“招式”生成模式。能达到这等境界，方可谓之大成。

对我来说，面对潮水般涌来的深度学习知识，想淡定从容应对是不可能的。也就是说，一开始就把书读薄是不可能的。所以，浓缩版的总结文章也慢慢被我否定了。那么

只剩下一个选择了，就是把书读厚，把文章写得拖沓冗长一些。于是我开始尝试用拖沓但细致的语言做文献的总结，经过一次次地尝试，文章的易读性确实有了明显提升，因为文章描述不清楚而回看原文的次数也越来越少。与此同时，我发现写作这个过程让我对问题有了更深入的理解。

拖沓式写法通过写作驱动，更容易发现知识点与知识点之间的沟壑。要想详细描述自己阅读的内容，就得确定文章中的逻辑自己是否真的明白了。有时探究这些逻辑所花费的时间远比自己想象得长，但是它确实让我感受到了进步与成长。

渐渐地，拖沓式文章越写越多，我也逐渐将其中一些文章公开，希望能够与大家分享：希望大家能和我对同一个问题产生共鸣，同时也可以在阅读中指出文章中的错误。不到一年，我见证了我写的文章被越来越多的人关注、讨论，在交流的过程中我收获了又一次成长。

在完成了几十篇文章后，本书的编辑郑柳洁女士联系到我，问我是否有兴趣将这些文章集结，变成一本出版印刷的书。在此之前我并没有仔细想过这件事儿，但是她的建议让我觉得这也许是又一次证明自己、使自己成长的机会，而且出版书籍在我心中也是一件神圣的事儿，于是我接受了这个建议，开始了整理书稿的工作。

整理书稿并没有想象中那么简单。网上的文章都是单独出现的，而书中的文章需要有一定的整体性；网上的文章可以用比较随意的语言叙述，而书中的语言需要尽量正式、客观。这些挑战使我修改了放在网络上的很多文字，为了确保表达清晰准确、语言通顺流畅，有些文章基本被重新写了一遍。整理这些文章花费了很多业余时间，但功夫不负有心人，这项工作被一点点地完成了。

本书主要介绍了深度学习，尤其是卷积神经网络相关的基础知识，以及计算机视觉的部分应用内容。书中既包含深度学习中的基础知识，也包含部分进阶内容，同时也包含一些较新颖的概念与模型，适合不同人群阅读。对初学者来说，本书十分详细地介绍了很多基本概念，对自学入门深度学习很有帮助；对有一定经验的从业人员来说，本书可以梳理领域内的知识点，也可以作为工具书使用。

本书中的示例代码由 C++ 和 Python 两种语言实现，读者只要对两种语言有基本的了解即可。本书主要使用了 Caffe 这个第三方开源框架，在此向 Caffe 的作者表示感谢。现在有很多优秀的开源框架，这些框架各有优劣，但本质上有许多共性。对没有使用过 Caffe 的读者来说，阅读本书时不会有大的困扰。

最后聊聊本书的书名。最初想使用“无痛”这个词，无奈这个词太容易引发联想，不适合作为一本严肃的出版物的书名。“学习”这个词自古以来就不是轻松的代名词。“学习”的“习”字曾写作“習”，意思为鸟类挥动翅膀一次次试飞，其中暗含了反复练习的过程；在日语中“学习”被写作“勉強する”，从大家能看懂的两个字就可以看出，学

习中的艰辛；在韩语中“学习”又被写作“공부하다”，它前两个字的发音和“恐怖”二字很像，也许当初这个词背后的含义和恐怖有关。这样看来，东北亚的这几个国家都学习这件事都充满了敬畏之心，学习这件事是绝对不会轻松的，更不会是无痛的，经历过多年教育的读者相信也会有体会。

学习的过程充满痛苦，而这是不可避免的。我们每个人在人生的道路上都有过不断探索、不断遇到挫折，然后改进自我，完成进化的体验。就像机器学习中被训练的模型一样，不断地完成预测、找到差距、更新参数，最终达到最优。在这条自我成长的道路上，每一次的失败、每一次的努力都充满了艰辛，然而没有痛苦的积累，就不会有最终快乐的进发。

人生苦短。虽然人生要经历许多的痛苦，但我们的目标并不是痛苦，而是走向彼岸的快乐。因此，如果能有减少痛苦的良方，大家还是愿意一试。社会的不断发展进步就是探寻减少痛苦的解药，而本书的写作目标也是希望本书能成为各位读者学习路上的解药，在减少痛苦的同时实现心中的目标。

感谢邓澍军、夏龙先生在本书内容创作期间对内容的严格审核并提出宝贵的意见和建议；感谢知乎上各位为《无痛的机器学习》专栏中系列文章指出错误、提出疑问的朋友，是你们让文章变得更严谨；由于本人才疏学浅，行文间难免有所纰漏，望各位读者多多包涵，不吝赐教。最后感谢所有关心、支持我完成这件不易的工作的亲人和朋友。我爱你们！

作者

阅读须知

本书图片

部分图片可能需要放大观察，纸面上无法呈现应有效果。为此，书中图片均在博文视点官方网站提供下载。

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/31713>



目录

1	机器学习与深度学习的概念	1
1.1	什么是机器学习	1
1.1.1	1.1.1 机器学习的形式	2
1.1.2	1.1.2 机器学习的几个组成部分	8
1.2	1.2 深度学习的逆袭	9
1.3	1.3 深层模型在视觉领域的应用	13
1.4	1.4 本书的主要内容	15
1.5	1.5 总结	17
2	数学与机器学习基础	18
2.1	2.1 线性代数基础	18
2.2	2.2 对称矩阵的性质	22
2.2.1	2.2.1 特征值与特征向量	22
2.2.2	2.2.2 对称矩阵的特征值和特征向量	23
2.2.3	2.2.3 对称矩阵的对角化	24
2.3	2.3 概率论	25
2.3.1	2.3.1 概率与分布	25
2.3.2	2.3.2 最大似然估计	28
2.4	2.4 信息论基础	31
2.5	2.5 KL 散度	33
2.6	2.6 凸函数及其性质	37
2.7	2.7 机器学习基本概念	39
2.8	2.8 机器学习的目标函数	42
2.9	2.9 总结	44

3 CNN 的基石：全连接层	45
3.1 线性部分	45
3.2 非线性部分	48
3.3 神经网络的模样	50
3.4 反向传播法	55
3.4.1 反向传播法的计算方法	55
3.4.2 反向传播法在计算上的抽象	58
3.4.3 反向传播法在批量数据上的推广	59
3.4.4 具体的例子	63
3.5 参数初始化	65
3.6 总结	68
4 CNN 的基石：卷积层	69
4.1 卷积操作	69
4.1.1 卷积是什么	69
4.1.2 卷积层效果展示	73
4.1.3 卷积层汇总了什么	76
4.1.4 卷积的另一种解释	77
4.2 卷积层的反向传播	79
4.2.1 实力派解法	80
4.2.2 “偶像派”解法	84
4.3 ReLU	88
4.3.1 梯度消失问题	89
4.3.2 ReLU 的理论支撑	92
4.3.3 ReLU 的线性性质	93
4.3.4 ReLU 的不足	93
4.4 总结	94
4.5 参考文献	94

5 Caffe 入门	95
5.1 使用 Caffe 进行深度学习训练	96
5.1.1 数据预处理	96
5.1.2 网络结构与模型训练的配置	100
5.1.3 训练与再训练	108
5.1.4 训练日志分析	110
5.1.5 预测检验与分析	112
5.1.6 性能测试	115
5.2 模型配置文件介绍	117
5.3 Caffe 的整体结构	122
5.3.1 SyncedMemory	124
5.3.2 Blob	125
5.3.3 Layer	125
5.3.4 Net	126
5.3.5 Solver	126
5.3.6 多 GPU 训练	127
5.3.7 IO	127
5.4 Caffe 的 Layer	128
5.4.1 Layer 的创建——LayerRegistry	128
5.4.2 Layer 的初始化	130
5.4.3 Layer 的前向计算	132
5.5 Caffe 的 Net 组装流程	133
5.6 Caffe 的 Solver 计算流程	139
5.6.1 优化流程	140
5.6.2 多卡优化算法	142
5.7 Caffe 的 Data Layer	145
5.7.1 Datum 结构	145
5.7.2 DataReader Thread	147
5.7.3 BasePrefetchingDataLayer Thread	148

5.7.4	Data Layer	149
5.8	Caffe 的 Data Transformer	150
5.8.1	C++ 中的 Data Transformer	150
5.8.2	Python 中的 Data Transformer	153
5.9	模型层扩展实践——Center Loss Layer	156
5.9.1	Center Loss 的原理	156
5.9.2	Center Loss 实现	160
5.9.3	实验分析与总结	164
5.10	总结	165
5.11	参考文献	165
6	深层网络的数值问题	166
6.1	ReLU 和参数初始化	166
6.1.1	第一个 ReLU 数值实验	167
6.1.2	第二个 ReLU 数值实验	169
6.1.3	第三个实验——Sigmoid	171
6.2	Xavier 初始化	172
6.3	MSRA 初始化	178
6.3.1	前向推导	178
6.3.2	后向推导	181
6.4	ZCA	182
6.5	与数值溢出的战斗	186
6.5.1	Softmax Layer	186
6.5.2	Sigmoid Cross Entropy Loss	189
6.6	总结	192
6.7	参考文献	192

7	网络结构	193
7.1	关于网络结构, 我们更关心什么	193
7.2	网络结构的演化	195
7.2.1	VGG: 模型哲学	195
7.2.2	GoogLeNet: 丰富模型层的内部结构	196
7.2.3	ResNet: 从乘法模型到加法模型	197
7.2.4	全连接层的没落	198
7.3	Batch Normalization	199
7.3.1	Normalization	199
7.3.2	使用 BN 层的实验	200
7.3.3	BN 的实现	201
7.4	对 Dropout 的思考	204
7.5	从迁移学习的角度观察网络功能	206
7.6	ResNet 的深入分析	210
7.6.1	DSN 解决梯度消失问题	211
7.6.2	ResNet 网络的展开结构	212
7.6.3	FractalNet	214
7.6.4	DenseNet	215
7.7	总结	217
7.8	参考文献	217
8	优化与训练	219
8.1	梯度下降是一门手艺活儿	219
8.1.1	什么是梯度下降法	219
8.1.2	优雅的步伐	220
8.2	路遥知马力: 动量	225
8.3	SGD 的变种算法	232
8.3.1	非凸函数	232
8.3.2	经典算法的弯道表现	233
8.3.3	Adagrad	234

8.3.4	Rmsprop	235
8.3.5	AdaDelta	236
8.3.6	Adam	237
8.3.7	爬坡赛	240
8.3.8	总结	242
8.4	L1 正则的效果	243
8.4.1	MNIST 的 L1 正则实验	244
8.4.2	次梯度下降法	246
8.5	寻找模型的弱点	251
8.5.1	泛化性实验	252
8.5.2	精确性实验	255
8.6	模型优化路径的可视化	255
8.7	模型的过拟合	260
8.7.1	过拟合方案	261
8.7.2	SGD 与过拟合	263
8.7.3	对于深层模型泛化的猜想	264
8.8	总结	265
8.9	参考文献	265
9	应用：图像的语意分割	267
9.1	FCN	268
9.2	CRF 通俗非严谨的入门	272
9.2.1	有向图与无向图模型	272
9.2.2	Log-Linear Model	278
9.2.3	条件随机场	280
9.3	Dense CRF	281
9.3.1	Dense CRF 是如何被演化出来的	281
9.3.2	Dense CRF 的公式形式	284
9.4	Mean Field 对 Dense CRF 模型的化简	285
9.5	Dense CRF 的推断计算公式	288

9.5.1	Variational Inference 推导	289
9.5.2	进一步化简	291
9.6	完整的模型: CRF as RNN	292
9.7	总结	294
9.8	参考文献	294
10	应用: 图像生成	295
10.1	VAE	295
10.1.1	生成式模型	295
10.1.2	Variational Lower bound	296
10.1.3	Reparameterization Trick	298
10.1.4	Encoder 和 Decoder 的计算公式	299
10.1.5	实现	300
10.1.6	MNIST 生成模型可视化	301
10.2	GAN	303
10.2.1	GAN 的概念	303
10.2.2	GAN 的训练分析	305
10.2.3	GAN 实战	309
10.3	Info-GAN	314
10.3.1	互信息	315
10.3.2	InfoGAN 模型	317
10.4	Wasserstein GAN	320
10.4.1	分布的重叠度	321
10.4.2	两种目标函数存在的问题	323
10.4.3	Wasserstein 距离	325
10.4.4	Wasserstein 距离的优势	329
10.4.5	Wasserstein GAN 的实现	331
10.5	总结	333
10.6	参考文献	334

1

机器学习与深度学习的概念

1.1 什么是机器学习

当你拿起本书时，我们已经携手踏上了探索机器学习的征程。机器学习作为人工智能的一个分支，近年来在科研和工业界有了很大的发展，同时也引起了圈内和圈外所有人的关注。现在，机器人的一些技能已经完全超出了旁观群众的认知，这使得全民又一次开始讨论“机器是否能够取代人类”这样的话题。

“机器取代人类”虽然是一个听上去让人不太舒服的话题，但是这个事情本身是令人向往的。最近几年，机器人在眼睛、耳朵、嘴巴和四肢方面的技术都有了很大提升，在某些领域，它的能力已经超过了一些专家的水平；在另一些领域，其能力也逼近专家的水平；在某些领域，其能力已经有了令人惊叹的重大突破。这些提升让人感到欣喜，因为机器人的表现离大家想象中的人工智能又近了一步。不过，从更深层的角度分析近几年机器学习的发展，我们又会发现一个残酷的事实：这几年深度学习的发展并不是完全得益于机器学习理论研究的新突破，主要是因为支撑理论的外部条件有了很大的提升。

让我们先回到机器学习本身，回到那些经过历史积淀的理论本身，看看它们究竟给我们设定了一个怎样的世界观，它们又有哪些启发我们的思维模式。其实，机器学习理论数十年没有本质变化这件事儿，从某种意义上说也是件好事，说明这个思维框架经得住时间的考验，也说明了它被学习的价值。

1.1.1 机器学习的形式

大家都知道，世界上的许多事物都可以从很多角度、用很多方法进行描述：我们可以十分详细地描述一个事物的细节，以图 1-1 所示的一朵花为例，我们可以说出这朵花的形状、颜色，花瓣有几朵，什么时节开花；也可以概括地描述一个事物，把一个抽象的名字冠在这种花上，以后见到这类花就统一叫它们这个名字。



图 1-1 花

同样，我们可以描述一间房子的地理位置、周围环境、房屋使用面积和配套设施等，如图 1-2 所示；也可以用一个价格标记这个房屋的租赁价格。除了这些描述，我们还可以从时间和空间等不同角度对同一个事物做出不同的描述。



图 1-2 房子的外观图

很显然，这些描述都是针对同一个事物，那么描述之间必然存在某种关联关系。人类经过长时间的学习事件，已经可以在自己熟悉的领域轻松地完成这些描述的转换，而对于计算机，想让它在短时间内完成这些转换几乎是不可能的。它要通过人类的操作和指引才能具备这样的能力。在现实中，我们通常会遇到这样的问题：对于一个事物，关于它的最直观的描述可以以一种比较容易的方式获得，我们希望利用描述间的相关性，将这种描述转换成另一种相对抽象且不容易直接获得的描述。更进一步地，我们希望这个分析运算的事情能够由计算机自动完成，因为计算机运算速度快，每秒可以进行大量的数值计算；同时表现稳定，不知疲倦，没有情绪波动，给电就能工作，成本比人低，所以一旦将计算机的计算潜力激发出来，它就可以更好、更快地完成人类可以解决或者人类不易解决的问题。

因为现代的计算机主要基于数字计算实现，所以我们需要把要解决的问题全部以数字的形式表达出来。数值计算是计算机的强项，但是想把所有信息都用数字的形式表达出来还是有点难度的。为此，前辈们想尽了各种办法。比方说，想表示一朵花的大小，可以把这朵花想象成一个圆形，花蕊是这个圆形的圆心，然后测量这个圆的半径，那么这个半径就是一个可以被计算机使用的描述信息，这个信息在机器学习中一般被称作**特征 (Feature)**。

大小这个属性比较好转化为特征，而且这个转化后的特征和我们的直觉吻合。然而，实际中有些事物就不那么好表示了。比方说，我们想把不同的中文词语用特征的方式表示出来，这件事情就会遇到不小的麻烦。由于特征最终需要以数值的方式表达，而数值天生具有可比较性，也具有度量这样的概念，那么词语与词语之间的度量该如何表示呢？如果像花朵大小那样，把所有的汉字映射到正整数数轴上，每一个汉字确实有了自己的数值表示方法，但是词语之间的比较就失去了意义。或者说，采用传统的定义在欧式距离下的距离测量方法变得不再适用。试想一下，如果我们有一个词语叫“如果”，它被映射到 200，接着又有一个词语叫“但是”，它被映射到 300，那么它们在欧式空间的距离：

$$\sqrt{(300 - 200)^2} = 100$$

表示什么意思呢？如果刚好有一个词语被映射到了 100，那么这个词是不是就是“如果”和“但是”的距离？还是说，这个结果有其他含义呢？这些问题一下子变得难以回答，至少和花朵大小这件事儿相比，这个距离难以解释。

那么，中文词语有什么可以表示的方法呢？一种比较简单的表示法被人们称为 **One Hot 编码 (One-Hot encoding)**，它的表示方法如下。

1. 可以假设中文词语的个数是有限的，或者说我们考虑的中文词语集合是全体中文词语的一个子集，这个子集是有限的，这里假设一共有 N 个词，于是我们定义一个维度为 N 的空间，这个空间的每一维只能取两个值：0 和 1。