

大 / 数 / 据 / 丛 / 书

Deep Learning: Methods and Applications

深度学习 方法及应用

[美] 邓力 (Li Deng) 俞栋 (Dong Yu) 著

谢磊 译



机械工业出版社
CHINA MACHINE PRESS

大数据丛书

深度学习：方法及应用

Deep Learning: Methods and Applications

[美] 邓力 (Li Deng) 著
俞栋 (Dong Yu)

谢磊 译



机械工业出版社

本书对深度学习方法以及它在各种信号与信息处理任务中的应用进行了阐述。书中所列举的应用是根据以下三个标准来选取的：①本书作者所具备的专业技能和知识；②由于深度学习技术的成功应用而取得了重大突破的领域，比如语音识别和计算机视觉等；③由于深度学习的应用而有可能发生巨变的应用领域，以及从深度学习的最新研究中获益的应用领域，其中包括自然语言和文本处理、信息检索和多任务深度学习驱动下的多模态信息处理等。

本书适用于对深度学习方法以及它在各种信号信息处理应用方面具有浓厚兴趣的学生与研究人员。

Deep Learning: Methods and Applications / now Publishers Inc. / 1-1 /
by Li Deng, Dong Yu / ISBN: 9781601988140

Original English Language Edition published by now Publishers Inc.

© China Machine Press 2014. Authorized translation of the English edition
© Li Deng and Dong Yu. This edition is published and sold by permission of
Now Publishers, Inc., the owner of all rights to publish and sell the same.

This title is published in China by China Machine Press with license from
now Publishers Inc.. This edition is authorized for sale in China only, excluding
Hong Kong SAR, Macao SAR and Taiwan. Unauthorized export of this edition is
a violation of the Copyright Act. Violation of this Law is subject to Civil and
Criminal Penalties.

本书由 now Publishers Inc. 授权机械工业出版社在中国大陆地区（不包括香港、澳门特别行政区以及台湾地区）出版与发行。未经许可之出口，视为违反著作权法，将受法律之制裁。

北京市版权局著作权合同登记 图字：01-2015-5213 号

图书在版编目（CIP）数据

深度学习：方法及应用/（美）邓力，（美）俞栋著；谢磊译.

—北京：机械工业出版社，2015.12

书名原文：Deep Learning: Methods and Applications

ISBN 978-7-111-52906-4

I. ①深… II. ①邓…②俞…③谢… III. ①人工智能

IV. ①TP18

中国版本图书馆 CIP 数据核字（2016）第 027177 号

机械工业出版社（北京市百万庄大街 22 号 邮政编码 100037）

策划编辑：王 康 责任编辑：王 康 汤 嘉

版式设计：霍永明 责任校对：闫玥红

封面设计：路恩中 责任印制：乔 宇

北京京丰印刷厂印刷

2016 年 3 月第 1 版第 1 次印刷

169mm×239mm·10.75 印张·201 千字

标准书号：ISBN 978-7-111-52906-4

定价：39.80 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务

网络服务

服务咨询热线：010-88361066

机工官网：www.cmpbook.com

读者购书热线：010-68326294

机工官博：weibo.com/cmp1952

010-88379203

金书网：www.golden-book.com

封面防伪标均为盗版

教育服务网：www.cmpedu.com

译者序

深度学习是目前人工智能、机器学习领域异常火热的研究方向，受到了学术界和工业界的高度关注，被《麻省理工学院技术评论》（MIT Technology Review）评为 2013 年十大突破性技术之首。深度学习已经在语音识别、图像识别、自然语言处理等诸多领域取得了突破性进展，对学术界和工业界产生了深远的影响。本书原著的作者——微软研究院的邓力博士和俞栋博士是最早将深度学习技术付诸于语音识别工业级实践的专家，他们和深度学习专家多伦多大学 Geoffrey Hinton 教授合作，最早将深度神经网络应用于大词汇量连续语音识别领域中，使相对识别错误率降低了 20% 以上。

作为多年的导师和好朋友，两位专家将这本书的中文翻译任务交付给我，我倍感荣幸。此次中译本是在忠于原著的基础上进行翻译的，既涉及深度学习的背景和基本概念，又涉及常用的模型与方法，同时包含深度学习在不同领域中的应用。本书共有 12 章，具体内容包括：引言、深度学习的历史、三类深度学习网络、深度自编码器、预训练的深度神经网络、深度堆叠网络及其变种、语音和音频处理中的应用、在语言模型和自然语言处理中的相关应用、信息检索领域中的应用、在目标识别和计算机视觉中的应用、多模态和多任务学习中的典型应用、结论。

本书的翻译除了受到原著作者的指导，也受到了张蓬副教授、陈小敏、吕航、丁闯、孙思宁、何长青、樊博、张弼弘、张彬彬、周祥增的帮助，在此表示感谢。同时感谢机械工业出版社的大力支持与推动。没有他们的帮助，本书的翻译是无法促成的。

由于深度学习技术是一个快速发展的方向，新的模型和应用层出不穷，加之本人学识有限以及中英文语言表达、术语翻译上的差异，书中难免存在错误，还请广大读者指正与原谅。建议读者在学习过程中和英文原著一起阅读，并参考本书引用的参考文献，以便提高学习和理解效果。

谢磊

原书序

“这本书对最前沿的深度学习方法及应用进行了全面的阐述，不仅包括自动语音识别（ASR），还包括计算机视觉、语言建模、文本处理、多模态学习以及信息检索。在深度学习这一领域，这是第一本，也是最有价值的一本书，能使读者对这一领域进行广泛而深入的学习。深度学习对信息处理的很多方面（尤其对语音识别）都具有重大的影响，甚至对整个科技领域的影响也不容忽视。因此，对于有意了解这一领域的学者，这本书是绝对不容错过的。”

——Sadaoki Furui，芝加哥丰田技术研究院院长，日本东京工业大学教授

目 录

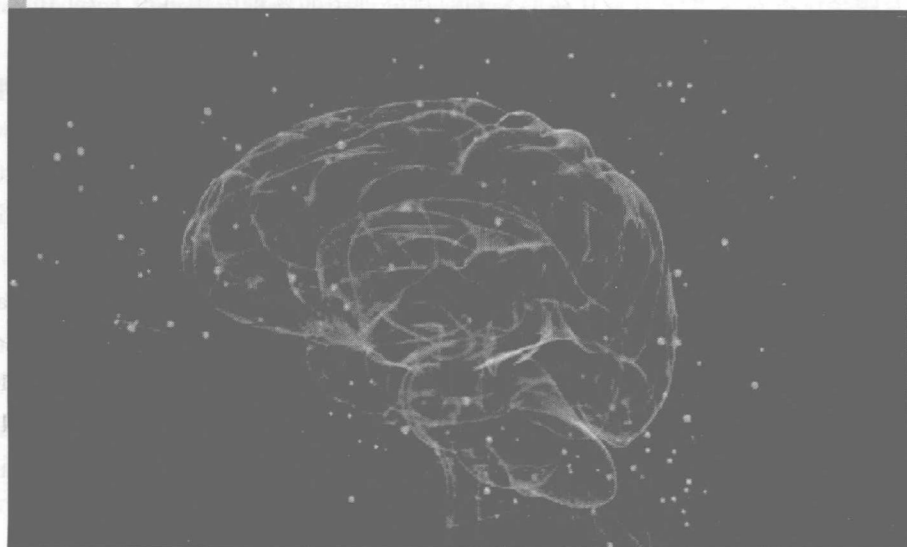
译者序

原书序

1	引言	1
1.1	深度学习的定义与背景	2
1.2	本书的结构安排	5
2	深度学习的历史	7
3	三类深度学习网络	17
3.1	三元分类方式	18
3.2	无监督和生成式学习深度网络	19
3.3	监督学习深度网络	23
3.4	混合深度网络	25
4	深度自编码器——一种无监督学习方法	29
4.1	引言	30
4.2	利用深度自编码器来提取语音特征	30
4.3	堆叠式去噪自编码器	37
4.4	转换自编码器	37
5	预训练的深度神经网络——一种混合方法	39
5.1	受限玻尔兹曼机	40
5.2	无监督逐层预训练	43
5.3	DNN 和 HMM 结合	45
6	深度堆叠网络及其变形——有监督学习	47
6.1	简介	48
6.2	深度堆叠网络的基本结构	49
6.3	一种学习 DSN 权值的方法	50
6.4	张量深度堆叠网络	51

6.5	核化深度堆叠网络	54
7	语音和音频处理中的应用	59
7.1	语音识别中声学模型的建立	60
7.2	语音合成	76
7.3	音频和音乐处理	77
8	在语言模型和自然语言处理中的相关应用	79
8.1	语言模型	80
8.2	自然语言处理	84
9	信息检索领域中的应用	93
9.1	信息检索简介	94
9.2	用基于深度自编码器的语义哈希方法对文档进行索引和检索	95
9.3	文档检索中的深度结构语义模型	95
9.4	信息检索中深度堆叠网络的应用	101
10	在目标识别和计算机视觉中的应用	103
10.1	无监督或生成特征学习	104
10.2	有监督特征学习和分类	106
11	多模态和多任务学习中的典型应用	113
11.1	多模态：文本和图像	114
11.2	多模态：语音和图像	118
11.3	在语音、自然语言处理或者图像领域的多任务学习	120
12	结论	125
附录		129
参考文献		131

引言



1.1 深度学习的定义与背景

2006 年,深度结构学习(经常被称作深度学习或分层学习)作为机器学习研究的一个新的领域出现了^[20,163]。在过去的几年里,由深度学习发展而来的一些科学技术对信号和信息处理的各个方面都产生了深远的影响,这种影响不仅存在于传统领域,也存在于诸如机器学习和人工智能等一些重要的新兴领域中;对于此类研究,文献[7, 20, 24, 77, 94, 161, 412]进行了概述,媒体报道[6, 237]也有所涉及。近年来,很多研讨会、教程、期刊专刊或专题会议都对深度学习及其在信号和信息处理中的各种应用进行了专门的研讨活动,其中包括:

- 2008 年 NIPS (Neural Information Processing Systems, 神经信息处理系统)深度学习研讨会(2008 NIPS Deep Learning Workshop);

- 2009 年 NIPS 关于深度学习的语音识别及相关应用的研讨会(2009 NIPS Workshop on Deep Learning for Speech Recognition and Related Applications);

- 2009 年国际机器学习大会(International Conference on Machine Learning, ICML)关于学习特征的研讨会(2009 ICML Workshop on Learning Feature Hierarchies);

- 2011 年国际机器学习大会关于语音和视觉信息处理中学习架构、表示和最优化的研讨会(2011 ICML Workshop on Learning Architectures, Representations, and Optimization for Speech and Visual Information Processing);

- 2012 年 ICASSP (International Conference on Acoustics, Speech and Signal Processing, 国际声学, 语音与信号处理会议)关于在信号和信息处理中深度学习应用的研讨会(2012 ICASSP Tutorial on Deep Learning for Signal and Information Processing);

- 2012 年国际机器学习大会关于学习表示的研讨会(2012 ICML Workshop on Representation Learning);

- 2012 年 IEEE《音频、语音和语言处理》(T-ASLP, 1 月)会刊中有关语音和语言处理中深度学习专栏(2012 Special Section on Deep Learning for Speech and Language Processing in IEEE Transactions on Audio, Speech, and Language Processing (T-ASLP, January));

- 2010, 2011 和 2012 年 NIPS 关于深度学习和无监督特征学习的研讨会(2010, 2011, and 2012 NIPS Workshops on Deep Learning and Unsupervised Feature Learning);

- 2013 年 NIPS 关于深度学习和输出表示学习的研讨会 (2013 NIPS Workshops on Deep Learning and on Output Representation Learning);
- 2013 年 IEEE《模式分析和机器智能》(T-PAMI, 9 月) 的杂志中有关学习深度架构的特刊 (2013 Special Issue on Learning Deep Architectures in IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI, September));
- 2013 年关于学习表示的国际会议 (2013 International Conference on Learning Representations);
- 2013 年国际机器学习大会关于表示学习面临的挑战研讨会 (2013 ICML Workshop on Representation Learning Challenges);
- 2013 年国际机器学习大会关于音频、语音和语言处理中深度学习的研讨会 (2013 ICML Workshop on Deep Learning for Audio, Speech, and Language Processing);
- 2013 年 ICASSP 关于语音识别中的新型神经网络以及相关应用专栏 (2013 ICASSP Special Session on New Types of Deep Neural Network Learning for Speech Recognition and Related Applications)

本书的作者一直从事深度学习的研究, 也组织或参与过上述中的一些重要会议以及特刊的编写工作。要特别提出的是, 本书作者频频受邀在众多重要会议上对深度学习进行专题报告, 而本书的部分内容也是基于这些报告内容整理而成的。

在开始详细介绍深度学习的内容之前, 我们有必要先了解一些基本概念, 下面是一些与深度学习密切相关的概念和描述:

- **定义 1:** “机器学习是一类利用多个非线性信息处理层来完成监督或者无监督的特征提取和转化, 以及模式分析和分类等任务的技术。”
- **定义 2:** “深度学习是机器学习的子领域, 它是一种通过多层表示来对数据之间的复杂关系进行建模的算法。高层的特征和概念取决于低层的特征和概念, 这样的分层特征叫做深层, 其中大多数模型都基于无监督的学习表示。” (2012 年 3 月维基百科对深度学习的定义。)
- **定义 3:** “深度学习是机器学习的子领域, 它是基于多层表示的学习, 每层对应一个特定的特征、因素或概念。高层概念取决于低层概念, 而且同一低层的概念有助于确定多个高层概念。深度学习是基于表示学习的众多机器学习算法中的一员。一个观测对象 (比如一张图片) 可以用很多种方式表示 (如像素的一个向量), 但是有的表示则可以使基于训练样本的学习任务变得更容易 (如判定某张图像是否为人脸图像)。这一研究领域试图解决一个问题: 哪些因素可以产生更好的表示, 以及对于这些表示应该如何学习。”

(2013年2月维基百科对深度学习的定义。)

● **定义4**：“深度学习是机器学习的一系列算法，它试图在多个层次中进行学习，每层对应于不同级别的抽象。它一般使用人工神经网络，学习到的统计模型中的不同层对应于不同级别的概念。高层概念取决于低层概念，而且同一低层的概念有助于确定多个高层概念。”(2013年10月维基百科对深度学习的最新定义。)

● **定义5**：“深度学习是机器学习研究的一个新领域，它的出现将机器学习向人工智能这一目标进一步拉近。深度学习是对多层表示和抽象的学习，它使一些包括如图像、声音和文本的数据变得有意义。”(参看网址：<https://github.com/lisa-lab/DeepLearningTutorials>)

应该注意的是，本书所讨论的深度学习是使用深度结构来对信号和信息进行处理，而不是对信号或信息的深度理解，尽管在有的情况下这两个方面可能会比较相似。在教育心理学中，是这样定义深度学习的：“深度学习是描述学习的一种方法，其特点是：主动参与、内在激励和个人对意义的探索。”(http://www.blackwellreference.com/public/tocnode?id=g9781405161251_chunk_g97814051612516_ss1-1)我们应该注意将深度学习与教育心理学中的这些被滥用的术语区别开来。

在上述多个不同的高层描述中有两个重要的共同点：(1)都包含多层或多阶非线性信息处理的模型；(2)都使用了连续的更高、更抽象层中的监督或无监督学习特征表示的方法。深度学习是包括神经网络、人工智能、图模型、最优化、模式识别和信息处理的交叉领域，它今天之所以如此受欢迎，有三个重要原因：其一，芯片处理性能的巨大提升(比如，通用图形处理器)；其二，用于训练的数据爆炸性增长；其三，近来，机器学习和信号/信息处理研究有了很大进展，这些都使深度学习方法可以有效利用复杂的非线性函数和非线性的复合函数来学习分布和分层的特征表示，并且可以充分有效地利用标注和非标注的数据。

近年来活跃在机器学习领域的研究机构包括众多高校，比如多伦多大学、纽约大学、加拿大蒙特利尔大学、斯坦福大学、加州大学伯克利分校、加州大学、伦敦大学学院、密歇根大学、麻省理工学院、华盛顿大学，还有一些企业，如微软研究院(从2009年开始)、谷歌(大概从2011年开始)、IBM研究院(大概从2011年开始)、百度(从2012年开始)、Facebook(从2013年开始)、IDIAP研究所、瑞士人工智能研究所等。参看网址：<http://deeplearning.net/deep-learning-research-groups-and-labs/>

这些研究机构将深度学习方法成功地用于计算机领域的众多应用中，其中包括：计算机视觉、语音识别、语音搜索、连续语音识别、语言与图像的特征

编码、语义话语分类、自然语言理解、手写识别、音频处理、信息检索、机器人学，甚至有一个关于分子生物学的研究指出在深度学习方法的引领下发现了新的药物^[237]。

本书在最后一部分列出了一些参考文献，如果需要了解这个领域的最新进展，推荐到以下网址获取：

- <http://deeplearning.net/reading-list/>
- http://ufldl.stanford.edu/wiki/index.php/UFLDL_Recommended_Readings

Readings

- <http://www.cs.toronto.edu/~hinton/>
- <http://deeplearning.net/tutorial/>
- http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial

1.2 本书的结构安排

本书后续章节按照以下结构进行编排：

在第2章中，我们将简要对深度学习的历史加以回顾，主要从以下三个问题入手：第一，深度学习对语音识别技术有哪些影响。第二，这一重大科技革命是如何开始的。第三，它是如何获得并保持如此强大动力的。

第3章讲述了深度学习中绝大多数研究所使用的三元分类法。其中包括：有监督、无监督和混合深度学习网络。在分类任务中，混合深度学习网络利用无监督学习（或称为预训练）来辅助下一个阶段的监督学习。有监督和混合深度神经网络通常都具有同一类型的深度网络体系或结构，但是无监督深度网络的结构却往往不同。

第4~6章分别集中介绍了深度结构的三种主流类型，这三种类型都来源于第3章中所提到的三元分类法。在第4章中，深度自编码器作为无监督深度学习网络的经典方法，我们将详细对其进行介绍并加以讨论。虽然其中巧妙地利用到了反向传播这样的监督学习算法，但是在学习的过程中并没有使用类别标签信息，而是将输入信号本身作为“监督”信号。

第5章作为混合深度网络分类的主要实例讲解部分，详细介绍了这种用无监督生成式的预训练方法来提高监督训练效率的深度学习网络。在训练数据有限，并且没有其他合适的正则化方法（如 dropout）可利用的情况下，混合深度学习网络是很有用的。这种独特的预训练方法是以受限玻尔兹曼机和本章所要学习的深度学习网络为基础的，它开启了深度学习在语音识别和其他信息处理任务中的早期应用，具有很重要的历史意义。除了回顾综述以外，我们也讨论了混合深度学习网络的后续发展和近期出现的一些不同观点。

第6章详细讨论了基于三元分类法的判别式、有监督深度神经网络的一些具体实例——基本的深度堆叠式网络及其扩展。这类深度网络的工作原理与深度神经网络在很多方面都有所不同。需要特别指出的是，它们在建立宏观深度网络的层或模块时采用目标作为标签来简化学习算法。另外，深度网络中的部分假设，如模块中输出单元呈线性的假设，也简化了网络的学习算法，使得我们可以构建和学习比第4章和第5章中网络更丰富的架构。

第7~11章选取了一些深度学习在信号和信息处理各个领域成功的典型应用。第7章回顾了深度学习在语音识别、语音合成和音频处理中的应用，文献综述中围绕语音识别这一主要议题的几个突出问题划分了几个小节。

第8章主要介绍了深度学习在语言模型和自然语言处理中的最新应用成果，其中强调了将符号实体（如词语）转化为低维连续向量的最新进展。

第9章主要集中于对深度学习在信息检索（包含网页搜索）中的突出应用的介绍。

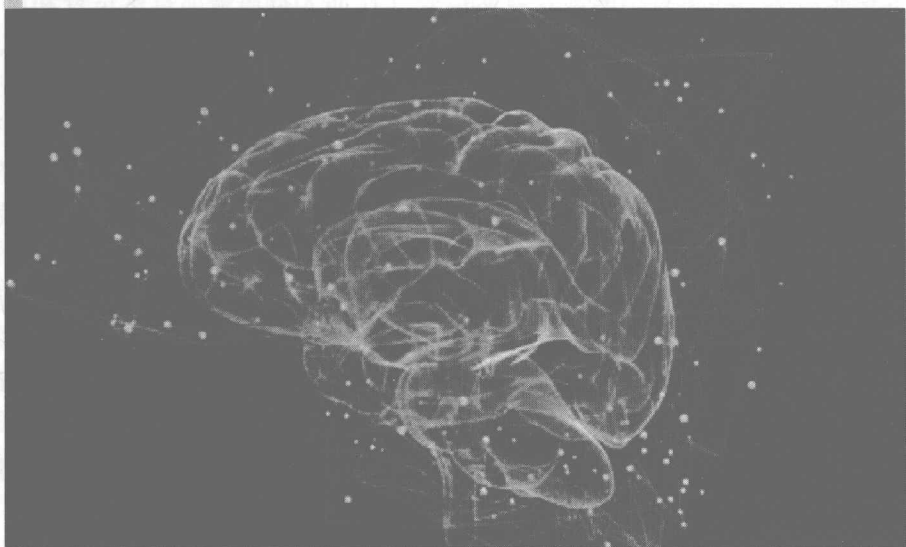
第10章涉及了深度学习在计算机视觉领域中有关图像目标识别的几大应用。这一章将深度学习的方法分为两大类：（1）无监督特征学习；（2）端对端的监督学习以及特征间的学习和分类。

第11章主要介绍了深度学习在多模态处理和多任务学习中的几大应用。我们根据输入到深度学习系统中的多模态数据特征将其分为三类。对于语音、文本或图像的单模态数据，本文也回顾了基于深度学习方法的一些多任务学习研究。

最后，第12章对本书内容进行了总结，并对深度学习将面临的挑战和它的发展方向进行了讨论。

本书内容精短，包括了几位作者提供的专题报告，一次是2011年10月APSIPA会议上的报告，另一次是2012年3月ICASSP会议上的报告。另外，本书也根据领域内的进展，提供了大量更新到2014年1月的内容（包括在2013年12月举办的NIPS-2013和IEEE-ASRU-2013两次会议中的一些资料），这些内容主要集中于近几年快速发展的深度学习研究和技术的应用层面。

深度学习的历史



以前，绝大多数机器学习和信号处理技术都利用浅层结构，这些结构一般包含最多一到两层的非线性特征变换，这种状况直到近几年才得以改变。浅层结构包括高斯混合模型（GMM）、线性或非线性动力系统、条件随机场（CRF）、最大熵模型（MaxEnt）、支持向量机（SVM）、逻辑回归（LR）、核回归以及多层感知器（MLP）（包括极限学习器而且只包含一个隐层）。例如，当使用核方法时，支持向量机就会使用一个只包含一个或零个特征转换层的浅层线性模式的分离模型（最近由深度学习发展而来的一些核方法尤其值得注意，请参见文献[9, 53, 102, 377]）。已经证明，浅层结构在解决很多简单的或者限制较多的问题上效果明显，但是由于其建模和表示能力有限，在遇到实际生活中一些更复杂的涉及自然信号（比如人类语音、自然声音和语言、自然图像和视觉场景）的问题时就会遇到各种困难。

然而，人类信息处理机制（比如视觉和听觉）总是需要深度结构从丰富的感官输入信息中提取复杂结构并构建内部表示。例如，由于人类语言的产出和感知系统都具有清晰的层结构，这就使得信息可以从波形层转换到语言层^[11, 12, 74, 75]。同理，人类视觉系统也有分层的特点，这些虽然基本都只是存在于感知层面，但有趣的是，有时候在产出时也有分层的特点^[43, 126, 287]。我们相信，如果能提出更高效且更有效的深度学习算法，那么用于处理这种自然信号的最前沿技术也将进一步得到提高。

深度学习的概念起源于对人工神经网络的研究（所以有时候可能会听到“新一代神经网络”的说法）。前馈神经网络或具有多隐层的多层感知器——也叫做深度神经网络（Deep Neural Network, DNN）——是深度结构模型中很好的范例。反向传播算法（back-propagation）流行于20世纪80年代，是广为人知的一种学习算法，在学习网络参数上很有用。遗憾的是，仅仅使用反向传播算法在实际学习隐层数目较少的网络时效果并不是很好^[20, 129]。在优化目标为非凸函数的深度神经网络中，来自局部最优化或其他最优化问题的挑战普遍存在，这些挑战通常是学习中面临的主要困难。反向传播算法基于局部梯度信息，并往往从一些随机的初始点开始，当使用批量梯度下降或随机梯度下降的反向传播算法时，目标函数经常会陷入局部最优的境地。随着网络层数的加深，局部最优的情况也就会变得越来越严重。之所以出现上述问题，部分原因在于：我们虽然对小规模的神经网络的探究从未间断过^[42, 45, 87, 168, 212, 263, 304]，但是大多数机器学习和信号处理研究方向有所偏离，人们将重点从对神经网络的研究转移到对具有凸损失函数的浅层模型（例如，支持向量机、CRF和Max-Ent模型）的研究，这类模型以降低建模能力为代价，达到快速高效地收敛到全局最优化的目的，所以深层网络常陷入局部最优的问题还有待解决。

2006年Hinton在研讨会上的两篇论文[163, 164]中介绍了一种高效的

无监督学习算法，它们经验性地缓解了与深度模型相关的最优化难题。这两篇论文介绍了一类叫作深度置信网络（Deep Belief Network, DBN）的深度产生式模型。DBN 是由一组受限玻尔兹曼机（RBMs）堆叠而成的，它的核心部分是贪婪的、逐层学习的算法，这种算法可以最优化深度置信网络的权重，它的时间复杂度与网络的大小和深度呈线性关系。使人意想不到的是，使用配置好的深度置信网络来初始化多层感知器的权重，常常会得到比随机初始化的方法更好的结果。包含多个隐层的多层感知器或深度神经网络，通过无监督的深度置信网络来进行预训练，然后通过反向传播微调来实现，在文献 [67, 260, 258] 中也称之为深度置信网络。最近，研究者对于 DNN 与 DBN 进行了更加细致的区分^[68, 161]，如果使用 DBN 去初始化 DNN 的训练时，这种网络可以被称为 DBN-DNN^[161]。

与受限玻尔兹曼机的发展相独立，在 2006 年，两个不同的、非概率的、非产生式的无监督的深度模型出现了。一个是自编码器的一种变体，使用与 DBN 训练相似的贪心分层进行训练。另一个是基于能量的模型，用稀疏的完备表示来进行非监督学习。与 DBN 相似，它们都可以对深度神经网络进行高效的预训练。

除了具有好的初始点，DBN 还有一些颇具吸引力的优点：第一，它的学习算法可以有效使用未标注的数据；第二，它可以看作是一个概率生成模型；第三，对于经常出现在诸如 DBN 这样的含有数百万个参数的模型中的过拟合问题，以及经常出现在深度网络中的欠拟合问题，都可以通过产生式预训练方法得到有效解决^[259]。

在 DNN 中，多神经元隐层的使用不仅显著提高了 DNN 的建模能力，而且产生出了许多接近的最优配置。即使参数学习过程陷入局部最优，但由于出现欠佳的局部最优的概率比网络中应用少数神经元的时候要低，所以最终的 DNN 仍然效果很好。然而，在训练过程中使用深而宽的神经网络需要强大的计算性能，这也就解释了为什么直到最近几年研究人员才开始认真探索这种既深又宽的神经网络的问题。

更好的学习算法和不同种的非线性关系也促使了 DNN 的成功。随机梯度下降（SGD）算法在大多数训练集较大且数据丰富的情况下是最有效的算法^[39]。最近，研究证实随机梯度下降（SGD）可以有效地实现并行，一种方法是通过异步模式^[69]使用多台机器，另一种方法是使用多 GPU 的流水线型的反向传播算法^[49]。此外，从单个或小批量样本中估计出的嘈杂梯度使得 SGD 通常能让训练跳出局部最优。其他学习算法如 Hessian free^[195, 238]或 Krylov subspace^[378]方法都表现出了类似的能力。

对于 DNN 学习的高度非凸优化问题，由于优化是从初始模型开始的，所

以很明显，更好的参数初始化技术将会打造出更好的模型。然而，不明显的是：如何有效和高效地初始化 DNN 参数以及如何使用大量的训练数据来缓解学习中的问题。对于这些问题，直到最近，文献 [28, 20, 100, 64, 68, 163, 164, 161, 323, 376, 414] 对其进行了探索和分析。此外，之前讨论中提出的无监督的预训练方法是最引人注目的 DNN 参数初始化技术。

DBN 预训练并不是唯一可以使 DNN 有效初始化的过程，另一种性能相当的无监督的方法是：对 DNN 进行逐层地预训练，通过将每两层视为一个除噪自编码器，该除噪自编码器通过将输入节点的随机子集设置为零而进行正则化^[20,376]。另一种方法则是使用压缩自编码器，它通过使输入变量具有更好的鲁棒性来达到同样的目的，例如，对于输入，它通过惩罚隐层单元的激活函数的梯度^[303]来达到目的。此外，Ranzato 等人^[294]开发了稀疏编码对称机 (SESM)，其在构建 DBN 模块中具有和 RBM 非常类似的架构，它也可以用来有效地初始化 DNN 训练。除了使用贪心的逐层过程进行无监督预训练^[28,164,295]，有监督的预训练（有时称为判别式预训练）也证明是很有效的^[28,161,324,432]，并且在有标签的训练数据充足的情况下比无监督的预训练技术表现得更好。判别式预训练的思想是：从一个经过 BP 算法训练的单个隐层 MLP 开始，每一次需要添加一个新的隐层时，用一个随机初始化的新的隐层和输出层替换原有输出层，并用 BP 算法训练全新的 MLP（或 DNN）。与无监督预训练技术不同的是，判别式预训练技术需要标签。

有研究人员将深度学习应用于语音和视觉，他们分析了 DNN 在语音和图像中捕获了哪些信息，比如，文献 [259] 用一种降维方法来可视化研究通过 DNN 学习到的特征向量之间的关系。他们发现，DNN 的隐藏激活向量保留了与多个尺度上的特征向量相似的结构，这一点对于滤波器组特征（filterbank feature）来说尤为如此。最近，Zeiler 和 Fergus^[436]精心设计了另一种可视化方法，该方法基于一个在分类网络的相反方向自上而下的生成过程，用来检测深度卷积网络从图像数据中捕获的特征。深层网络的强大之处在于，它们拥有在提取合适特征的同时做判别的能力^[210]。

可以从另一个角度来了解这个发展历程，即用“成熟度曲线”来回顾人工神经网络的历史。这是一种用图来对一种特定技术的成熟期、接受程度和社会应用进行描述的方式。图 2.1 显示的是由高德纳咨询公司（Gartner）绘制的 2012 版的成熟度曲线图，用来展示一项技术或应用是如何随着时间推移而发展的（按五个阶段：科技诞生的促动期，过高期望的峰值期，泡沫化的底谷期，稳步爬升的光明期和实质生产的高峰期），也提供了一种管理技术布置的方法。