

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。

© 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.

北京市版权局著作权合同登记 图字 01-2015-5213 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：

封面设计： 责任印制：

印刷厂印刷

年 月第 版第 次印刷

169mm × 239mm · 印张 · 千字

— 册

标准书号：ISBN 978-7-111- -

定价： 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

封面无防伪标均为盗版

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——有监督学习权值	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