

NEURAL NETWORK DESIGN

Second Edition

神经网络设计

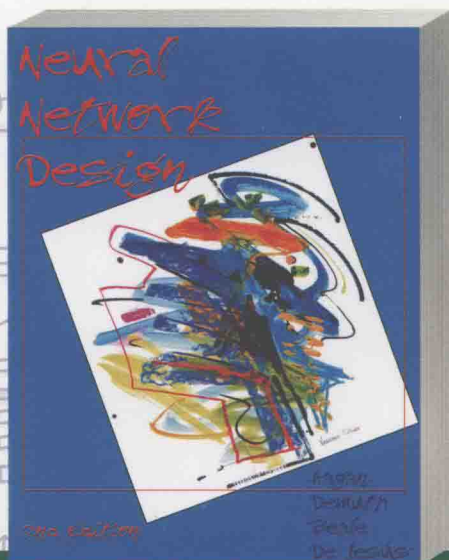
(原书第2版)

马丁 T. 哈根 (Martin T. Hagan)
霍华德 B. 德姆斯 (Howard B. Demuth)
马克 H. 比勒 (Mark H. Beale)
奥兰多·德·赫苏斯 (Orlando De Jesús)

[美]

著

章毅 等译



机械工业出版社
China Machine Press

智能科学与技术丛书



NEURAL NETWORK DESIGN

Second Edition

神经网络设计

(原书第2版)



[美]

马丁 T. 哈根 (Martin T. Hagan)
霍华德 B. 德姆斯 (Howard B. Demuth)
马克 H. 比勒 (Mark H. Beale)
奥兰多·德·赫苏斯 (Orlando De Jesús)

著

章毅 等译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

神经网络设计 (原书第 2 版) / (美) 马丁 T. 哈根 (Martin T. Hagan) 等著; 章毅等译.
—北京: 机械工业出版社, 2017.11

(智能科学与技术丛书)

书名原文: Neural Network Design, Second Edition

ISBN 978-7-111-58674-6

I. 神… II. ①马… ②章… III. 人工神经网络—设计 IV. TP183

中国版本图书馆 CIP 数据核字 (2017) 第 305617 号

本书版权登记号: 图字 01-2017-0150

Authorized translation from the English language edition entitled Neural Network Design, Second Edition (ISBN: 978-0971732117) by Martin T. Hagan, Howard B. Demuth, Mark H. Beale, Orlando De Jesús. Copyright by Martin T. Hagan and Howard B. Demuth.

Chinese simplified language edition published by China Machine Press.

Copyright © 2018 by China Machine Press.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanic, including photocopying, recording, or by any information storage retrieval system, without the prior permission of the publisher.

本书中文简体字版由 Martin T. Hagan 和 Howard B. Demuth 授权机械工业出版社独家出版。未经出版者预先书面许可, 不得以任何方式复制或抄袭本书的任何部分。

本书是一本易学易懂的神经网络教材, 主要讨论网络结构、学习规则、训练技巧和工程应用, 紧紧围绕“设计”这一视角组织材料和展开讲解, 强调基本原理和训练方法, 概念清晰, 数学论述严谨, 包含丰富的实例和练习, 并配有课件和 MATLAB 演示程序。

本书要求读者具备线性代数、概率论和微分方程的基础知识, 可作为高年级本科生或一年级研究生的神经网络导论课程教材, 也可供有兴趣的读者自学或参考。

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: 曲 熠

责任校对: 殷 虹

印 刷: 北京市荣盛彩色印刷有限公司

版 次: 2018 年 1 月第 1 版第 1 次印刷

开 本: 185mm×260mm 1/16

印 张: 27.25

书 号: ISBN 978-7-111-58674-6

定 价: 99.00 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) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@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

多年前，因要给研究生讲授神经网络这门课，我选择了本书第1版作为教材。当时神经网络的书五花八门，要从众多书中选择一本合适的教材并非易事。最终选择这本书的一个重要原因是：本书是难得的数学上比较严谨、完整的神经网络教科书。我们曾用本书为多届研究生讲授神经网络，许多学生反映该书易学易懂。

本书内容涵盖了神经网络领域的许多基础理论。除了严谨的数学论述外，本书另一个重要特点是从神经网络设计的角度阐述方法，使得学生能够快速理解神经网络的本质。书中提供了大量的例题解析、课后练习等。此外，书中还叙述了一些神经网络的相关历史，读来有趣。

感谢机械工业出版社邀请我们翻译本书第2版。和第1版相比，第2版增加了不少新的内容。同时，第1版优秀的中译本也为我们翻译第2版提供了非常有价值的参考。在此，我们要向第1版的译者致以深深的感谢！

参加本书翻译的人员有四川大学章毅、陈媛媛、贺喆南、郭际香、于佳丽、彭玺、陈盈科、张海仙、罗川、吕建成、张蕾、彭德中、桑永胜、陈杰、屈鸿、毛华，以及实验室许多在读博士生如高振涛、戚晓峰等。本书内容较多，翻译时间仓促，难免有不足之处，恳请读者批评指正！

译者

2017年12月于成都

本书介绍基本的神经网络结构和学习规则，重点阐述网络的数学分析、训练方法，以及网络在非线性回归、模式识别、信号处理、数据挖掘和控制系统等领域实际工程问题中的应用。

我们尽最大努力以清晰和一致的方式安排本书内容，以期本书易懂易用。书中使用了许多例子来解释每个讨论的主题。在书的最后几章，我们提供了一些学习实例，以展示神经网络在实际应用中可能遇到的问题。

由于本书是关于神经网络设计的，因此在内容的选择上依据了两个原则。首先，我们希望提供最有用且实际的神经网络结构、学习规则和训练技巧。其次，我们希望本书能够自成体系，并且章节间的过渡自然流畅。为了实现这个目的，我们将有关应用数学的各种介绍材料和章节放在了需要用到这些材料的特定主题之前。简言之，一些内容的选择是因为它们对神经网络的实际应用有重要作用，而另一些内容的选择则是由于它们对解释神经网络的运行机制有重要意义。

我们省略了许多本可以包含的内容。例如，我们没有把本书写成一个涵盖所有已知神经网络结构和学习规则的目录或纲要，而是集中精力介绍基础概念。其次，我们没有讨论神经网络的实现技术，如 VLSI、光学器件和并行计算机等。再有，我们没有提供神经网络的生物学和心理学方面的基础知识。虽然这些内容都是重要的，不过，我们希望重点讨论那些我们认为对神经网络设计最有用的内容，并进行深入阐释。

本书可作为高年级本科生或一年级研究生一个学期的神经网络导论课程教材（也适合短期课程、自学或参考）。读者需要具有一定的线性代数、概率论和微分方程的基础知识。

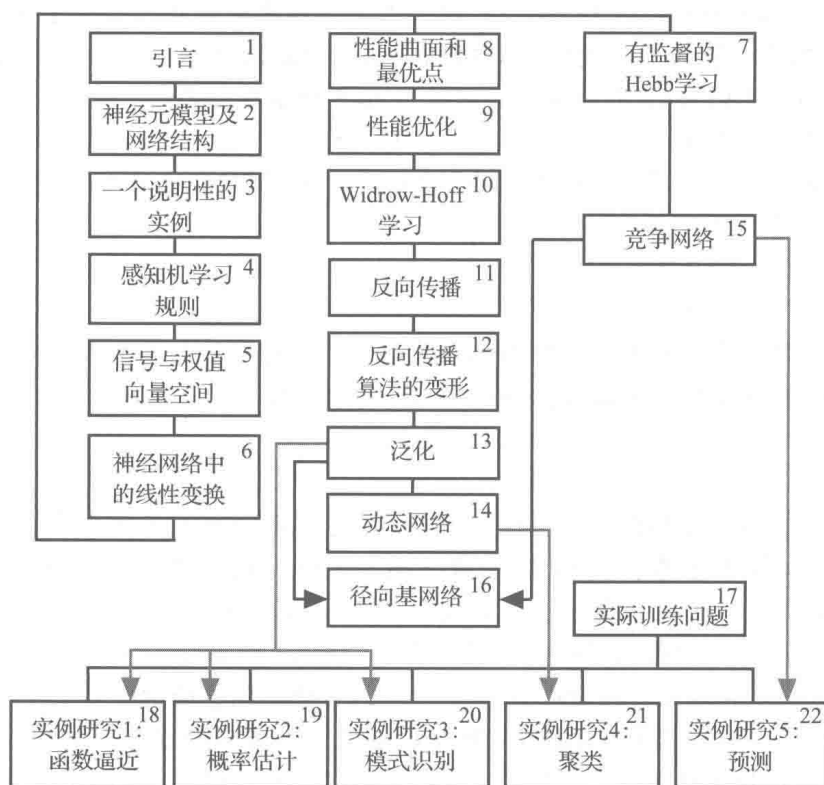
书中每章按如下方式分节：目标、理论与例子、小结、例题、结束语、扩展阅读和习题。理论与例子是每章的主要部分，包含基本思想的发展以及例子解析。小结部分提供一个方便的列表，包含重要方程和概念，以易于工程参考。每章将大约 1/3 的篇幅用在例题部分，为所有关键的概念提供详细的例题解析。

下页的框图说明了各章之间的依赖关系。

第 1~6 章涵盖后面所有章节需要的基本概念。第 1 章是本书的引言，简要介绍神经网络的历史背景和一些生物学基础。第 2 章描述基本的神经网络结构，这一章引入的记号体系将贯穿全书。第 3 章给出一个简单的模式识别问题，并展示该问题可以分别由三类不同的神经网络来解决。这三类网络是本书后面的网络类型的代表。此外，这里的模式识别问题也提供了本书解决问题的一般思路。

本书主要专注于训练神经网络完成各种任务的方法。在第 4 章中，我们介绍学习算法并给出第一个实际算法：感知机学习规则。感知机网络虽然有其本质的局限性，但却具有重要的历史意义，并且可作为一个有用的工具来引入后面章节中强大神经网络的关键概念。

本书的主要目的之一是阐述神经网络的工作原理。为此，我们将神经网络的内容和一些重要的辅助材料组织在一起。例如，在第 5、6 章中提供了线性代数的相关材料，这是理解神经网络所需的核心数学知识。这两章讨论的概念将广泛用于本书后面章节。



第 7 章和第 15 章阐述主要受生物学和心理学启发的神经网络及其学习规则。它们分为两类：联想网络和竞争网络。联想网络安排在第 7 章，而竞争网络安排在第 15 章。

第 8~14 章以及第 16 章论述一类称为性能学习的方法，该方法用于训练网络以优化其性能。第 8、9 章介绍性能学习的基本概念。第 10~13 章中将这些概念用于逐步强大和复杂的前馈神经网络。第 14 章将这些概念用于动态神经网络。在第 16 章中，这些概念则被用于径向基神经网络，这类网络也用到了来自竞争学习的概念。

第 17~22 章和前面的章节不同。前面的章节主要关注每种网络的基础理论和相应的学习规则，要点是理解核心概念。在第 17~22 章中，我们讨论神经网络在实际应用中的具体问题。第 17 章阐述许多实际的训练技巧，第 18~22 章给出一系列实例研究，这些实例包括将神经网络用于函数逼近、概率估计、模式识别、聚类和预测。

软件

MATLAB 软件并非是使用本书所必需的，上机练习可以用任何编程语言来完成。另外，本书提供的神经网络设计演示(Neural Network Design Demonstrations)程序虽然有助于理解书中内容，但非关键所在。

不过，我们还是采用 MATLAB 软件包来提供本书的辅助材料。该软件使用广泛，而且由于它的矩阵/向量表示法和图形化显示，为神经网络实验提供了方便的环境。我们用两种不同的方法使用 MATLAB。第一种，我们为读者准备了一些用 MATLAB 去完成的练习。神经网络的一些重要特性仅在大型问题中才能体现出来，这需要大量计算，不适合手工演算。利用 MATLAB 可以快速实现神经网络算法，也可以方便地测试大型问题。(如果没有 MATLAB 软件，可以使用任何其他编程语言来完成这些练习。)

第二种使用 MATLAB 的方法是利用 Neural Network Design Demonstrations 软件，

你可以从网站 hagan.okstate.edu/nnd.html 下载它。这些交互式的演示程序解释了每章中的重要概念。把这个软件加载到你计算机的 MATLAB 目录下之后(或者放在 MATLAB 路径上),就可在 MATLAB 提示符下通过输入 `nnd` 命令进行调用。所有的演示程序都可以方便地由一个主菜单访问。

这些演示需要 2010a 或更高版本的 MATLAB,也可使用 MATLAB 学生版。附录 C 给出了关于这个演示软件的具体信息。

课件

为了帮助使用本书的教师,我们为本书配备了一套课件。每章的课件(PPT 格式或者 PDF 格式)可以从网站 hagan.okstate.edu/nnd.html 下载。

致谢

我们对审稿人致以深切的感谢!他们花费了大量时间审阅本书全部或部分书稿,并测试了本书软件的各个版本。我们要特别感谢 Canterbury 大学的 John Andreae 教授、AT&T 的 Dan Foresee、Oklahoma 州立大学的 Carl Latino 博士、MCI 的 Jack Hagan、SRI 的 Gerry Andeen 博士以及 Idaho 大学的 Joan Miller 和 Margie Jenks。来自 Oklahoma 州立大学 ECEN 5733、Canterbury 大学 ENEL 621、法国国立应用科学学院 INSA 0506 和 Colorado 大学 ECE 5120 课程班的研究生也提出了许多建设性意见,他们阅读了本书初稿,测试了软件,并在过去几年中提出了许多建议以进一步改进本书。我们还要感谢为本书提供过许多建议的匿名审稿人。

感谢 Peter Gough 博士邀请我们加入位于新西兰 Christchurch 的 Canterbury 大学电气电子工程系,感谢 Andre Titli 博士邀请我们加入位于法国 Toulouse 的法国国家科学研究中心的系统分析与体系结构实验室。Oklahoma 州立大学的学术休假和 Idaho 大学的一年假期给了我们写作本书的时间。感谢 Texas Instruments、Halliburton、Cummins、Amgen 以及 NSF 对我们在神经网络研究上的资助。感谢 The Mathworks 许可我们使用 Neural Network Toolbox 中的材料。

译者序
前言

第 1 章 引言 1

- 1.1 目标 1
- 1.2 历史 1
- 1.3 应用 3
- 1.4 生物学启示 4
- 1.5 扩展阅读 5

第 2 章 神经元模型及网络结构 8

- 2.1 目标 8
- 2.2 理论与例子 8
 - 2.2.1 记号 8
 - 2.2.2 神经元模型 8
 - 2.2.3 网络结构 11
- 2.3 小结 15
- 2.4 例题 17
- 2.5 结束语 18
- 2.6 习题 18

第 3 章 一个说明性的实例 20

- 3.1 目标 20
- 3.2 理论与例子 20
 - 3.2.1 问题描述 20
 - 3.2.2 感知机 21
 - 3.2.3 Hamming 网络 23
 - 3.2.4 Hopfield 网络 26
- 3.3 结束语 27
- 3.4 习题 28

第 4 章 感知机学习规则 31

- 4.1 目标 31

4.2 理论与例子 31

- 4.2.1 学习规则 31
- 4.2.2 感知机结构 32
- 4.2.3 感知机的学习规则 35
- 4.2.4 收敛性证明 39
- 4.3 小结 41
- 4.4 例题 42
- 4.5 结束语 48
- 4.6 扩展阅读 49
- 4.7 习题 49

第 5 章 信号与权值向量空间 53

- 5.1 目标 53
- 5.2 理论与例子 53
 - 5.2.1 线性向量空间 53
 - 5.2.2 线性无关 54
 - 5.2.3 生成空间 55
 - 5.2.4 内积 56
 - 5.2.5 范数 56
 - 5.2.6 正交性 56
 - 5.2.7 向量展开式 58
- 5.3 小结 60
- 5.4 例题 61
- 5.5 结束语 66
- 5.6 扩展阅读 67
- 5.7 习题 67

第 6 章 神经网络中的线性变换 71

- 6.1 目标 71
- 6.2 理论与例子 71
 - 6.2.1 线性变换 71
 - 6.2.2 矩阵表示 72
 - 6.2.3 基变换 74
 - 6.2.4 特征值与特征向量 76

6.3 小结	79	9.2.3 共轭梯度法	139
6.4 例题	79	9.3 小结	142
6.5 结束语	85	9.4 例题	142
6.6 扩展阅读	85	9.5 结束语	150
6.7 习题	86	9.6 扩展阅读	150
第7章 有监督的 Hebb 学习	90	9.7 习题	151
7.1 目标	90	第10章 Widrow-Hoff 学习	153
7.2 理论与例子	90	10.1 目标	153
7.2.1 线性联想器	91	10.2 理论与例子	153
7.2.2 Hebb 规则	91	10.2.1 ADALINE 网络	153
7.2.3 伪逆规则	93	10.2.2 均方误差	154
7.2.4 应用	95	10.2.3 LMS 算法	156
7.2.5 Hebb 学习的变形	96	10.2.4 收敛性分析	157
7.3 小结	97	10.2.5 自适应滤波器	159
7.4 例题	98	10.3 小结	164
7.5 结束语	105	10.4 例题	165
7.6 扩展阅读	105	10.5 结束语	174
7.7 习题	106	10.6 扩展阅读	174
第8章 性能曲面和最优点	108	10.7 习题	175
8.1 目标	108	第11章 反向传播	179
8.2 理论与例子	108	11.1 目标	179
8.2.1 泰勒级数	108	11.2 理论与例子	179
8.2.2 方向导数	110	11.2.1 多层感知机	179
8.2.3 极小点	111	11.2.2 反向传播算法	182
8.2.4 优化的必要条件	113	11.2.3 例子	186
8.2.5 二次函数	114	11.2.4 批量训练和增量训练	188
8.3 小结	119	11.2.5 使用反向传播	188
8.4 例题	120	11.3 小结	192
8.5 结束语	127	11.4 例题	193
8.6 扩展阅读	127	11.5 结束语	201
8.7 习题	128	11.6 扩展阅读	201
第9章 性能优化	131	11.7 习题	202
9.1 目标	131	第12章 反向传播算法的变形	210
9.2 理论与例子	131	12.1 目标	210
9.2.1 最速下降法	131	12.2 理论与例子	210
9.2.2 牛顿法	136	12.2.1 反向传播算法的缺点	210

12.2.2 反向传播算法的启发式改进	215	15.2.5 学习向量量化	310
12.2.3 数值优化技术	218	15.3 小结	314
12.3 小结	226	15.4 例题	315
12.4 例题	228	15.5 结束语	322
12.5 结束语	235	15.6 扩展阅读	322
12.6 扩展阅读	236	15.7 习题	323
12.7 习题	237		
第 13 章 泛化	241	第 16 章 径向基网络	329
13.1 目标	241	16.1 目标	329
13.2 理论与例子	241	16.2 理论与例子	329
13.2.1 问题描述	242	16.2.1 径向基网络	329
13.2.2 提升泛化能力的方法	243	16.2.2 训练 RBF 网络	333
13.3 小结	257	16.3 小结	343
13.4 例题	258	16.4 例题	344
13.5 结束语	265	16.5 结束语	347
13.6 扩展阅读	265	16.6 扩展阅读	347
13.7 习题	266	16.7 习题	348
第 14 章 动态网络	270	第 17 章 实际训练问题	352
14.1 目标	270	17.1 目标	352
14.2 理论与例子	270	17.2 理论与例子	352
14.2.1 分层数字动态网络	271	17.2.1 训练前的步骤	353
14.2.2 动态学习的基本原则	273	17.2.2 网络训练	359
14.2.3 动态反向传播	276	17.2.3 训练结果分析	362
14.3 小结	288	17.3 结束语	368
14.4 例题	290	17.4 扩展阅读	368
14.5 结束语	296		
14.6 扩展阅读	296	第 18 章 实例研究 1: 函数逼近	370
14.7 习题	297	18.1 目标	370
第 15 章 竞争网络	302	18.2 理论与例子	370
15.1 目标	302	18.2.1 智能传感系统描述	370
15.2 理论与例子	302	18.2.2 数据收集与预处理	371
15.2.1 Hamming 网络	303	18.2.3 网络结构选择	372
15.2.2 竞争层	304	18.2.4 网络训练	372
15.2.3 生物学中的竞争层	307	18.2.5 验证	373
15.2.4 自组织特征图	308	18.2.6 数据集	374
		18.3 结束语	375
		18.4 扩展阅读	375

第 19 章 实例研究 2: 概率估计 ...	376	21.2 理论与例子	391
19.1 目标	376	21.2.1 森林覆盖问题描述	391
19.2 理论与例子	376	21.2.2 数据收集与预处理	392
19.2.1 CVD 过程描述	376	21.2.3 网络结构选择	392
19.2.2 数据收集与预处理	377	21.2.4 网络训练	393
19.2.3 网络结构选择	378	21.2.5 验证	394
19.2.4 网络训练	379	21.2.6 数据集	396
19.2.5 验证	381	21.3 结束语	396
19.2.6 数据集	382	21.4 扩展阅读	396
19.3 结束语	382		
19.4 扩展阅读	383	第 22 章 实例研究 5: 预测	398
		22.1 目标	398
第 20 章 实例研究 3: 模式识别 ...	384	22.2 理论与例子	398
20.1 目标	384	22.2.1 磁悬浮系统描述	398
20.2 理论与例子	384	22.2.2 数据收集与预处理	399
20.2.1 心肌梗死识别问题		22.2.3 网络结构选择	399
描述	384	22.2.4 网络训练	401
20.2.2 数据收集与预处理	384	22.2.5 验证	402
20.2.3 网络结构选择	387	22.2.6 数据集	404
20.2.4 网络训练	387	22.3 结束语	404
20.2.5 验证	388	22.4 扩展阅读	405
20.2.6 数据集	389		
20.3 结束语	390	附录 A 参考文献	406
20.4 扩展阅读	390	附录 B 记号	413
		附录 C 软件	417
第 21 章 实例研究 4: 聚类	391	索引	420
21.1 目标	391		

引言

1.1 目标

当你翻开本书时，大脑正在调用复杂的生物神经网络用于阅读。你拥有大约 10^{11} 个高度连接的神经元，它们帮助你完成阅读、呼吸、运动和思考。每一个生物神经元都含有丰富的组织与化学成分，它的复杂程度相当于一个低速微处理器。你的一部分神经结构是与生俱来的，而其他部分则是后期随着人生经历而建立起来的。

科学家对于生物神经网络如何工作的认识尚处于起步阶段。一般认为，所有的生物神经功能，包括记忆，都存储在神经元及神经元之间的连接中。学习，被认为是神经元之间建立新的连接或者修改已有连接的过程。如此，便引出了这样一个问题：尽管目前对生物神经网络的理解尚处在初级阶段，但我们能否构造一些简单的人工“神经元”，或许还可以训练它们，去实现一些有用的功能？这个答案是肯定的。本书正是一本关于人工神经网络的著作。

我们这里谈论的神经元并不是指生物神经元，而是指通过对生物神经元的高度简单的抽象而产生的人工神经元，它们可以由程序或者硅电子电路来实现。由这种人工神经元构成的网络不具备人类大脑那样的能力，但可以通过训练实现一些有用的功能。本书将论述这类神经元，以及由它们构成的网络和相应的网络训练方法。

1-1

1.2 历史

神经网络的历史充满传奇色彩，来自不同领域的充满创造力的先驱们奋斗了数十载，为我们探索出许多今天看来已是理所当然的概念。这段历史曾被多位作者在书中描述过，其中特别有趣的是由 John Anderson 和 Edward Rosenfeld 合著的《Neurocomputing: Foundations of Research》（《神经计算：研究基础》）。该书收集了 43 篇具有特殊历史意义的文章，每篇文章前面都有一段从历史角度对这篇文章的介绍。

本书每章的开篇都将介绍对神经网络做出贡献的主要人物的历史，这里就不再重复。下面我们将简要介绍一下神经网络的主要发展脉络。

对一项技术的发展而言，至少有两个方面是必需的：概念与实现。首先，我们必须要有概念，它提供一种思考问题的方式，其中包含的观点会澄清之前的认识。概念可能是一个简单的想法，也可能是比较具体的且包含数学的描述。为了进一步阐述这点，让我们回顾一下人类认识心脏的历史。在历史上的不同时期，心脏曾被认为是灵魂的归宿或者是热量的源泉。直到 17 世纪，医学家终于开始认识到心脏是一个泵，并设计了实验来研究心脏作为泵的活动。这些实验革新了我们对循环系统的认识。如果不是因为有泵的概念，就无法把握对心脏的理解。

仅有概念和它们的数学描述还不足以促进一项技术走向成熟，还需要有某种方法来实现与之相关的系统。例如，用于对计算机辅助断层(CAT)扫描的图像进行重构所需要的数学早在许多年前就存在了，然而，直到高速计算机和相关高效率的算法出现后，可用的

CAT 系统才得以实现。

神经网络的历史是通过概念创新和系统实现两个方面发展起来的。然而，发展的道路是崎岖的。

神经网络的一些基础性工作始于 19 世纪后期和 20 世纪初期。这些工作主要来自物理学、心理学、神经生理学等交叉领域的科学家，如 Hermann von Helmholtz、Ernst Mach、Ivan Pavlov。这些早期工作致力于学习、视觉、条件反射等方面的一般理论研究，没有包括神经元活动的数学模型。

现代神经网络起源于 Warren McCulloch 和 Walter Pitts[McPi43]在 20 世纪 40 年代的工作。他们展示了由人工神经元组成的网络原则上能完成任何算术或者逻辑运算。这项工作通常被认为是神经网络领域的起源。

在 McCulloch 和 Pitts 之后，Donald Hebb[Hebb49]指出经典的条件反射(如 Pavlov 所发现的)源于个体神经元的性质。他提出了一套关于生物神经元学习机制的学说(参见第 7 章)。

人工神经网络的第一个实际应用出现于 20 世纪 50 年代后期，这源于 Frank Rosenblatt[Rose58]发明了感知机以及相应的学习算法。Rosenblatt 和他的同事构建了感知机网络，并展示了这种网络的模式识别能力。这一早期的成功应用掀起了神经网络研究的巨大热潮。遗憾的是，稍后的工作显示了原始的感知机网络仅能解决一小类问题(更多 Rosenblatt 及感知机学习规则的内容详见第 4 章)。

几乎同时，Bernard Widrow 和 Ted Hoff[WiHo60]提出了一种新的学习算法用于训练自适应线性网络，这种网络在结构和能力上都类似于 Resenblatt 的感知机。Widrow-Hoff 学习算法直到今天还在使用(更多内容见第 10 章 Widrow-Hoff 学习)。

不幸的是，Rosenblatt 和 Widrow 的网络都具有同样的内在局限性。这一局限性通过 Marvin Minsky 和 Seymour Papert[MiPa69]的著作被广泛传播。虽然 Rosenblatt 和 Widrow 当时也意识到了这种局限性，并提出了能够克服这种局限性的新网络，然而，他们没能成功地改进感知机学习算法以训练更为复杂的网络。

受 Minsky 和 Papert 的影响，许多人认为进一步研究神经网络是死路一条，加之当时没有足够计算能力的计算机帮助实验，许多人放弃了神经网络的研究，致使神经网络的研究停顿了约 10 年时间。

虽然对神经网络的兴趣骤然停止，不过，20 世纪 70 年代仍然有一些重要的研究工作持续出现。1972 年，Teuvo Kohonen[Koho72]和 James Anderson[Ande72]独立发明了具有记忆功能的新型神经网络。Stephen Grossberg[Gros76]也在那个年代积极从事自组织神经网络的研究。进入 20 世纪 80 年代，计算设备和研究路线这两个困难均被克服，对神经网络的研究兴趣显著增加。新式个人电脑和工作站的计算能力快速增长并广泛应用。此外，一些重要的新概念也涌现出来。

有两个新的概念对神经网络的重生起着重要作用。第一个是使用统计机制去解释一类回复神经网络的运行，这种网络可以用于联想记忆。这一奠基性的工作是由物理学家 John Hopfield[Hopf82]做出的。

第二个重要的发展是 20 世纪 80 年代的用于训练多层感知机网络的反向传播算法，该算法由多位科学家独立提出。最具影响的关于反向传播算法的著作是由 David Rumelhart 和 James McClelland[RuMc86]合著的。这个算法是对 20 世纪 60 年代 Minsky 和 Papert 对感知机批评的回答(详见第 11 章反向传播算法的发展)。

这些新的发展重新点燃了人们对神经网络的热情。自 20 世纪 80 年代以来,人们发表了数以千计的神经网络论文,开发了无数神经网络应用,整个领域忙碌于新的理论和实践工作。

上面给出的简短历史并非要细数为本领域做出主要贡献的所有科学家,而是想让读者感受一下神经网络领域的知识是如何发展起来的。大家可能已感受到,这个发展过程并不是持续稳定的,它有急剧发展的时期,也有平淡无奇的日子。

神经网络的许多进步都和新的概念密切相关,比如新颖的结构和训练方法。与之同等重要的是新型计算机的出现,强大的计算能力能够测试新的概念。

有关神经网络的历史就讲到这里。一个现实的问题是:“未来会发生什么?”神经网络作为一种数学或者工程工具,已占据牢不可破的地位。它们不会解答所有问题,而是某些适当情形下的重要工具。此外,不要忘记我们对大脑是如何工作的还知之甚少,神经网络最重要的发展毫无疑问将是在未来。

神经网络大量而广泛的应用是非常鼓舞人心的。下节,我们将介绍一些各方面的应用。

1-4

1.3 应用

报刊曾报道过 Aston 大学将神经网络用于文学研究。它这样写道:“网络可以经训练用于个体写作风格识别,研究者们用它来比较莎士比亚及其同龄人的工作。”一档科普电视节目曾报道,在意大利的一个研究所,他们采用神经网络对橄榄油进行纯度检测。Google 将神经网络用于图像标记(自动识别图像并给出关键字),Microsoft 发展出能把英文语音转换为中文语音的神经网络。来自瑞典 Lund 大学和 Skåne 大学医学院的研究人员利用神经网络,通过识别接受者与捐献者的最优匹配,提高心脏移植患者的长期存活率。这些例子展示了神经网络广阔的应用领域。因为神经网络善于解决问题,其应用还在不断扩张,不仅在工程、科学和数学上,而且在医学、商业、金融以及文学等领域。神经网络在许多领域的广泛应用,使得其极具吸引力。另一方面,更高速计算机和更高效算法的出现,使得利用神经网络去解决过去需要大量计算的复杂工业问题成为可能。

下面列出的神经网络应用来自于 MATLAB 神经网络工具箱,已获得 MathWorks 公司许可。

DARPA[DARP88]在 1988 年的一份神经网络研究中列出了神经网络的多项应用,其中第一个是大约出现于 1984 年的自适应频道均衡器。这是一个非常成功的商业产品,它是一个单神经元网络,用于稳定长途电话系统的声音信号。DARPA 的报告还列出了其他商业应用产品,包括单词识别器、过程监控器、声呐分类器和风险分析系统。

在 DARPA 报告发布后的很多年里,数以千计的神经网络被用于数百个不同的领域,部分应用列举如下。

- 航空。高性能飞行器自动驾驶,飞行路线模拟,飞行器控制系统,自动驾驶增强,飞行器部件模拟,飞行器部件故障检查。
- 汽车。汽车自动导航系统,燃油喷射器控制,自动刹车系统,熄火检测,虚拟排放传感器,保修活动分析。
- 银行。支票和其他文档阅读器,信用评估,现金流预测,公司分类,汇率预测,贷款回收率预测,信用风险评估。
- 国防。武器操控,目标追踪,目标辨识,人脸识别,新型传感器,声呐,雷达和图像信号处理(包含数据压缩、特征提取、噪声抑制、信号/图像识别)。

1-5

- 电子。代码序列预测，集成电路芯片布局，过程控制，芯片故障分析，机器视觉，语音合成，非线性建模。
- 娱乐。动画，特效，市场预测。
- 金融。不动产评估，借贷咨询，抵押贷款审查，企业债券分级，信贷额度使用分析，组合交易规划，公司财务分析，货币价格预测。
- 保险。决策评估，产品优化。
- 制造。制造流程控制，产品设计与分析，过程与机器诊断，实时微粒识别，可视化质量监测系统，啤酒检测，焊接质量分析，纸张质量预测，计算机芯片质量分析，磨床运转分析，化工产品设计分析，机器维护分析，项目投标，计划与管理，化工过程系统动态建模分析。
- 医疗。乳腺癌细胞分析，EEG 和 ECG 分析，假肢设计，移植时间优化，医疗费用节约，医疗质量改进，急诊室检测建议。
- 石油与天然气。勘探，智能传感器，油藏建模，油井处理决策，地震解释。
- 机器人。轨迹控制，铲车机器人，操作器控制，视觉系统，无人驾驶汽车。
- 语音。语音识别，语音压缩，元音分类，文本语音合成。
- 证券。市场分析，自动债券评级，股票交易咨询系统。
- 电信。图像和数据压缩，自动信息服务，实时口语翻译，客户支付处理系统。
- 交通。卡车制动诊断系统，车辆调度，路径规划。

1-6

可见，神经网络应用的数量、投入到神经网络软硬件方面的资金，以及人们对这些神经网络器件兴趣的广度和深度，都是巨大的。

1-7

1.4 生物学启示

本书所描述的人工神经网络与对应的生物神经网络仅有很少的联系。本节，我们将简要阐述一些大脑的功能特征，它们启发了人工神经网络的发展。

大脑含有大量(大约 10^{11} 个)高度连接(每个神经元大约有 10^4 个连接)的神经元。对我们来说，这些神经元都由三个主要部分构成：树突、胞体和轴突。树突是一种用于接收电信号并传给胞体的树状神经纤维接收网。胞体有效地叠加这些传入的信息并用阈值控制电信号信息的输出。轴突是一条长的神经纤维，用于将信息从胞体传送到其他神经元。一个神经元的轴突和另一个神经元树突的连接点称为突触。正是因为神经元的结构，以及由复杂化学过程决定的每个突触的连接强度，建立起了神经网络的功能。图 1.1 是一个简化的两个神经元连接的示意图。

当我们生下来时，有些神经结构就存在了。另外一些则是后天通过学习，伴随神经元之间新连接的建立和旧连接的删减而发展出来的。这种发展在每个人的幼儿时期尤其明显。例如，科学家们发现，如果一只小猫在其成长的关键期不使用它的一只眼睛，那这只眼睛就再也发展不出正常的视力。语言学家们发现，婴儿在前六个月中必须要接触某些发音，否则他们之后就再也无法区分这些发音[WeTe84]。

1-8

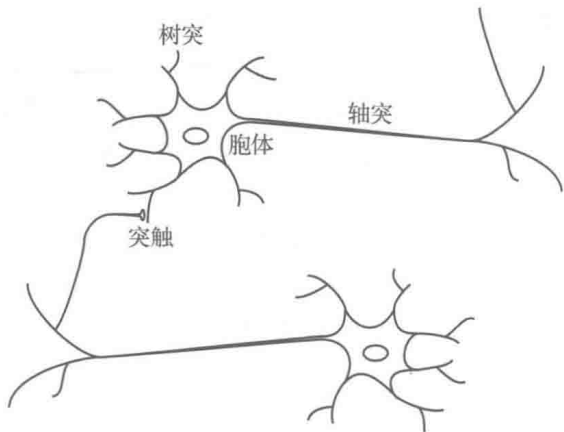


图 1.1 生物神经元示意图

神经结构在一生中会持续改变。这些变化倾向于加强或减弱突触的连接。例如,新的记忆被认为是通过这些突触连接强度的修改而产生的。所以,记住一个新朋友面孔的过程实际是由许多突触连接的改变而来的。又如,神经科学家们发现[MaGa2000],伦敦出租车司机的海马体明显大于平均水平,这是因为他们必须记住大量的导航信息,这个过程需要两年以上的时间。

人工神经网络并非旨在模仿大脑的复杂性,不过,在生物神经网络和人工神经网络之间有两个基本的相似性。首先,两个网络的组成模块都是由简单的计算单元(虽然人工神经网络比起生物神经网络要简单得多)通过高度连接构成的。其次,网络的功能是由神经元之间的连接决定的。本书的基本目的,就是去设计适当的连接进而解决一些实际问题。

需要指出的是,虽然生物神经元相比于电子电路要慢得多(10^{-3} 秒相比于 10^{-10} 秒),但大脑完成许多任务的速度却比任何传统的计算机要快许多。部分原因在于生物神经网络是大规模的并行计算结构,所有的神经元同时进行运算。人工神经网络继承了这种并行结构。虽然目前大部分的人工神经网络是由传统的数字计算机来实现的,然而,这样的并行结构使其非常适合由 VLSI、光学器件、并行处理器来实现。

下面的章节中,我们将介绍基本的人工神经元,以及如何由神经元组合成神经网络。这将为第 3 章提供背景基础,在那里我们将首次看到神经网络粉墨登场。

1-9

1.5 扩展阅读

[Ande72] J. A. Anderson, "A simple neural network generating an interactive memory," *Mathematical Biosciences*, Vol. 14, pp. 197-220, 1972.

Anderson 提出了一个“线性联想器”模型用于联想记忆。这一模型使用推广的 Hebb 假说来训练,用于学习输入向量和输出向量之间的关联关系。它强调网络在生理学层面的合理性。与此同时, Kohonen[Koho72]发表了一篇内容相近的论文,他们的工作是独立进行的。

[AnRo88] J. A. Anderson and E. Rosenfeld, *Neurocomputing: Foundations of Research*, Cambridge, MA: MIT Press, 1989.

该书是一本基础参考书,收录了 40 余篇神经网络计算领域的重要文献。每篇文章都附有一个成果简介,以及该文在领域历史上的地位描述。

[DARP88] DARPA Neural Network Study, Lexington, MA: MIT Lincoln Laboratory, 1988.

这是到 1988 年为止,据作者所知关于神经网络知识的最全面的摘要总结。其中介绍了神经网络的理论基础,并讨论了它们当前的应用。书中包含线性联想器、递归神经网络、视频、语音识别、机器人等内容,最后还讨论了仿真工具和实现技术。

[Gros76] S. Grossberg, "Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors," *Biological Cybernetics*, Vol. 23, pp. 121-134, 1976.

Grossberg 描述了一个基于视觉系统的自组织神经网络。这一网络含有短期记忆和长期记忆两种记忆机制,是一种连续时间竞争网络。它是自适应谐振理论(ART)网络的基础。

1-10

[Gros80] S. Grossberg, "How does the brain build a cognitive code?" *Psychological Review*, Vol. 88, pp. 375-407, 1980.

Grossberg 于 1980 年发表的这篇论文提出了神经结构和机制，它可以解释很多生理行为，如空间频率自适应和双目竞争等。该系统可以在无外界帮助的情况下完成自我校正。

[Hebb49] D. O. Hebb, *The Organization of Behavior*. New York: Wiley, 1949.

这本重要著作的核心思想是：行为可以由生物神经元的活动来解释。在书中，Hebb 提出了第一个神经网络学习规则，这是一个在细胞级别上的学习机制假说。

Hebb 指出生物学中的传统条件反射现象的出现是由个体神经元的性质造成的。

[Hopf82] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, Vol. 79, pp. 2554-2558, 1982.

Hopfield 提出了按内容寻址的神经网络。他清楚地描述了该网络是如何工作的，以及该网络可以胜任何种工作。

[Koho72] T. Kohonen, "Correlation matrix memories," *IEEE Transactions on Computers*, vol. 21, pp. 353-359, 1972.

Kohonen 为联想记忆提出了相关矩阵模型。这种模型用外积规则(也称 Hebb 规则)训练，用于学习输入向量和输出向量之间的相关性，强调网络的数学结构。与此同时，Anderson[Ande72]也发表了一篇内容相近的论文，他们的工作是独立进行的。

[MaGa00] E. A. Maguire, D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. J. Frackowiak, and C. D. Frith, "Navigation-related structural change in the hippocampi of taxi drivers," *Proceedings of the National Academy of Sciences*, Vol. 97, No. 8, pp. 4398-4403, 2000.

伦敦出租车司机必须经历在数千个城市地标中准确导航的广泛训练和学习，完成这种俗称“经历学习”的训练大约需要 2 年时间。这项研究表明出租车司机的后海马体与这些导航控制有重要相关性。

1-11

[McPi43] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115-133, 1943.

这篇论文介绍了神经元的第一个数学模型。这个模型通过比较多个输入信号的加权和与阈值来决定是否激发该神经元。这是基于当时对计算单元的认识，第一次尝试描述大脑的工作原理。论文展示了这种简单的神经网络可以计算任何算术或逻辑函数。

[MiPa69] M. Minsky and S. Papert, *Perceptrons*, Cambridge, MA: MIT Press, 1969.

这本标志性著作第一次严谨地研究了感知机的学习能力。通过严格的论述，阐述了感知机的局限性，以及克服该局限性的方法。但是，该书悲观地认为感知机的局限性意味着神经网络研究是没有前景的。这一失实的观点为后续若干年的神经网络研究和基金资助造成了极大的负面影响。

[Rose58] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, Vol. 65, pp. 386-408, 1958.

Rosenblatt 提出了第一个实际的神经网络模型：感知机。

[RuMc86] D. E. Rumelhart and J. L. McClelland, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, Cambridge, MA: MIT Press, 1986.

20 世纪 80 年代重新燃起人们对神经网络研究兴趣的两大重要论著之一。书中包含了许多主题，训练多层网络的反向传播算法是其中之一。