

神经网络新方法 新理论与方法

张代远 著



清华大学出版社



神经网络 新理论与方法

张代远 著

清华大学出版社
北京

内 容 简 介

这本专著介绍了作者所提出的人工神经网络的新理论与算法(样条权函数神经网络算法与代数算法),这些理论与算法彻底克服了困扰学术界多年的传统算法的困难(例如局部极小、收敛速度慢、不收敛、难以求得全局最优点等困难),其中样条权函数神经网络算法还具有很好的泛化能力。

全书共有7章,讨论了人工神经网络的概论、基本概念、实神经网络的代数算法、全局最小值分析、复数神经网络的代数算法、样条权函数神经网络及其学习算法和神经网络的统计灵敏度分析。

本书逻辑严谨,论述清晰,数学推导详细,对于相关领域的广大科技工作者、理工科博士、硕士研究生以及高年级本科生都有很好的参考价值,也可以作为相关领域的教材。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

神经网络新理论与方法 / 张代远著. —北京:清华大学出版社, 2006.11

ISBN 7-302-13938-5

I. 神… II. 张… III. 人工神经元网络-研究 IV. TP183

中国版本图书馆 CIP 数据核字(2006)第 120377 号

出 版 者: 清华大学出版社 地 址: 北京清华大学学研大厦

<http://www.tup.com.cn> 邮 编: 100084

社 总 机: 010-62770175 客户服务: 010-62776969

组稿编辑: 丁 岭

文稿编辑: 徐跃进

印 装 者: 北京市清华园胶印厂

发 行 者: 新华书店总店北京发行所

开 本: 185 × 260 印张: 8.5 插页: 1 字数: 205 千字

版 次: 2006 年 11 月第 1 版 2006 年 11 月第 1 次印刷

书 号: ISBN 7-302-13938-5/TP · 8379

印 数: 1 ~ 2000

定 价: 28.00 元



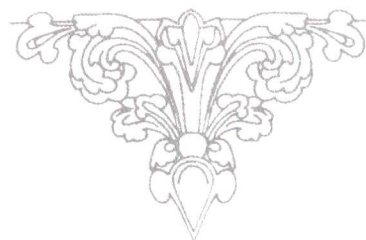
作者简介



张代远 男 博士学位，长期从事神经网络和计算机体系结构方面的科研和教学工作，出版专著、教材及教辅书4部，发表学术论文60余篇，其中不少被国际权威机构检索。作者的主要科研成果如下：

神经网络研究领域 作者提出了样条权函数神经网络算法与代数算法，这些理论与方法彻底克服了困扰学术界多年的传统算法的局部极小、收敛速度慢、难以求得全局最优点等困难。特别是样条权函数神经网络理论与算法具有网络结构简单、泛化能力强等突出优点，值得推广。

计算机体系结构领域 作者提出了迭加原理，给出了迭加原理的一般形式和递推形式。根据迭加原理可以自然地导出状态图的画法规则。引入了流水线渐近参数的概念，导出了计算非线性流水线渐近参数的公式。作者提出的计算非线性流水线参数的公式给出了反映各个参数内在关系的数学表达式。



前言

本书介绍了作者提出的人工神经网络的新理论与算法，主要是实神经网络的代数算法、复神经网络的代数算法、样条权函数神经网络算法（这些算法的名称是作者取的）。

这些理论与算法彻底克服了困扰学术界多年的传统算法的困难（例如 BP 或 RBF 算法中，局部极小、收敛速度慢、不收敛、难以求得全局最优点等困难）。

本书凝聚着作者近几年的创新成果，希望能够起到抛砖引玉，继往开来的作用。

神经网络作为人工智能的一个重要分支，备受关注，它的研究成果对于新一代计算机的研究开发具有重要的意义。作者真诚希望广大读者提出宝贵建议（作者的邮箱地址是 dyzhang@njupt.edu.cn），共同推动这一研究领域的发展，踏出一条创新型的独立发展道路。

全书共分 7 章，具体是：

第 1 章概论，这一章简要介绍计算机、生物神经网络与人工神经网络，本书的目的与价值，神经网络的发展历史等。

第 2 章基本概念，包括生物学的启示、人工神经元模型（单输入单输出人工神经元、多输入单输出人工神经元），简要介绍了作者提出的样条函数神经元以及人工神经网络，人工神经网络结构等基本知识。通过生物神经元的启示，介绍了人工神经元以及人工神经网络的基本概念。这些基本概念是理解人工神经网络的基础。

第 3 章实神经网络的代数算法，介绍新型前馈神经网络学习算法及其理论基础。包括引言、BP 算法的缺点与困难、代数算法的基本原理（符号与神经网络的拓扑结构、代数算法的理论与实现）、计算实例、结论等内容。在这一章里，作者从理论上证明了代数算法可以求得全局最优点。作者在隐层引入了类支集函数，可以使得求解线性方程组的过程稳定。代数算法在神经网络的训练开始之前就能准确给出隐层神经元个数，并具有很高的精度、较快的运算速度。作者在本章中提出了许多新的概念，并采用了一套新的符号。

第 4 章全局最小值分析，包括引言、三层前馈网络代价函数全局最小值分析、数值仿真实验、结论等内容。这一章的重要成果是作者研究了隐层神经元个数的变化与神经网络的映射精度之间的关系，给出了一个重要的精度估计公式。这一重要的精度估计公式在神经网络训练之前就可根据已知的目标样本和隐层神经元个数来确定神经网络代价函数的全局最小值，并指出代价函数全局最小值随隐层神经元个数的增加而单调减小，并且最终成为 0。

第 5 章复数神经网络的代数算法，包括引言、复数神经网络的概念、复数神经网络的构造、复数前馈神经网络的全局最优学习算法的基本原理、隐层神经元个数的确定以及线性方程组的求解（解存在的充分必要条件及隐层神经元个数的确定、隐层神经元函数的选择和复自由权的选择）、结论与数值仿真实验等内容。作者从理论上证明了复数代数算法可以求得复数前馈神经网络的全局最优点。作者提出了隐层采用复数类支集函数训练复数神经网络，保证了求解过程的良态性。另一方面，复数神经网络的代数算法在网络训练开始之前就能准确给出隐层神经元个数，即容易确定获得全局最小点时所需要的隐层神经元的个数。复数神经网络的代数算法具有很高的训练精度和学习速度。在这一章里，作者引入了复数神经网络等许多概念。

第 6 章样条权函数神经网络及其学习算法，第 6 章的内容很重要，希望读者能够将第

6 章的理论与算法应用到实际工作当中去。对于传统的前馈神经网络学习算法(如 BP 算法、RBF 算法等),在训练后神经网络的权值是常数,很难把这些训练后的常数权值与训练过的样本之间建立起内在的对应关系,即人们很难从存储在神经网络的权值中提取有用的规则。对于希望得到规则的学习训练目标来说,这些常数权值似乎没有什么“意义”。为了克服传统神经网络训练的权值难以反映训练样本信息的缺陷,作者提出了一种新型结构的神经网络及其训练算法,作者将其命名为人工神经网络的样条权函数学习(训练)算法,简称为样条权函数学习(训练)算法。作者给出了新颖人工神经网络的拓扑结构,其结构比代数算法以及传统算法(如 BP 算法等)采用的神经网络结构都简单得多。作者提出了第一类样条权函数的神经网络拓扑结构和第二类样条权函数的神经网络拓扑结构。样条权函数学习算法使得训练后的神经网络的权值是输入样本的函数(称为权函数),由三次样条函数构成,而不是传统方法的常数。样条权函数训练算法能够精确记忆训练过的样本,即能够求得全局最小值,而且训练后的权函数可以很好地反映样本的信息特征。不仅如此,随着样本个数的增加,网络的泛化能力也在增强。样条权函数学习算法的主要计算工作量是求解样条函数,它速度快,不存在传统算法(如 BP 等梯度下降类算法)的局部极小、收敛速度慢、初值敏感性等问题。作者详细分析了在 Чебышев 范数意义下,第一类、第二类样条权函数训练算法的误差估计。误差估计对于神经网络的泛化能力的确定以及实际应用都是非常重要的。另外,为了研究样条权函数训练算法,作者在这一章中提出了许多新概念。

这一章的内容包括引言,学习曲线,投影概念与投影方程,第一类权函数的神经网络拓扑结构与训练算法(包括第一类权函数的神经网络拓扑结构、样条权函数方程的建立与求解,第一类权函数的神经网络训练算法又称为和函数的训练算法),第一类权函数的训练算法(和函数的训练算法)的误差分析,第一类权函数的神经网络拓扑结构与训练算法的一般情况(和函数的训练算法),第二类权函数的神经网络拓扑结构与训练算法(又称为积函数的训练算法,还叙述了第二类权函数训练算法的基本原理,第二类权函数的训练算法的误差分析),奇异样本的概念、数值仿真实验、结论等。

第 7 章神经网络的统计灵敏度分析,包括引言、符号约定、灵敏度分析(神经网络误差的一般计算公式、三层神经网络的输出层误差与灵敏度的计算、INI 网络的输出层误差与灵敏度计算)、数值仿真实验、结论等内容。第 7 章研究当训练好的神经网络受到噪声干扰时,神经网络的输出行为将如何改变。采用统计灵敏度概念来分析这样的变化及其影响(能够采用统计灵敏度进行分析的前提是训练后的神经网络的权值必须是精确映射的权值,即这些权值必须是全局最优点)。因此,建议读者采用本书所介绍的神经网络学习算法训练神经网络,因为本书算法得到的权值都是全局最优解。如果采用传统的算法(如 BP 等梯度下降类算法)训练神经网络,则得到的权值通常只能是近似值,此时统计灵敏度已经失去意义。第 7 章讨论了权值和输入样本的加性扰动灵敏度计算的理论公式。当输入扰动较小时,隐层取不同函数时灵敏度的值是不同的,但数值基本上稳定不变。当输入扰动值较大时,网络的灵敏度数值迅速下降,但是网络输出的平均相对误差却迅速增加。

本书逻辑严谨、论述清晰。为使广大本科(理工科)以上学历的读者能够理解本书的主要内容,所有的数学推导都力求尽量详细。

作者特别感谢清华大学出版社能够提供这样一个出版学术专著的机会。

作 者

2006 年 7 月

读者意见反馈

亲爱的读者：

感谢您一直以来对清华版计算机图书的支持和爱护。为了今后为您提供更优秀的图书，请您抽出宝贵的时间来填写下面的意见反馈表，以便我们更好地对本书做进一步改进。同时如果您在使用本书的过程中遇到了什么问题，或者有什么好的建议，也请您来信告诉我们。

地址：北京市海淀区双清路学研大厦 A 座 602 室 计算机与信息分社营销室 收
邮编：100084 电子邮件：jsjic@tup.tsinghua.edu.cn
电话：010-62770175-4608/4409 邮购电话：010-62786544

专著名称：神经网络新理论与方法

ISBN：7-302-13938-5/TP·8379

个人资料

姓名：_____ 年龄：_____ 所在院校/专业：_____

文化程度：_____ 通信地址：_____

联系电话：_____ 电子信箱：_____

您使用本书是作为：☐ 科学研究 ☐ 专题教材 ☐ 其他

您对本书封面设计的满意度：

☐ 很满意 ☐ 满意 ☐ 一般 ☐ 不满意 改进建议_____

您对本书印刷质量的满意度：

☐ 很满意 ☐ 满意 ☐ 一般 ☐ 不满意 改进建议_____

您对本书的总体满意度：

从语言质量角度看 ☐ 很满意 ☐ 满意 ☐ 一般 ☐ 不满意

从科技含量角度看 ☐ 很满意 ☐ 满意 ☐ 一般 ☐ 不满意

本书最令您满意的是：

☐ 创新性 ☐ 指导明确 ☐ 内容充实 ☐ 讲解详尽 ☐ 实例丰富

您认为本书在哪些地方应进行修改？（可附页）

您希望本书在哪些方面进行改进？（可附页）

目 录

第 1 章 概论	1
1.1 计算机、生物神经网络与人工神经网络	1
1.2 本书的目的与价值	2
1.3 神经网络的发展历史简介	4
参考文献	6
第 2 章 基本概念	7
2.1 生物学的启示	7
2.2 人工神经元模型	8
2.2.1 单输入单输出人工神经元	8
2.2.2 多输入单输出人工神经元	10
2.3 人工神经网络结构	11
第 3 章 实神经网络的代数算法	14
3.1 引言	14
3.2 BP 算法的缺点与困难	15
3.3 代数算法的基本原理	22
3.3.1 符号与神经网络的拓扑结构	22
3.3.2 代数算法的理论的实现	26
3.4 计算实例	31
3.5 结论	32
参考文献	33
第 4 章 全局最小值分析	34
4.1 引言	34
4.2 三层前馈网络代价函数全局最小值分析	34
4.3 数值仿真实验	38
4.4 结论	40
参考文献	40
第 5 章 复数神经网络的代数算法	41
5.1 引言	41
5.2 复数神经网络的概念	41
5.3 复数神经网络的构造	42
5.4 复数前馈神经网络的全局最优学习算法的基本原理	42
5.5 隐层神经元个数的确定以及线性方程组的求解	44
5.5.1 解存在的充分必要条件及隐层神经元个数的确定	44
5.5.2 隐层神经元函数的选择	45

5.5.3 复自由权的选择	45
5.6 本章结论与数值仿真实验	49
参考文献	50
第6章 样条权函数神经网络及其学习算法	51
6.1 引言	52
6.2 学习曲线	52
6.3 投影概念与投影方程	55
6.4 第一类权函数的神经网络拓扑结构与训练算法	59
6.4.1 第一类权函数的神经网络拓扑结构	59
6.4.2 样条权函数方程的建立与求解	62
6.5 第一类权函数的训练算法的误差分析	66
6.6 第一类权函数的神经网络拓扑结构与训练算法的一般情况	88
6.7 第二类权函数的神经网络拓扑结构与训练算法	94
6.7.1 第二类权函数的训练算法(积函数的训练算法)的基本原理	94
6.7.2 第二类权函数的训练算法(积函数的训练算法)的误差分析	104
6.8 奇异样本的概念	106
6.9 数值仿真实验	110
6.10 结论	112
参考文献	113
第7章 神经网络的统计灵敏度分析	114
7.1 引言	114
7.2 符号约定	115
7.3 灵敏度分析	116
7.3.1 神经网络误差的一般计算公式	116
7.3.2 三层神经网络的输出层误差与灵敏度的计算	118
7.3.3 INI 网络的输出层误差与灵敏度计算	121
7.4 数值仿真实验	122
7.5 结论	124
参考文献	125

第1章 概 论

1.1 计算机、生物神经网络与人工神经网络

神经网络作为人工智能的一个分支,在近二十年来,受到人们的广泛关注。工业革命以来,人类大量采用机器来减轻人们的体力劳动,并获得了巨大的效益。同样,人类为了通过使用某种机器来减轻人类的脑力劳动,也一直进行着不懈的努力。到了20世纪40年代,由于计算机的发明,使得人类的文明进入到计算机时代。通过使用计算机,人们可以解决科学计算和工程设计中的一些复杂的问题,在一定程度上减轻了人们的脑力劳动。然而,计算机需要在人们事先编制好的程序的指挥下才能工作,从这个意义上讲,计算机并没有真正意义上的智能。目前,计算机的主要应用仍然是信息处理和科学计算,对于智能计算,像分析、推理、判断、综合等方面,现代计算机仍然显得能力低下。

因此,智能计算机的开发研究成了一个十分引人关注的问题,科学家们也为此投入了巨大的研究热情,一些发达国家也投入大量的人力物力来开发智能计算机。然而,由于人类对于智能的理解还十分肤浅,开发智能计算机也遇到了巨大的挑战,可以说,到目前为止,智能计算机还仅仅是一个梦想。揭开智能之谜还需要进行大量的研究,这里面涉及到诸多方面的理论知识,包括计算机理论、信息处理、语言学、认知科学、数学、生理学、解剖学、哲学等方面的知识。因此,智能计算机的发展还需有一个相当漫长的过程。

在应用方面,对于智能计算和智能计算机的发展也是非常迫切的。例如,一些危险行业以及恶劣的工作环境,人们需要机器人来帮助工作,然而,机器人技术的发展离不开智能计算和智能计算机。可以说,智能计算机在当今人类社会发展的进程中,已经被提到了议事日程,是当今人类社会所面临的一项迫切而又重大的科技问题。

自从20世纪40年代人类发明计算机以来,可以说其发展速度一日千里。计算机改变了人们的生活方式,带来了信息技术的革命。传统计算机的存储能力、计算速度的发展非常快,主要得益于硬件的发展。相对而言,计算机体系结构的发展比较缓慢,基本上还是冯·诺依曼体系。它只能在人的指挥下工作,没有学习、创造等反映智能特征的能力。对于许多模式识别的问题,现代计算机的工作能力和效率还远不如人。笔者认为,智能计算及其发展需要突破传统的冯·诺依曼体系,建立新型的计算机体系结构,当然,任重道远,需要科学家们的不懈努力。

智能计算的核心问题是关于人脑功能的模拟问题,这需要回答什么是智能、什么是计算、什么是智能计算等复杂的问题。然而,遗憾的是,人类对于这些问题尚缺乏深刻的认识,还处在探索阶段。尽管如此,人类并没有在研究智能计算机的道路上止步不前,目前认为,人类的大脑中的神经元对于人脑的智能起着关键的作用,这些神经元的数量非常多,组成了十分复杂的生物神经网络。神经网络系统理论是以人脑的智能功能为研究对象,研究人类大脑的信息处理能力与方法,特别是研究与人类大脑的智能信息处理能力相关的信

息处理理论与方法,建立智能计算的理论与方法,为智能计算机的研究开发奠定理论基础。

由于生物神经网络的复杂性,目前人们主要是通过人工神经网络的方法进行研究,所以本书中考虑的神经元不是生物神经元,而是模拟生物神经元的人工神经元,人工神经元是对生物神经元极其简单的抽象,可以使用现代电子、光学等技术制造出来。所谓人工神经网络是由人工神经元按照一定的拓扑结构互连而成的网络,用来模拟人的大脑的一些行为。虽然由这些人工神经元组成的网络的能力远远不及人脑的那么强大,但是可对其进行训练,可以实现一些有用的功能。目前主要是利用人工神经网络模拟一些比较低级的行为,至于如何模拟人的智能中比较高级的功能,还需要进行大量的深入研究。

人类是智慧生命,生命的每一天都要处理许多复杂的事物,要完成生理的、心理的等复杂的信息处理,比如呼吸、运动、阅读、思考等,而完成这些功能需要使用一个复杂的生物神经网络。人的大脑有约为 10^{11} 个神经元,每一个生物神经元都是生物组织和化学物质的有机结合。某些神经结构是与生俱来的,而其他一些则是在实践中形成的。这些神经元高度互连的集合,构成了生物神经网络,网络中的每个神经元都是基本的处理单元,一种观点认为,每一个生物神经元都是一个复杂的微处理器。

人类大脑的工作机理十分复杂,目前,人类对大脑的认识还很肤浅,揭开大脑思维之谜任重道远,需要坚持不懈的努力。一般认为,所有生物神经功能,包括记忆在内,都存储在神经元以及各个神经元之间的相互连接上,这种连接称为连接权或简称为权。生物神经网络的学习过程被看作是在生物神经网络的神经元之间建立新的连接或对已有的连接进行修改的过程。

正是基于对生物神经网络的这样一个基本认识,使得人们想到利用一些简单的“人工神经元”构造一些人工系统,然后对其进行训练,目的是使这些系统具有一定的智能。

本书主要是对于大脑的学习能力进行了一些人工模拟,本书中的所有算法都是作者提出的,这些算法不同于传统的算法,希望能够起到抛砖引玉的作用。

1.2 本书的目的与价值

创新是本书的目的。本书中提出了一系列全新的前馈神经网络理论与算法,这些新的理论与算法主要分成两类:一类作者将其取名为代数算法,另一类作者将其取名为样条函数神经网络算法。

衡量一个新理论与方法的价值在于把它和已有的传统方法相比较,如果新理论与方法能革除传统理论方法的缺点,那么新理论方法就 valuable,革除得越彻底,价值就越大。衡量本书的价值应与传统的优化类算法进行比较(例如 BP 算法等梯度下降类最优化算法)相比较。

1. 代数算法

代数算法不再采用以梯度下降为核心思想的误差反向传播学习算法,只要隐层神经元个数与样本个数满足一个简单的代数关系,代数算法便能快速实现给定样本的精确映射(或者说,准确地“回想”起所学过的样本)。

(1) 作者提出的代数算法在理论和应用上能使代价函数为 0。对 BP (反向传播) 算法而言, 通常代价函数大于 0, 这意味着代数算法的精度远高于 BP 算法。或者更确切地说, 代数算法能准确地获得全局最优解, 而 BP 算法通常无法获得全局最优解。

(2) 从时间复杂度上来看, 代数算法无须迭代计算, 是一种多项式阶算法, 而 BP 算法的时间复杂度目前尚未见到理论上的报道, 但通过一些实验研究认为其时间复杂度为指数阶 (见第 3 章)。多项式阶算法的时间效率远优于指数阶算法, 因而本书算法较 BP 算法要快得多, 可求解问题的规模也要大得多。

(3) 从工程应用上看, 代数算法给出了隐层所需神经元个数的准确计算方法, 而 BP 算法只能给出一些经验数据, 这意味着本书新方法在工程应用中具有比 BP 算法优越得多的指导作用。

综上所述, 可以认为代数算法开辟了多层前馈人工神经网络算法的一个新领域, 在理论和应用上都具有重要的价值。

另外, 本书的另一个重要创新是对工程上常用的三层神经网络的极限逼近能力进行了深入的分析, 得出了一系列重要的结果。这些结果指出, 必然存在代价函数最小值为 0 的三层神经网络 (此时对应的解为全局最优解); 如果三层神经网络的隐层神经元个数固定 (即为某一确定的不可调整的常数), 则对一些给定的样本, 该三层神经网络的代价函数最小值将大于 0, 这意味着此时三层神经网络的逼近精度是有限的, 它不可能使神经网络以任意精度趋近于 0。这一结论深刻地揭示了隐层神经元个数的选择将直接影响到三层神经网络的极限逼近能力, 或者说, 将直接影响到三层神经网络的全局最小代价函数值的大小。不仅如此, 还给出了一个非常有实用价值的、便于计算的估计公式。利用这一估计公式, 可使网络在训练之前就知道该网络对给定训练样本的极限逼近能力, 这一结果对前馈神经网络学习算法的停机准则有重要的理论指导意义。

2. 样条函数神经网络算法

虽然代数算法克服了传统算法的主要缺点 (例如速度慢, 难以求得全局最优值, 无法确定隐层神经元个数等), 但由于代数算法实现精确映射的一个充分条件是隐层神经元的个数等于训练样本的个数 (见第 3 章), 这使得当训练的样本数量很多时会使得神经网络隐层神经元的个数太多, 使得神经网络的结构复杂。另外, 代数算法与传统算法的共同缺点是训练后的权值是常数, 难以反映样本的内在信息。为了克服代数算法的缺点, 本书作者提出了样条函数神经网络算法。样条函数神经网络算法首先改造了传统神经网络的结构, 使得神经网络的结构与训练样本个数无关。另外, 样条函数神经网络算法将常数权改成了权函数, 即看成输入样本的函数, 采用三次样条函数来实现。这样做将权与样本之间建立起了联系, 可以反映训练样本的信息。训练后的神经网络, 其权函数可以很好地反映样本的特征信息。

样条函数神经网络算法同样可以实现代价函数为 0 的精确映射, 能够方便地求得全局最优解。不仅如此, 样条函数神经网络算法还具有很好的泛化能力 (见第 6 章)。

另外, 从时间复杂度上来看, 样条函数神经网络算法的主要计算工作是求线性方程组, 确定三次样条函数, 是一种多项式阶算法, 因而样条函数神经网络算法较 BP 算法要快得多。

多,可求解问题的规模也要大得多。

总之,样条函数神经网络算法既具有代数算法的优点,又克服了代数算法和传统算法的缺点,是一种非常好的算法,笔者认为值得推广。这些创新构成了本书的核心与精华。

1.3 神经网络的发展历史简介

人工神经网络的发展起源于何时,说法不一。一般认为,其起源可追溯到 Warren McCulloch 和 Walter Pitts 于 1943 年的论文^[1],这篇文献根据已知的神经细胞生物过程原理,构造了一个简单的人工神经元模型,后来,人们称它为 M-P 模型。文献[1]介绍了第一个神经元数学模型,其神经元的输入信号加权与与阈值比较再决定神经元是否输出。这是人类历史上第一次对大脑工作原理描述的尝试,从原理上证明了人工神经网络可以计算任何算术和逻辑函数。

20 世纪 40 年代末期,Donala O.Hebb 在文献[2]中首先定义了一种调整权的方法,称为 Hebb 学习规则(Hebbian rule)。这本著作的主要假定是行为可以由神经元的行为来解释。在这本书中 Hebb 提出的第一个学习规则,可以认为是在细胞一级进行学习的基本原理。Donala Hebb 指出,经典的条件反射是由单个神经元的性质引起的。他提出了生物神经元的一种学习机制。

20 世纪 50 年代后期, Frank Rosenblatt 定义了一种以后常用的神经网络结构,称为感知器(perceptron)。Frank Rosenblatt^[3]提出了感知机网络和联想学习规则。Rosenblatt 构造了一个感知机网络,演示了进行模式识别的能力。这是人工神经网络第一个实际应用,这次成功应用引起了许多学者对神经网络的研究兴趣。但后来的研究表明基本的感知机网络只能解决有限的几类问题。

20 世纪 60 年代,两个训练自适应神经元的方法——感知机规则及 LMS 算法规则相继出现。Bernard Widrow 和 Ted Hoff 在文献[4]中介绍了能进行快速、准确学习的神经网络,文中假设系统的输入和对每一个输入所期望的输出,然后神经网络计算实际输出与所希望的输出的误差,采用梯度下降法调整权值,以达到最小均方误差(最小均方误差算法或 LSM 算法)。另外, Bernard Widrow 和 Ted Hoff 提出了一个新的学习算法用于训练自适应线性神经网络。

在这些方法出现后的若干年内,神经网络领域的一些新技术得到迅速发展,如 Steinbuch 的学习矩阵, Widrow 的 Madaline 规则,竞争学习,等等。

20 世纪 70 年代, Grossberg 提出了自适应共振理论。Grossberg 研究了两种记忆机制(短期记忆和长期记忆),提出了一种基于可视系统的自组织神经网络^[5],这是一种连续时间竞争网络,是构成自适应谐振理论(ART)网络的基础。

另外, Kohonen 创立了自组织理论。

20 世纪 80 年代, Hopfield 及一些学者提出了 Hopfield 网络模型^[6],这是一种全连接的反馈型网络。Hopfield 提出了按内容寻址的神经网络,同时还给出了有关他的网络是如何工作的以及该网络能够做什么的完整描述。

另外, Hinton 等提出了 Boltzman 机。Kosko 发展了 Hopfield 及 Grossberg 的一些概念, 提出了双向自联想记忆网络, 该网络中使用了微分 Hebb 及竞争学习规则。另外, 20 世纪 80 年代末期, Chua 等人提出了细胞神经网络。

多层前馈网络是目前应用最为广泛的神经网络模型之一, 其神经元结点分层排列, 组成输入层、隐层和输出层, 每层的神经元只接收前层神经元的输出信号。多层前馈神经网络的反向误差传播训练算法(可简称为 BP 算法)首先是由 Werbos 在他的博士毕业论文中提出^[7], 但不幸的是, 他首先发表于 1974 年的这一成果并没有引起当时科技界的重视。直到 1982 年, Parker 重新发现了这一方法, 不久以后 Rumelhart、Hinton 和 Williams 也重新报道了这一方法^[8], 才使得该算法广为流传。

Rumelhart 等人使用的神经元不同于以往人们在 Madaline 网络中使用的神经元。早期的 Madaline 网络中使用的神经元是硬限幅(Signums)神经元, 而 BP 算法中网络所使用的神经元是可微分的非线性神经元或 Sigmoid 函数神经元。Rumelhart 等人提出的 BP 算法(或如他们自己称作的广义 S 规则)克服了 Minsky 和 Papert (1969)所指出的感知机算法的局限性^[9]。BP 算法似乎再次打开了 Minsky 和 Papert 早已关闭的研究大门。

BP 算法很快在神经网络中占据了主导地位。一些学者转为企业家开发、销售廉价的 BP 算法软件, 而另一些人却平静地用 BP 算法预测股票价格, 成百上千的科技工作者在科技学会上报道 BP 算法的应用成果。一些企业投资者与大公司资助个人(PC)机及工作站上的 BP 算法加速板的开发。大众媒体也把 BP 算法喻为人们期待已久的机器智能研究的突破, 它能从经验中学习, 并预示着制造业的自动化; 如不同语言间的翻译, 能自我修复的机器人, 等等。生理学家们在寻找人脑中中枢神经系统中的 BP 算法, 而理论家们猜想 BP 算法能学会任意样本函数。

BP 算法的普及也引起了人们激烈的批评, BP 算法的主要缺点(这些缺点是本质的, 也是致命的)是常不收敛, 或收敛于局部极小, 即使收敛其速度也很慢, 这使得 BP 算法只能解决小规模的问题。由于 BP 算法的这些缺点, 人们对它进行了改进研究。以往, 人们的改进研究大都集中在变学习率, 加惯性项, 寻求良好的初始条件等方面。应该承认, 这些改进研究有一定的效果, 但它们都没有摆脱 BP 算法思想框架的束缚——梯度下降思想, 正因为如此, 它们不可能从根本上克服 BP 算法的本质缺点。

20 世纪 90 年代以来, 神经网络在向纵深发展的同时, 也在与模糊技术、遗传算法、进化计算等智能方法相结合的方向上发展。

虽然在几十年的发展过程中, 神经网络理论取得了一些进步, 但其发展过程并非一帆风顺, 曾经历了较长一段时间的低潮期, 直到 20 世纪 80 年代后期才进入发展的高潮期。但是, 正当人们满怀热情地应用神经网络的已有成果去解自己的实际工作中的问题时, 人们发现了大量原有模型和算法所存在的问题, 因此, 必须从理论上再做进一步的研究。本书正是在这样的背景下对多层前馈网络算法进行了进一步深入的研究, 取得了重要的成果。

本书作者首次提出了一系列全新的理论和算法, 并将其称为代数算法(详见第 3 章)和样条函数神经网络算法(详见第 6 章), 这些理论和算法摆脱了梯度下降思想的束缚, 使复杂的非线性映射问题转变为线性代数方程组的求解问题, 因而从根本上克服了 BP 算法的本质缺点。具体内容请读者见书中的各章。

参考文献

- [1] McCulloch W and Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, V01.5, p115~133, 1943
- [2] Hebb O. The Organization Behavior. New York: Wiley, 1949
- [3] Rosenblatt P. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, V01.65, p388~408, 1958
- [4] B Widrow and Hoff M E. Adaptive switching circuits. 1960 IRE WESCON Convention Record. New York: IRE Part 4, p96~104, 1960
- [5] Grossberg S. Adaptive pattern classification and universal recoding: I.Parallel development and coding of neural feature detectors. Biological Cybernetics, Vo1.23, p121~134, 1976
- [6] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, V01. 79, p2554~2558, 1982
- [7] Werbos P J. Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph D Thesis, Harvard University, Cambridge, MA, 1974
- [8] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. Nature(London), 323, 533~536, 1986
- [9] Minsky M L, Papert S. Perceptrons: An introduction to computational geometry. M I T Press, Cambridge, MA, 1969

第2章 基本概念

本章通过生物神经元的引入，介绍了人工神经元以及人工神经网络的基本概念。这些基本概念是理解人工神经网络的基础，对于已经具备这些基础知识的读者可以略过这一章。本章也简要介绍了作者提出的样条函数神经元以及神经网络，关于样条函数理论预算法的详细内容请读者阅读第6章。

2.1 生物学的启示

解剖学的研究告诉人们，人的大脑由大量（约 10^{11} 个）高度互连的单元（每个单元约有 10^4 个连接）组成。这些单元被称为神经元。而这些神经元组成了十分复杂的生物神经网络，目前人类对于生物神经网络的整体认识还很肤浅。从解剖学上看，生物神经元由树突、细胞体和轴突三部分组成。

树突是树状的神经纤维接收网络，它将电信号传送到细胞体，细胞体对这些输入信号进行处理。

轴突是单根长纤维，它把细胞体的输出信号传给其他神经元。

突触是一个神经细胞的轴突和另一个神经细胞树突的结合点。

图 2.1 给出了两个生物神经元的简化示意图。

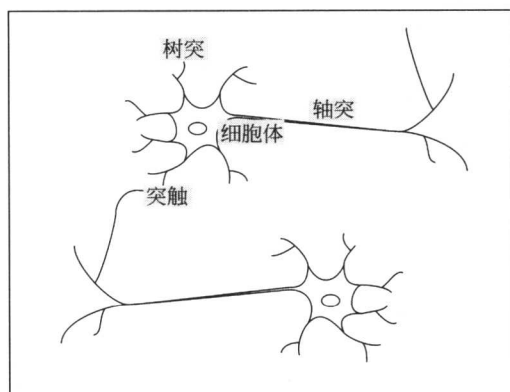


图 2.1 两个生物神经元的简化示意图

一种观点认为，生物体的功能与生物个体的生物神经网络结构有关。一些神经结构是与生俱来的，而另外一些结构则是在学习的过程中形成的。

与生俱来的生物神经结构使得生物刚刚出生时就具备某种功能，也许就是人们常说的本能，而这些都是与生俱来的生物神经结构是通过生物遗传获得的。

在后天的学习过程中会改变生物神经结构。例如，可能会产生一些新的连接，而原有的一些连接也可能会消失。改变生物神经结构的过程在生命早期最为显著。这说明婴幼儿的早期教育非常重要，早期教育使得婴幼儿的生物神经结构的改变更加容易。

生物神经结构在整个生命期内不断地进行着改变。生物神经结构后期的改变与前期的改变有所不同，主要是加强或减弱突触连接。目前的研究已经确认，新记忆的形成是通过改变突触连接强度而实现的。

本书研究的是人工神经网络，它是生物神经网络的高度简化的版本，因此只能在较为低级的层次上模拟人的大脑的功能。由于人类目前对于生物神经网络的了解甚少，因此还需要全世界科学家的共同努力，不断提高人工神经网络的智能模拟能力。

人工神经网络虽然没有生物神经网络（以人脑为例）那么复杂，但它们之间有两个关键相似之处。第一个相似之处是人工神经网络和生物神经网络（以人脑为例）的结构都可以看成是神经元（可计算单元）的高度互连而成的网络，当然，人工神经元比生物神经元简单得多。第二个相似之处是神经元之间的连接决定了网络的功能。

一个非常有趣但是又非常值得注意的现象是，虽然生物神经元的运算速度（大约是毫秒级）相对于电子电路（至少是纳秒级）来说非常慢，但是在许多方面，人脑却能以比现有计算机快得多的速度完成各种复杂任务。那么为什么会这样？一种观点认为生物神经网络具有巨大的并行性，即所有的神经元能同时操作。如果这一观点成立的话，就意味着神经网络与巨型系统的并行操作密切相关。巨型系统的并行操作向传统的冯·诺依曼计算机体系结构提出了新的挑战。

人工神经网络中的一个主要研究领域就是建立神经网络的拓扑结构并采用合适的连接（称为权）来解决特定的实际问题。

2.2 人工神经元模型

2.2.1 单输入单输出人工神经元

一个单输入单输出的人工神经元如图 2.2 所示。

从图 2.2 中可以看出，标量输入 x_0 经过预处理函数单元 I ，预处理单元 I 的输出为

$$y_0 = I(x_0) \quad (2.2.1)$$

标有 a 的方框称为输入函数或激活函数，其输入输出关系为

$$x_1 = a(w, y_0, \theta) \quad (2.2.2)$$

其中输入函数（激活函数） a 可以是任意的函数， w 称为权（或权值，在这里是标量）， θ 称为阈值。权值和阈值都是可以调整的参数。如果不想在神经元中使用阈值，可以忽略它。在后面的章节中将会出现这样的情况。输入函数的输出 x_1 通常被称为净输入，它被送入一个变换函数（或传输函数） f ，在 f 中产生神经元的标量输出 y_1 。

标有 f 的圆圈称为人工神经元的变换函数，其输入输出关系为