

混合神经网络技术

田雨波 编著



科学出版社

www.sciencep.com

混合神经网络技术

田雨波 编著

科学出版社

北京

内 容 简 介

本书在论述神经网络基本概念和基本原理的基础上,重点介绍了混合神经网络技术,同时,给出各种混合神经网络技术在电磁建模和优化问题中的应用。全书共分12章,内容主要包括神经网络的基本概念、基础知识、BP神经网络、RBF神经网络、Hopfield神经网络、随机神经网络、遗传神经网络、粒子群神经网络、模糊神经网络、混沌神经网络、小波神经网络和神经网络集成等。同时,书后附录给出相关程序。

本书可供从事神经网络理论与技术、计算电磁学、电磁场工程等领域研究和开发工作的科技人员和高校教师参考阅读,也可作为高等院校相关专业的高年级本科生和研究生的教学用书。

图书在版编目(CIP)数据

混合神经网络技术/田雨波编著.—北京:科学出版社,2009

ISBN 978-7-03-024813-8

I.混… II.田… III.人工神经-神经网络 IV.TP18

中国版本图书馆CIP数据核字(2009)第101013号

责任编辑:王志欣 孙芳 于宏丽 / 责任校对:陈丽珠

责任印制:赵博 / 封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

印刷

科学出版社发行 各地新华书店经销

*

2009年6月第一版 开本:B5(720×1000)

2009年6月第一次印刷 印张:24 1/2

印数:1—3 000 字数:480 000

定价:70.00元

(如有印装质量问题,我社负责调换(双青))

前 言

人工神经网络是在对人脑认识的基础上,以数学和物理方法及从信息处理的角度对人脑生物神经网络进行抽象并建立起来的某种简化模型,它是计算智能和机器学习研究的最活跃的分支之一。近年来,神经网络在理论研究、实现技术和应用研究等方面取得了引人注目的成果,为此,国内外已经出版了有关神经网络方面的著作、教材、论文集等。信息科学与包括生命科学在内的其他智能技术的相互交叉、相互渗透和相互促进是现代科学技术发展的一个显著特点。神经网络与各种智能信息处理方法有机结合具有很大发展前景,如与模拟退火算法、遗传算法、粒子群算法、模糊理论、混沌理论和小波分析等相结合,即形成所谓的“混合神经网络技术”,目前已经成为一大研究热点。人们希望通过这些理论和算法与神经网络相互混合,获得具有柔性信息处理功能的系统,但关于这方面的书籍并不是很多,这正是本书的主要特点之一。

计算电磁学是在 20 世纪 60 年代随着电子计算机技术的发展而诞生的,它是在电磁学、计算数学和计算机科学的基础上产生的边缘交叉学科。计算电磁学实质上是以电磁场理论为基础,以高性能计算技术为手段,运用计算数学提供的各种方法解决复杂电磁场理论和工程问题的应用科学。经过几十年的发展,计算电磁学内容已经非常丰富,影响非常深广,以致所有与电磁场相关的领域都因其发展而受益,其中,不少领域由于运用了计算电磁学的方法而使其面貌完全改观。然而,对于复杂的电磁系统,对其进行严格的电磁仿真耗时而费力,在保证计算精度的情况下对其进行快速而精确的建模和优化必将成为计算电磁学的发展趋势,而这在一定程度上又是建立在神经网络技术基础之上的。本书在讲述混合神经网络技术的同时,重点给出各种混合神经网络在电磁学方面的应用,这也是本书的另外一个特点。

本书是作者从事神经网络理论与技术和计算电磁学建模及优化的教学和科研工作的系统总结,并从国内外相关文献资料中提取最主要的理论及成果,力图反映最新的研究动态,清楚阐述混合神经网络技术及这些技术在电磁问题的数值仿真、高效建模和优化设计中的具体应用。全书共分 12 章。第 1 章介绍神经网络的基本概念、基本功能、基本性质及性能指标、研究内容、发展趋势等,同时,综述神经网络在电磁方面的应用。第 2 章介绍神经网络的基础知识,包括神经网络的基本模型、训练与学习、泛化能力等。第 3 章介绍 BP 神经网络,它是最具代表性的前馈神经网络模型之一,其中,包括 BP 神经网络的网络结构、学习算法、应用要

点、不足及改进等,同时,应用 BP 神经网络对微带天线进行结构设计。第 4 章介绍 RBF 神经网络,它是另外一种具有代表性的前馈神经网络模型,包括 RBF 神经网络的网络结构、学习算法、网络特点等,并与 BP 神经网络作了对比说明。第 5 章介绍 Hopfield 神经网络,它是最常见的反馈神经网络,包括神经动力学和 Lyapunov 定理、连续 Hopfield 神经网络、离散 Hopfield 神经网络等。第 6 章介绍随机神经网络,包括 Boltzmann 机、神经网络的随机训练、模拟退火算法等。第 7 章介绍遗传神经网络,包括遗传算法基本原理及遗传神经网络的实现等,同时,应用遗传 RBF 神经网络解决了自适应波束形成问题。第 8 章介绍粒子群神经网络,包括粒子群优化算法基本原理及粒子群神经网络的实现等。第 9 章介绍模糊神经网络,包括模糊理论基本知识及模糊神经网络的实现等,同时,应用模糊神经网络进行了波导匹配负载设计和微带天线谐振频率计算。第 10 章介绍混沌神经网络,包括混沌理论基本知识及混沌神经网络的实现等,同时,应用混沌神经网络进行了移动通信系统信道分配和自适应雷达目标信号处理。第 11 章介绍小波神经网络,包括小波分析基本知识及小波神经网络的实现等,同时,应用小波神经网络解决了飞机图像识别和微带不连续问题。第 12 章介绍神经网络集成,包括神经网络集成的基本概念、实现方法、理论分析等,同时,应用神经网络集成进行了股市预测、肺癌诊断和谐振频率计算。书末附录给出本书中应用的大部分程序,方便读者理解及使用。另外,本书在编著过程中参阅了大量的国内外文献,在此对作者深表感谢。

本书是在江苏省“青蓝工程”优秀青年骨干教师项目、江苏省高校自然科学基金基础研究项目(项目编号:07KJB510032)、江苏科技大学研究生部出版基金的资助下完成的,在此对上述资助单位表示诚挚谢意。同时,对在本书创作过程中给予作者大力支持的江苏科技大学电子信息学院的领导及各位同仁由衷地表示感谢。

由于作者水平有限,不妥之处在所难免,敬请读者批评指正。

作者

2009 年 3 月

目 录

前言

第 1 章 绪论	1
1.1 神经网络的概念与分类	1
1.2 神经网络的基本特征和基本功能	3
1.3 神经网络的基本性质、优点及应用	5
1.4 神经网络的性能指标及研究内容	6
1.5 神经网络的发展简史、存在问题及发展趋势	7
1.6 神经网络的电磁应用	11
参考文献	12
第 2 章 基础知识	16
2.1 神经网络模型	16
2.2 神经网络的训练和学习	22
2.3 神经网络的泛化能力	26
2.4 神经网络训练用样本	27
参考文献	32
第 3 章 BP 神经网络	33
3.1 BP 神经网络结构	33
3.2 BP 学习算法	35
3.3 BP 神经网络应用要点	38
3.4 BP 算法的不足及改进	42
3.5 应用 BP 神经网络进行微带贴片天线设计	48
参考文献	57
第 4 章 RBF 神经网络	60
4.1 网络结构和工作原理	60
4.2 网络的生理学基础和数学基础	62
4.3 常用的学习算法	66
4.4 网络的特点及注意事项	71

4.5 RBF神经网络与BP神经网络的比较	72
参考文献	73
第5章 Hopfield神经网络	74
5.1 Hopfield神经网络简介	74
5.2 神经动力学	74
5.3 Lyapunov定理	75
5.4 连续Hopfield神经网络	77
5.5 离散Hopfield神经网络	80
5.6 Hopfield神经网络应用	82
5.7 Hopfield神经网络特点	84
参考文献	84
第6章 随机神经网络	86
6.1 Boltzmann机	86
6.2 神经网络的随机训练	95
6.3 模拟退火算法	99
参考文献	103
第7章 遗传神经网络	104
7.1 遗传算法	104
7.2 遗传神经网络原理及实现	125
7.3 遗传神经网络应用	138
参考文献	144
第8章 粒子群神经网络	148
8.1 粒子群优化算法	148
8.2 粒子群神经网络原理及实现	171
8.3 粒子群神经网络应用	176
参考文献	179
第9章 模糊神经网络	183
9.1 模糊理论	183
9.2 模糊神经网络原理及实现	194
9.3 模糊神经网络应用	206
参考文献	230

第 10 章 混沌神经网络	234
10.1 混沌理论	234
10.2 混沌神经网络原理及实现	253
10.3 混沌神经网络应用	266
参考文献	279
第 11 章 小波神经网络	283
11.1 小波分析	283
11.2 小波神经网络原理及实现	297
11.3 小波神经网络应用	308
参考文献	318
第 12 章 神经网络集成	321
12.1 神经网络集成基本知识	321
12.2 神经网络集成的应用	328
参考文献	343
附录	346
附录 1 BP 神经网络源程序	346
附录 2 基于梯度算法的 RBF 神经网络源程序	349
附录 3 基于聚类法的 RBF 神经网络源程序	351
附录 4 基于正交最小二乘算法的 RBF 神经网络源程序	353
附录 5 遗传算法源程序	355
附录 6 粒子群算法源程序	359
附录 7 粒子群算法优化神经网络源程序	361
附录 8 粒子群算法和 BP 算法相结合优化神经网络源程序(1)	367
附录 9 粒子群算法和 BP 算法相结合优化神经网络源程序(2)	369
附录 10 小波神经网络源程序	372
附录 11 基于十进制粒子群优化算法的神经网络集成源程序	376
附录 12 基于二进制粒子群优化算法的神经网络集成源程序	381

第 1 章 绪 论

人类具有高度发达的大脑,大脑是思维活动的物质基础,而思维是人类智能的集中体现。长期以来,人们想方设法了解人脑的工作机理和思维本质,向往构造出人工智能系统来模仿人脑的功能,其中的一个重要成果就是人工神经网络(artificial neural networks, ANN)。本章主要讲述人工神经网络的概念、基本性质、基本功能和应用,以及人工神经网络的性能指标、研究内容和发展趋势,最后介绍人工神经网络的电磁应用。

1.1 神经网络的概念与分类

1.1.1 神经网络的概念

人工神经网络也叫神经网络(neural network)、人工神经系统(artificial neural systems)、自适应系统(adaptive systems)、自适应网(adaptive networks)、连接模型(connectionism)、神经计算机(neurocomputer)等,它是在对人脑认识的基础上,以数学和物理方法及从信息处理的角度对人脑生物神经网络(biological neural network, BNN)进行抽象并建立起来的某种简化模型。它是对人类大脑系统特性的一种描述,是由多个非常简单的处理单元彼此按某种方式连接而形成的计算机系统,该系统是靠其状态对外部输入信息的动态响应来处理信息的。简单地讲,人工神经网络是一个数学模型,可以用电子线路来实现,也可以用计算机程序来模拟,是人工智能研究的一种方法。

1.1.2 神经网络的分类

到目前为止,神经网络已经有几十种不同的模型,按照不同的原则,可以对神经网络进行不同的分类,通常有以下 5 类^[1~4]。

- (1) 按照网络的结构进行分类:前馈网络、反馈网络 and 自组织网络。
- (2) 按照学习方式进行分类:有导师学习网络和无导师学习网络。
- (3) 按照网络的性能进行分类:连续型网络和离散型网络、随机型网络和确定型网络。
- (4) 按照突触性质进行分类:一阶线性并联网络和高阶非线性并联网络。
- (5) 按照对生物神经系统的层次模拟分类:神经元层次模型、组合式模型、网

络层次模型、神经系统层次模型和智能型模型。

在人工神经网络的设计与应用过程中,人们较多地考虑神经网络的互联结构,包括5种典型结构^[5],如图1.1所示。

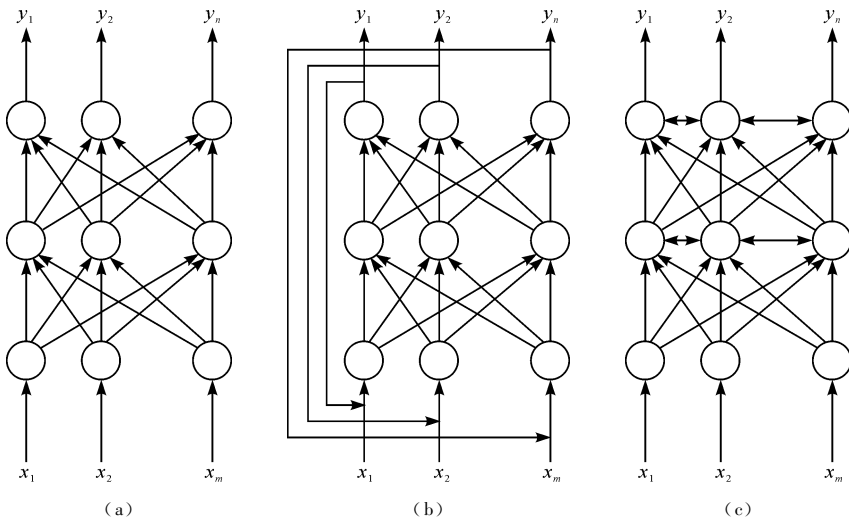
(1) 前馈网络。前馈网络中,神经元是分层排列的,每个神经元只与前一层的神经元相连,如图1.1(a)所示。最上一层为输出层,最下一层为输入层,还有中间层,中间层也称为隐层,隐层的层数可以是一层或多层。

(2) 输入输出有反馈的前馈网络。如图1.1(b)所示,在输出层上存在一个反馈回路到输入层,而网络本身还是前馈型的。该种神经网络的输入层不仅接受外界的输入信号,也接受网络自身的输出反馈信号。输出反馈信号可以是原始输出信号,也可以是经过转化的输出信号;可以是本时刻的信号,也可以是经过一定延迟的信号。此种网络经常用于系统控制、实时信号处理等需要根据系统当前状态进行调节的场合。

(3) 前馈内层互连网络。如图1.1(c)所示,在同一层内存在互相连接,它们可以形成互相制约,而从外部看还是一个前向网络,很多自组织网络大都存在内层互连的结构。

(4) 反馈型全互连网络。图1.1(d)所示的网络是一种单层全互连网络,每个神经元的输出都与其他神经元相连,如Hopfield神经网络和Boltzmann机都是属于这一类网络。

(5) 反馈型局部连接网络。图1.1(e)所示的网络是一种单层网络,它的每个神经元的输出只与其周围的神经元相连形成反馈的网络,这类网络也可发展为多层的金字塔形的结构。



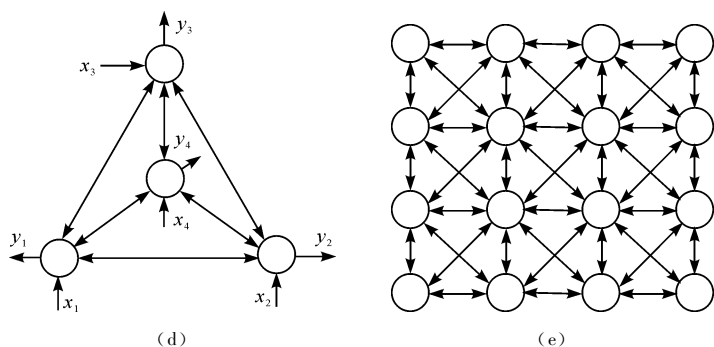


图 1.1 神经网络的拓扑结构

目前,最具代表性的前馈神经网络模型有:①BP(back propagation,反向传播)神经网络。它是一种多层前馈网络,采用最小均方差的学习方式,是使用最广泛的网络。这个网络的缺点是为有导师训练,训练时间较长,易于陷入局部极小等。②RBF(radial basis function,径向基函数)神经网络。它是一种非常有效的多层前馈网络,其神经元基函数具有仅在微小局部范围内才产生有效的非零响应的局部特性,因而可以在学习过程中获得高速化。这个网络的缺点是由于高斯函数的特性,该网络难以学习映射的高频部分。最具代表性的反馈网络模型是 Hopfield 神经网络,该网络是由相同的神经元构成的单层网络,并且是不具学习功能的自联想网络,它需要对称连接。这个网络可以完成制约优化和联想记忆(associative memory, AM)等功能。

1.2 神经网络的基本特征和基本功能

1.2.1 神经网络的基本特征

神经网络的基本特征可归结为结构特征和能力特征。

(1) 结构特征——并行处理、分布式存储与容错性。人工神经网络是由大量的简单处理单元相互连接构成的高度并行的非线性系统,具有大规模并行性处理特征。虽然每个处理单元的功能十分简单,但大量简单处理单元的并行活动使网络呈现出丰富的功能,并具有较快的速度。结构上的并行性使神经网络的信息存储必然采用分布式方式,即信息不是存储在网络的某个局部,而是分布在网络所有的连接权中。神经网络内在的并行性与分布性表现在其信息的存储与处理都是在空间上分布、时间上并行的,这两个特点必然使神经网络在两个方面表现出良好的容错性:一方面由于信息的分布式存储,当网络中部分神经元损坏时不会对系统的整体性能造成影响,这一点就像人脑中每天都有神经细胞正常死亡而不

会影响大脑的功能一样;另一方面当输入模糊、残缺或变形的信息时,神经网络能够通过联想恢复出完整的记忆,从而实现对不完整输入信息的正确识别,这一点就像人可以对不规则的手写字进行正确识别一样。

(2) 能力特征——自学习、自组织与自适应性。自适应性是指一个系统能够改变自身的性能以适应环境变化的能力,它是神经网络的一个重要特征。自适应性包含自学习与自组织两层含义。神经网络的自学习是指当外界环境发生变化时,经过一段时间的训练或感知,神经网络能够通过自动调整网络结构参数,对给定输入能产生期望的输出。训练是神经网络学习的途径,因此,经常将学习与训练两个词混用,神经系统能在外部刺激下按一定规则调整神经元之间的突触连接,逐渐构建起神经网络,这一构建过程称为网络的自组织。神经网络的自组织能力与自适应性相关,自适应性是通过自组织实现的。

1.2.2 神经网络的基本功能

(1) 联想记忆。由于神经网络具有分布存储信息和并行计算的性能,因此,它具有对外界刺激信息和输入模式进行联想记忆的能力,这种能力是通过神经元之间的协同结构及信息处理的集体行为实现的。神经网络是通过其突触权值和连接结构来表达信息的记忆,这种分布式存储使神经网络能存储较多的复杂模式和恢复记忆的信息。神经网络通过预先存储信息和学习机制进行自适应训练,可以从不完整的信息和噪声干扰中恢复原始的完整信息,这一能力使其在图像复原、图像和语音处理、模式识别、分类等方面具有巨大的潜在应用价值。

(2) 非线性映射。在客观世界中,许多系统的输入与输出之间存在复杂的非线性关系,对于这类系统,往往很难用传统的数理方法建立其数学模型。设计合理的神经网络通过对系统输入输出样本对进行自动学习,能够以任意精度逼近任意复杂的非线性映射。神经网络的这一优良性能使其可以作为多维非线性函数的通用数学模型,该模型的表达是非解析的,输入输出数据之间的映射规则由神经网络在学习阶段自动抽取并分布式存储在网络的所有连接中。

(3) 分类与识别。神经网络对外界输入样本具有很强的识别和分类能力。对输入样本的分类实际上是在样本空间找出符合分类要求的分割区域,每个区域内的样本属于一类。传统分类方法只适合解决同类相聚、异类分离的识别与分类问题,但客观世界中,许多事物在样本空间上的区域分割曲面是十分复杂的,相近的样本可能属于不同的类,而远离的样本可能同属一类。神经网络可以很好地解决对非线性曲面的逼近,因此,比传统的分类器具有更好的分类和识别能力。

(4) 优化计算。优化计算是指在已知的约束条件下寻找一组参数组合,使由该组合确定的目标函数达到最小值。某些类型的神经网络可以把待求解问题的可变参数设计为网络的状态,将目标函数设计为网络的能量函数,神经网络经过

动态演变过程达到稳定状态时对应的能量函数最小,从而其稳定状态就是问题的最优解。这种优化计算不需要对目标函数求导,其结果是网络自动给出的。

(5) 知识处理。神经网络获得知识的途径与人类相似,也是从对象的输入输出信息中抽取规律而获得关于对象的知识,并将知识分布在网络的连接中予以存储。神经网络的知识抽取能力使其能够在没有任何先验知识的情况下自动地从输入数据中提取特征,发现规律,并通过自组织过程将自身构建成适合于表达所发现的规律。另一方面,人的先验知识可以大大提高神经网络的知识处理能力。两者相结合会进一步提升神经网络的智能。

1.3 神经网络的基本性质、优点及应用

1.3.1 神经网络的基本性质

神经网络的基本性质主要包括收敛性、容错性、鲁棒性及推广性等。

神经网络的收敛性是指神经网络的训练算法在有限次迭代之后可收敛到正确的权值或权向量。神经网络良好的容错性保证网络将不完整的、污损的、畸变的输入样本恢复成完整的原型。容错性的研究归结于神经网络动力系统记忆模式吸引域的大小,吸引域越大,网络从部分信息恢复全部信息的能力越大,表明网络的容错性越大。神经网络的高度鲁棒性使网络中的神经元或突触遭到破坏时网络仍然具有学习和记忆能力,从而使网络表现出高度的自组织性。研究表明,如果记忆模式的吸引域越“规则”,网络抵抗干扰、噪声或自身损害的能力就越强,即鲁棒性越好。训练好的神经网络应能够对不属于训练样本集合的输入样本正确识别或分类,这种现象常称为神经网络具有良好的推广性。

人工神经网络的操作有两种过程:一是训练学习,二是正常操作或称回忆。训练时,把要教给网络的信息(外部输入)作为网络的输入和要求的输出,使网络按某种规则(训练算法)调节各处理单元间的连接权值,直到加上给定输入后网络就能产生给定输出为止。这时,各连接权已经调节好,网络的训练完成。所谓正常操作,就是对训练好的网络输入一个信号,它就可以正确回忆出相应输出。不论是训练网络还是操作网络,人工神经网络的状态总是变化的。所谓神经网络的状态,是指神经网络所有节点的输出信号值。状态变化可以指某个节点的状态变化,也可指所有节点的状态变化。神经网络的这种动态特性受两种性能的约束,即系统的整体稳定性和收敛性。所谓稳定的神经网络,定义为这样一种非线性动态系统:当在该系统上加入一初始输入时,系统的状态发生变化,但最后达到一固定点(收敛点或均衡点),这些固定点就是存储信息的点。虽然稳定的神经网络总能保证所有输入被映射到固定点,但不能保证该固定点就是要求的固定点。不难

理解,神经网络的稳定性是与反馈网络的回忆操作相联系的,这种反馈网络的稳定性可以用 Lyapunov 准则进行判定。收敛性是指在训练过程中,输出节点的实际输出值与要求的输出值之间的误差,最后能达到可接受的最小值。一般要求收敛过程迅速和精确,即输出能尽快趋于目标值。显然,收敛性是与有指导的训练操作相联系,收敛过程严格依赖于所采用的具体训练算法和训练参数。

1.3.2 神经网络的优点

- (1) 很强的鲁棒性和容错性。这是因为信息是分布存储于网络内的神经元中。
- (2) 并行处理方法。人工神经网络在结构上是并行的,而且网络的各个单元可以同时进行类似的处理过程,使计算快速。
- (3) 自学习、自组织、自适应性。神经元之间的连接多种多样,各神经元之间连接强度具有一定可塑性,使得神经网络可以处理不确定或不知道的系统。
- (4) 可以充分逼近任意复杂的非线性关系。
- (5) 具有很强的信息综合能力。能同时处理定量和定性的信息,能很好地协调多种输入信息关系,适用于处理复杂非线性和不确定对象。

1.3.3 神经网络的应用

神经网络以其独特的结构和处理信息的方法,在许多实际应用领域中取得了显著的成效。主要应用于自动控制、处理组合优化问题、模式识别、图像处理、传感器信号处理、机器人控制、信号处理、卫生保健、医疗、经济、化工、焊接、地理、数据挖掘、电力、交通、军事、矿业、农业和气象等领域。

1.4 神经网络的性能指标及研究内容

1.4.1 神经网络的性能指标

与人脑的作用机理类似,一个神经网络完成任务的过程包括学习(训练)过程和使用(回忆或联想)过程。对一个神经网络学习算法来说,衡量其性能优劣的指标有以下几个方面。

- (1) 泛化能力。一个训练好的神经网络到实际中使用是否有好的效果,这是神经网络最重要的性能指标。
- (2) 时间复杂性。训练一个固定结构的神经网络所需要的时间。
- (3) 空间复杂性。算法计算机实现时所占用的内存空间,一般与神经网络的结构复杂程度密切相关。
- (4) 在线学习能力。如果神经网络的学习过程和使用过程是分别进行的,即

先学习后使用,则称为离线学习;如果这两个过程是同时进行的,即边学习边使用,则称为在线学习能力。

(5) 其他指标。包括能否用硬件实现、算法的稳定性、神经网络模型的鲁棒性等。

1.4.2 神经网络的研究内容

当前,神经网络研究内容主要包括神经网络理论研究、神经网络实现技术研究和神经网络应用研究三个方面。

(1) 神经网络理论研究。神经网络理论研究侧重于寻找合适的神经网络模型和学习算法。其中,模型研究是指构造合适的单个神经元模型及确定神经元之间的连接方式,并探讨它所适用的场合;学习算法研究是指在神经网络模型的基础上找出一种调整神经网络结构和权值的算法,并满足学习样本的要求,同时具有较快的学习速度。神经网络理论研究的另一个重要内容是从理论上分析常用的神经网络设计方法对泛化能力的影响。

(2) 神经网络实现技术研究。神经网络实现技术研究主要是探讨利用电子、光学、光电、生物等技术实现神经计算机的途径,包括利用传统计算机技术实现模拟神经计算机及新型神经计算机体系结构的研究等。

(3) 神经网络应用研究。神经网络应用研究是探讨如何利用神经网络解决实际工程问题。人们可以在几乎所有的领域中发现神经网络应用的影子。当前,神经网络的主要应用领域有模式识别、故障检测、智能机器人、非线性系统辨识和控制、市场分析、决策优化、物资调用、智能接口、知识处理和认知科学等。

1.5 神经网络的发展简史、存在问题及发展趋势

1.5.1 神经网络的发展简史

神经网络是一门活跃的边缘性交叉学科,研究它的发展过程和前沿问题具有重要意义。目前,神经网络的理论和应用研究得到了极大的发展,而且已经渗透到几乎所有的工程应用领域。但是,人工神经网络的发展过程并不是一帆风顺的,大致经历了以下几个阶段。

(1) 初始时期。1943年,McCulloch和Pitts提出了MP模型^[6],从而给出了神经元的最基本模型及相应的工作方式。1949年,神经生物学家Hebb发现,脑细胞之间的通路在参与某种活动时将被加强^[7],这个重要规则给出了生理学与心理学间的联系,被称为Hebb学习规则,该规则至今还被许多神经网络学习算法所使用。1958年,Rosenblatt提出了感知器模型^[8],这是一个由线性阈值神经元组

成的前馈神经网络,可用于分类。1960年,Widrow和Hoff提出了自适应线性单元^[9],这是一种连续取值的神经网络,可用于自适应系统。

(2) 低潮时期。1969年,人工智能的创始人Minsky和Papert出版了*Perceptrons*^[10],在该书中,他们指出单层感知器只能作线性划分,多层感知器不能给出一种学习算法,因此无实用价值。由于Minsky和Papert在人工智能领域的地位,该书在人工神经网络研究人员间产生了极大的反响,神经网络研究受到了严重影响,自此陷入低潮。但是,即便在神经网络研究的低潮时期,也仍有一些人在兢兢业业地研究神经网络,并取得了一些重要成果。其中,最著名的是1982年加利福尼亚理工大学教授Hopfield提出的Hopfield神经网络^[11]。在这个用运算放大器搭成的反馈神经网络中,Hopfield借用Lyapunov能量函数的原理,给出了网络的稳定性判据,并为著名的组合优化问题——旅行商问题(TSP)提供了一个新的解决方案。Hopfield神经网络可用于联想存储、优化计算等领域。

(3) 高潮时期。1986年,Rumelhart等给出了多层感知器的权值训练BP算法^[12],从而解决了Minsky认为不能解决的多层感知器的学习问题,自此引导了神经网络的复兴,神经网络的研究也进入了一个崭新的发展阶段。

1.5.2 神经网络的一些问题

随着对神经网络研究的广泛关注,其中的一些问题逐渐暴露出来,并已成为该学科进一步发展的障碍。目前,亟待从以下几个方面予以改进^[13]。

(1) 加快神经网络的学习速度。目前,绝大多数神经网络算法都需要进行耗时的迭代训练,其计算开销相当大,训练速度太慢,难以满足实时性要求较高的在线学习任务的需要。此外,太慢的速度还使得神经计算技术很难用于数据挖掘等领域,因为等到网络训练完成时,数据库的内容可能已经发生了更新变动,网络学习到的知识将无法反映出当前事物的特点。

(2) 增强神经网络的可理解性。神经网络模型的一大特点是其分布式知识表示,即网络中单一的神经元或连接并没有明确的意义,这就决定了神经网络是一种典型的“黑箱”模型,其学习到的知识隐藏在大量的连接权值中,用户无法知道某一个具体的网络能做什么,也无法知道它是如何做的。一般来说,“可解释性”是可靠系统的必备特性,由于通常的神经网络模型都是“不可解释”的,这在一定程度上影响了用户对通过神经计算技术构建智能系统的信心。虽然Baum和Hausler指出^[14]，“如果一个神经网络可以为大量的训练例产生正确的解答,那么,可以相信它们也能为类似于训练例的未知例产生正确解答”,但这并没有抵消用户对可理解性的偏好。此外,训练好的神经网络学习到的知识不能以容易理解的方式提交给决策者,这也是神经计算技术难以用于数据挖掘领域的主要原因之一。

(3) 设计出易于使用的工程化神经计算方法。神经计算由于缺少一个统一的理论框架,经验性成分相当高。虽然 Hornik 等证明^[15],仅有一个非线性隐层的前馈网络就可以任意精度逼近任意复杂度的函数,但一些研究者指出,对网络的配置和训练是 NP 问题^[16,17]。这就使得在利用神经计算解决问题时,只能采取具体问题具体分析的方式,通过大量费力耗时的实验摸索,确定出合适的神经网络模型、算法及参数设置,其应用效果完全取决于使用者的经验。即使采用同样的方法解决同样的问题,由于操作者不同,其结果很可能大相径庭。在实际应用中,操作者往往是缺乏神经计算经验的普通工程技术人员,如果没有易于使用的工程化神经计算方法,神经计算技术的应用效果将很难得到保证。

(4) 更好地模拟生物神经系统。由于神经网络产生于对生物神经系统的模拟,因此,人们希望它能具有生物神经系统的各种优良特性。然而,目前的神经计算模型却没能做到这一点。例如,从容错性的角度来说,生物神经系统的容错性相当好,尽管每天约有 10^4 个脑细胞死亡,人脑仍然能正常工作。虽然神经网络一般采用分布式知识表示,网络由多个功能相同或相似的神经元组成,但是,正如一些研究者所指出的^[18,19],现有的神经网络学习算法,尤其是前馈神经网络学习算法,并没有充分利用分布式知识表示中的冗余信息,由它们训练出的网络在本质上并不具有容错性,网络的容错能力需要通过额外的机制加以改善。

(5) 将神经计算与传统人工智能技术相结合。传统人工智能技术在逻辑推理等许多方面都是很有有效的,神经计算绝不可能完全替代它们,而只能在某些方面与之互补。Minsky 指出了利用不同的组件构建智能系统的必要性^[20]:“人工智能研究必须从其传统关注的特殊模式走出来。世界上并不存在一种最佳的知识表示或问题求解方法。当前机器智能的局限性在很大程度上是由以下两方面造成的,即力图寻找统一的理论,或者试图弥补那些在理论上很漂亮但在概念上却很虚弱的方法之不足……我们所需的多功能性智能在更大规模的结构中找到,这些结构应能够同时利用和管理若干种知识表示的优势,使得各种类型的表示可以相得益彰。”目前,很多研究者对将符号学习与神经学习相结合非常关注,因为如果这两者能够很好地结合起来,就可以在一定程度上模拟人类逻辑思维和直觉思维的统一,这将是人工智能领域的重大突破。

1.5.3 神经网络的发展趋势

近年来,神经网络理论与实践有了引人注目的进展,它再一次拓展了计算概念的内涵,使神经计算、进化计算成为新的学科,神经网络的软件模拟得到了广泛应用。近几年来,科技发达国家的主要公司对神经网络芯片、生物芯片情有独钟。展望未来,神经网络的研究趋势将主要侧重于以下几个方面^[21~23]。

(1) 促进神经生理学、神经解剖学的研究。通过神经网络研究的发展,人们

对人脑一些局部功能的认识已经有所提高,如对感知器的研究、对视觉处理网络的研究、对存储与记忆问题的研究等都取得一定的成功。遗憾的是,这些成功一方面还远不够完善,另一方面在对人脑作为一个整体的功能的解释上几乎起不到任何作用。科学家已经积累了大量关于大脑组成、大脑外形、大脑运转基本要素等知识,但仍无法解答有关大脑信息处理的一些实质问题。整体功能绝不是局部功能的简单组合,而是一个巨大的质的飞跃,人脑的知觉和认知等过程是包含着一个复杂的动态系统中对大量神经元活动进行整合的统一性行动。由于人们对人脑完整工作过程几乎没有什么认识,连一个稍微完善的可令人接受的假设也没有,造成了神经网络研究始终缺乏一个明确的大方向。这方面如果不能有所突破,神经网络研究将始终限于模仿人脑局部功能的缓慢摸索过程当中,而难以达到研究水平的质的飞跃。

(2) 发展与之相关的数学领域。神经元以电为主的生物过程在认识上一般采用非线性动力学模型,其动力学演变过程往往是非常复杂的,神经网络这种强的生物学特征和数学性质要求有更好的数学手段。而对解决非线性微分方程这样的问题,稍微复杂一些的便没有办法利用数学方法求得完整的解。这使得在分析诸如一般神经网络的自激振荡、稳定性、混沌等问题时常常显得力不从心,更不用说当我们面对人脑这样的由成千上万个神经网络子系统组成的巨系统,而每个子系统(具有某种特定功能)又可能由成千上万个神经元组成,每个神经元本身是一个基本的非线性环节。因此,当今神经网络理论的发展,已经客观要求有关数学领域必须有所发展,并预期一种更简洁、更完善和更有效的非线性系统表达与分析的数学方法是这一领域数学发展的主要目标之一。

(3) 扩大神经网络结构和神经元芯片的作用。神经网络结构体现了算法和结构的统一,是硬件和软件的混合体。未来的研究主要是针对信息处理功能体,将系统、结构、电路、器件和材料等方面的知识有机结合起来,建构有关的新概念和新技术,如结晶功能体、高分子功能体等。生物芯片由于元件是分子大小的,其包装密度可成数量级增加,它的信号传播方式是孤电子,将不会有损耗,并且几乎不产生热量。因此,随着大量神经计算机和神经元芯片应用于高科技领域,它有着更诱人的前景。

(4) 增强对智能和机器关系问题的认识。神经网络是由大量处理单元组成的非线性、自适应、自组织系统,它是在现代神经科学研究成果的基础上提出的,试图模拟神经网络加工、记忆信息的方式,设计一种新的机器,使之具有人脑风格的信息处理能力。对于智能和机器的关系,应该从进化的角度,把智能活动看成动态发展的过程,并合理地发挥经验的作用。同时,还应该从环境与社会约束及历史文化约束的角度加深对它的理解与分析。鉴于此,智能理论的发展方向是把基于连接主义的神经网络理论、基于符号主义的人工智能专家系统理论和基于进

化论的人工生命理论这三大研究领域,在共同追求的总目标下,自发而有机地结合起来。

(5) 发展神经计算和进化计算的理论与应用。20 世纪 80 年代以后,神经网络理论在计算理论方面取得了引人注目的成果,形成了神经计算和进化计算新概念,激起了许多理论家的强烈兴趣。离散符号计算、神经计算和进化计算相互促进,也许最终导致这三种计算统一起来,这应是我们无法回避的一个重大难题。在 21 世纪,关于这个领域的研究会产生新的概念和方法。

(6) 促进信息科学与生命科学的相互融合。信息科学与生命科学的相互交叉、相互渗透和相互促进是现代科学技术发展的一个显著特点。神经网络与各种智能信息处理方法有机结合具有很广阔的发展前景,如与模糊逻辑、混沌理论、遗传进化算法、粒子群算法、免疫算法等相结合,即所谓“混合神经网络”方法。由于这些理论和算法都是属于仿效生物体信息处理的方法,人们希望通过它们之间的相互结合,能够获得具有柔性信息处理功能的系统,这也是本书的主要特点。

1.6 神经网络的电磁应用

在微波集成电路设计中,现代计算机辅助设计(computer aided design, CAD)方法的效果取决于有源及无源电路元件模型的准确性。随着电路密度及工作频率的增高,传统建模技术的准确性变得不可靠。例如,微波电路中存在大量的不连续性,为这些结构建立电磁场边界值方程和求解这样的方程均十分困难。若采用数值计算方法,则计算量、存储量与准确性成正比。在优化设计过程中,往往要反复计算电路的特性,若都用数值计算,耗时太多。建立近似等效电路模型,则不可能在所有条件下均合适。再如,典型的电路模拟软件所提供的无源元件模型不能准确地考虑寄生和耦合效应。为补救这一缺陷,人们采用建立无源元件库的办法制作、测量、存储成百上千组元件数据在一个表中,但建表成本高,占用内存资源大。另外,在任何一个高速 VLSI 系统中,为保证系统正常工作,必须进行互联结构的仿真与优化设计。信号时延、串扰、接地板反弹噪声等重要的信号整体特性的优劣取决于系统中的互连网络及互连电路。目前,高速互连系统的分析是一项高强度的计算工作,耗费很长计算时间,占用大量的内存资源。目前在工程设计时,一般采用由测量值拟合的经验公式或经验图表,准确性受到限制。其他的还有插值(quadratic interpolation)、GMDH(group method of data handling)、统计方法等,由于问题的模型要考虑的变量很多,且所需特性与诸多变量之间的关系往往是高度非线性的,上述方法的效果大多不甚理想。

从数学上看,一个电磁场问题的 CAD 模型就是一种映射关系 F ,即

$$Y = F(X) \quad (1.1)$$

式中, Y 为目标函数矢量; X 为输入矢量。通常, X 和 Y 之间的函数关系是多参量的、高度非线性的, 难以用简单函数直接给出。而神经网络模型却能有效、准确地描述这种映射关系, 并且计算方便、快速, 非常适合于面向 CAD 优化过程的复杂系统电磁场建模。

神经网络具有学习功能, 可模拟复杂的非线性函数映射关系, 许多类型的神经网络是通用逼近器, 能以适合的精度逼近任意输入输出映射, 已成为许多工程领域的有力工具。建立神经网络模型来逼近电路的输入输出响应, 在此基础上进行优化, 可克服传统设计优化中的困难。

神经网络近年来在电磁场领域也得到了广泛的应用^[24,25], 相关的著作已经出版^[26~29]。美国 Colorado 大学教授 Gupta 指出神经网络技术是电磁场复杂系统 CAD 的发展趋势之一^[30], 加拿大 Carleton 大学的 Zhang 教授对神经网络在射频和微波方面的设计从理论到实践进行了阐述^[31], 文献[32]、[33]也对神经网络在电磁场中应用的具体问题进行了说明。可以说, 神经网络已经被应用到电磁场领域的各个方面。文献[34]将 RBF 神经网络与全波三维电磁模拟软件 Quick-Wave 3D 相结合, 分析了波导弯头、T 形波导及开槽波导的 S 参数。文献[35]建立了微波神经网络数据库和微波神经网络工具箱, 使用它可以对微波电路进行仿真。文献[36]详细地讨论了知识人工神经网络, 并得到基于新型传输线方程的知识人工神经网络模型。文献[37]应用神经网络求解微波电磁场正向数值计算和微波电磁场激励源反演问题。神经网络可以应用于电磁成像^[38~40]、微波电路设计^[41~46]、微波器件设计^[47,48]、微波测量^[49]、波导匹配负载设计^[50]、频率选择表面设计^[51,52]、微带贴片天线设计^[53,54]、微带天线 (microstrip antenna) 谐振频率设计^[55]、天线阵来波方向估计^[56]、磁共振成像^[57]、放大器设计^[58]、滤波器设计^[59,60]等。

参 考 文 献

- [1] 胡守仁. 神经网络导论. 长沙: 国防科技大学出版社, 1993.
- [2] 焦李成. 神经网络系统理论. 西安: 西安电子科技大学出版社, 1996.
- [3] 魏海坤. 神经网络结构设计的理论与方法. 北京: 国防工业出版社, 2005.
- [4] 钟珞, 饶文碧, 邹承明. 人工神经网络及其融合应用技术. 北京: 科学出版社, 2007.
- [5] 张立明. 人工神经网络的模型及其应用. 上海: 复旦大学出版社, 1993.
- [6] McCulloch W S, Pitts W H. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 1943, (5): 115-133.
- [7] Hebb D O. *The Organization of Behavior*. New York: Wiley, 1949.
- [8] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958, (65): 386-458.
- [9] Widrow B, Hoff M E. *Adaptive Switching Circuits*. New York: IRE, 1960:94-104.
- [10] Minsky M, Papert S. *Perceptrons*. Cambridge: MIT Press, 1969.

-
- [11] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science*, 1982, (79): 2554-2558.
- [12] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagation error. *Nature*, 1986, (323): 533-536.
- [13] 周志华. 神经计算中若干问题的研究[博士学位论文]. 南京: 南京大学, 2000.
- [14] Baum E B, Haussler D. What size net gives valid generalization? *Neural Computation*, 1989, 1(1): 151-160.
- [15] Hornik K M, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 1989, 2(2): 359-366.
- [16] Judd J S. Learning in networks is hard//*Proceedings of the 1st IEEE International Conference on Neural Networks*, San Diego, 1987, 2: 685-692.
- [17] Blum A, Rivest R L. Training a 3-node neural networks is NP-complete. *Neural Networks*, 1992, 5(1): 117-127.
- [18] Nijhuis J, Hofflinger B, Schaik A, et al. Limits to fault-tolerance of a feedforward neural network with learning//*Digest of Papers of the International Symposium on Fault-Tolerant Computing*, Los Alamitos, 1990: 228-235.
- [19] Segee B E, Carter M J. Fault tolerance of pruned multilayer networks//*Proceedings of the International Joint Conference on Neural Networks*, Los Alamitos, 1991, 2: 447-452.
- [20] Minsky M. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Magazine*, 1991, 12(2): 35-51.
- [21] 刘永红. 神经网络理论的发展与前沿问题. *信息与控制*, 1999, 28(1): 31-46.
- [22] 周志华, 陈世福. 神经网络国际研究动向. *模式识别与人工智能*, 2000, 13(4): 415-418.
- [23] 朱大奇. 人工神经网络研究现状及其展望. *江南大学学报(自然科学版)*, 2004, 3(1): 103-110.
- [24] Burrascano P, Fiori S, Mongiardo M. A review of artificial neural networks applications in microwave computer-aided design. *International Journal of RF and Microwave Computer-Aided Engineering*, 1999, 9(3): 158-174.
- [25] Mishra R K. An overview of neural network methods in computational electromagnetics. *International Journal of RF and Microwave Computer-Aided Engineering*, 2002, 12(1): 98-108.
- [26] Zhang Q J, Gupta K C. *Neural Networks for RF and Microwave Design*. Norwood: Artech House, 2000.
- [27] Christodoulou C, Georgiopoulos M. *Applications of Neural Networks in Electromagnetics*. Norwood: Artech House, 2001.
- [28] 王秉中. *计算电磁学*. 北京: 科学出版社, 2002.
- [29] 田雨波, 钱鉴. *计算智能与计算电磁学*. 北京: 科学出版社, 2008.
- [30] Gupta K C. Emerging trends in millimeter-wave CAD. *IEEE Trans. on Microwave Theory and Technology*, 1998, 46(6): 747-755.
- [31] Zhang Q J, Gupta K C, Devabhaktuni V K. Artificial neural networks for RF and microwave design from theory to practice. *IEEE Trans. on Microwave Theory and Technology*, 2003, 51(4): 1339-1349.
- [32] Devabhaktuni V K, Yagoub M, Zhang Q J. A robust algorithm for automatic development of neural network models for microwave applications. *IEEE Trans. on Microwave Theory and Technology*, 2001, 49(12): 2282-2291.

-
- [33] Wang F, Devabhaktuni V K, Zhang Q J, et al. Neural network structures and training algorithms for microwave applications. *International Journal of RF and Microwave Computer-Aided Engineering*, 1999, 9(3): 216-240.
- [34] Murphy E K. Radial-basis-function neural network optimization of microwave systems[Master Thesis]. Worcester: Worcester Polytechnic Institute, 2002.
- [35] 刘钊. 微波神经网络技术研究[博士学位论文]. 天津: 天津大学, 2004.
- [36] 洪劲松. 新型传输线方程及知识人工神经网络模型的研究[博士学位论文]. 成都: 电子科技大学, 2005.
- [37] 刘洋. 基于神经网络的微波电磁场计算问题的研究[硕士学位论文]. 大连: 大连理工大学, 2006.
- [38] Mydur R, Michalski K A. A neural-network approach to the electromagnetic imaging of elliptic conducting cylinders. *Microwave and Optical Technology Letters*, 2001, 28 (5): 303-306.
- [39] Rekanos I T. Neural-network-based inverse-scattering technique for online microwave medical imaging. *IEEE Trans.on Magnetics*, 2002, 38 (2): 1061-1064.
- [40] Wang Y M, Gong X. A neural network approach to microwave imaging. *International Journal of Imaging Systems and Technology*, 2000, 11 (3): 159-163.
- [41] Xu J, Yagoub M, Ding R, et al. Neural-based dynamic modeling of nonlinear microwave circuits. *IEEE Trans.on Microwave Theory and Technology*, 2002, 50(12): 2769-2780.
- [42] Horng T, Wang C, Alexopoulos N G. Microstrip circuit design using neural networks. *IEEE Trans.on Microwave Theory and Technology*, 1997, 45(5): 794-802.
- [43] Creech G L, Paul B J, Lesniak C D, et al. Artificial neural networks for fast and accurate EM-CAD of microwave circuits. *IEEE Trans.on Microwave Theory and Technology*, 1997, 45(5): 794-802.
- [44] Fang Y, Yagoub M, Zhang Q J. A new macromodeling approach for nonlinear microwave circuits based on recurrent neural networks. *IEEE Trans. on Microwave Theory and Technology*, 2000, 48(12): 2335-2344.
- [45] Vai M, Prasad S. Microwave circuit analysis and design by a massively distributed computing network. *IEEE Trans.on Microwave Theory and Technology*, 1995, 43(5): 1087-1094.
- [46] Bila S, Harkouss Y, Ibrahim M, et al. An accurate wavelet neural-network-based model for electromagnetic optimization of microwave circuits. *International Journal of RF and Microwave Computer-Aided Engineering*, 1999, 9(3): 297-306.
- [47] Watson P M, Gupta K C, Mahajan R L. Development of knowledge based artificial neural network models for microwave components. *IEEE MTT-S International Microwave Symposium Digest*, 1998: 9-12.
- [48] Watson P M, Gupta K C. Design and optimization of CPW circuits using EM-ANN models for CPW components. *IEEE Trans.on Microwave Theory and Technology*, 1997, 45(12): 2515-2523.
- [49] Jargon J A, Gupta K C, DeGroot D C. Application of artificial neural networks to RF and microwave measurements. *International Journal of RF and Microwave Computer-Aided Engineering*, 2002, 12: 3-24.
- [50] 田雨波, 殷毅敏, 钱鉴, 等. 基于多层感知器神经网络的波导匹配负载设计. *电波科学学报*, 2004, 19 (2): 143-147.
- [51] Christodoulou C, Huang J, Georgiopoulos M, et al. Design of grating and frequency selective surfaces using fuzzy ARTMAP neural networks. *Journal of Electromagnetic Waves and Applications*, 1995,

- 9(1,2): 17-36.
- [52] Huang J. Theoretical analysis of ART neural networks and their applications in frequency selective surfaces[PhD Dissertation].Orlando; University of Central Florida, 1994.
- [53] 张小秋. 神经网络泛化能力研究及其电磁应用[硕士学位论文].镇江:江苏科技大学,2008.
- [54] Passos M G, Silva P H. Applications of modular RBF/MLP neural networks in the modeling of microstrip photonic bandgap structures. *PIERS*, 2007, 3(5): 695-700.
- [55] Guney K, Sarikaya N. A hybrid method based on combining artificial neural network and fuzzy inference system for simultaneous computation of resonant frequencies of rectangular, circular, and triangular microstrip antennas. *IEEE Trans.on Antennas and Propagation*, 2007, 55(3): 659-668.
- [56] Shieh C S, Lin C T. Direction of arrival estimation based on phase differences using neural fuzzy network. *IEEE Trans.on Antennas and Propagation*, 2000, 48(7): 1115-1124.
- [57] Chuang K H, Chiu M J, Lin C C, et al. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy c-means. *IEEE Trans.on Medical Imaging*, 1999, 18(12): 1117-1128.
- [58] Rodriguez N, Cubillos C. Wavelet network with hybrid algorithm to linearize high power amplifiers. *Lecture Notes in Computer Science*, 2007; 1016-1023.
- [59] Mirafatab V, Mansour R R. Computer-aided tuning of microwave filters using fuzzy logic.*IEEE Trans. on Microwave Theory and Technology*, 2002, 50(12): 2781-2788.
- [60] Kabir H, Wang Y, Yu M, et al. Neural network inverse modeling and applications to microwave filter design. *IEEE Trans.on Microwave Theory and Technology*, 2008, 56(4): 867-879.

第 2 章 基础知识

神经网络是指用计算机仿真人脑的结构,用许多小的处理单元仿真生物的神经元,用算法实现人脑的识别、记忆、思考过程。本章主要讲述神经网络的基本知识,包括神经网络模型、神经网络的训练和学习、神经网络的泛化能力及神经网络训练用样本等。

2.1 神经网络模型

研究者从神经生理学和心理学角度出发研究脑的某一部分结构,构造并研究神经网络,称之为生物神经网络;当研究者从工程的角度,应用适当的算法把任务作为一种数学问题来构造合适的神经网络,称之为人工神经网络。这里主要讨论人工神经网络。为更好地理解 and 讨论人工神经网络,本书首先简单介绍生物神经网络中的生物神经元。

2.1.1 生物神经元模型

人的大脑由多达 10^{11} 个不同种类的神经元(神经细胞)组成,神经元的主要功能是传输信息。信息在一个神经元上是用电脉冲的形式传输的,这种电脉冲称为动作电位。一个神经元产生动作电位的最大次数约等于 500 次/s。

生物神经元由细胞体、树突和轴突三部分组成,其结构如图 2.1 所示。树突是细胞的输入端,轴突是细胞的输出端。树突通过连接其他细胞体的“突触”接受周围细胞由轴突的神经末梢传出的神经冲动;轴突的端部有众多神经末梢作为神经信号的输出端子,用于传出神经冲动。生物神经元具有兴奋与抑制两种状态,当传入的神经冲动使细胞膜电位升高到阈值(约为 40mV)时,细胞进入兴奋状态,产生神经冲动,由轴突输出;若传入的神经冲动使细胞膜电位低于阈值时,则细胞进入抑制状态,没有神经冲动输出。

神经元作为信息处理的基本单元,具有如下重要功能。

(1) 可塑性。可塑性反映在新突触的产生和现有神经突触的调整上,可塑性使神经网络能够适应新的环境。

(2) 时空整合功能。时间整合功能表现在不同时间、同一突触上,空间整合功能表现在同一时间、不同突触上。

(3) 兴奋与抑制状态。传入神经冲动时,使细胞膜电位升高,超过被称为动

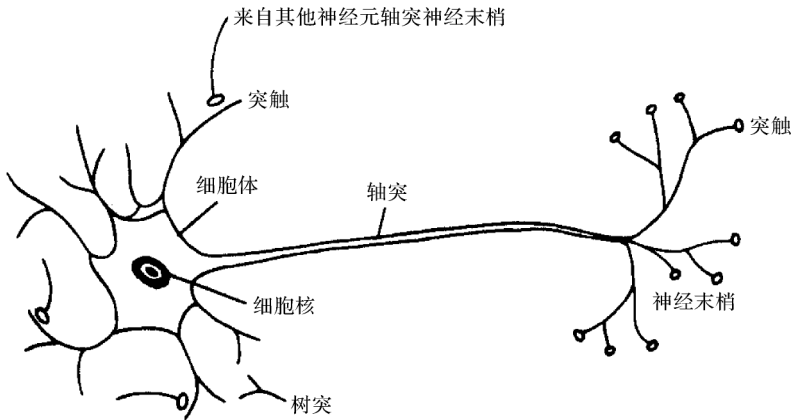


图 2.1 生物神经元

作电位的阈值,细胞进入兴奋状态,产生神经冲动,由轴突输出。当细胞膜电位低于阈值时,无神经冲动输出,细胞进入抑制状态。

(4) 脉冲与电位转换。沿神经纤维传递的电脉冲为等幅的离散脉冲信号,而细胞电位变化为连续信号,在突触接口处进行“D/A”转换。神经元中的轴突非常长和窄,具有电阻高、电压大的特性,因此,轴突可以建模成阻容传播电路。

(5) 突触的延时和不应期。突触对神经冲动的传递具有延时和不应期,在相邻的两次冲动之间需要一个时间间隔,在此期间对激励不响应,不能传递神经冲动。

(6) 学习、遗忘和疲劳。突触的传递有学习、遗忘和疲劳过程。

2.1.2 人工神经元模型

2.1.2.1 MP 模型

人工神经网络的第一个数学模型是由 McCulloch 和 Pitts 建立的 MP 模型^[1],该模型是基于这样一种思想:神经细胞的工作方式或者是兴奋或者是抑制。基于这个思想,McCulloch 和 Pitts 在神经元模型中引入了硬极限函数,该函数形式后来被其他神经网络(多层感知器、离散 Hopfield 神经网络)采用。

由于神经元之间的信号连接强度取决于突触状态,因此,在 MP 模型中,神经元的每个突触的活动强度用一个固定的实数即权值模拟。于是,每个神经元模型都可以从数十个甚至数百个其他神经元接收信息,产生神经兴奋和冲动;同时,在其他条件不变的情况下,不论何种刺激,只要达到阈值以上,就能产生一个动作电位。但如果输入总和低于阈值,则不能引起任何可见的反应。

图 2.2 所示为 MP 模型示意图。图中, x_1, x_2, \dots, x_n 为神经元的输入; $\omega_1, \omega_2, \dots, \omega_n$ 为相应的连接权值; T 为神经元的兴奋阈值; y 为神经元的输出。神经元的输出取二值函数,即

$$y = \begin{cases} 1, & \sum_{i=1}^n \omega_i x_i \geq T \\ 0, & \sum_{i=1}^n \omega_i x_i < T \end{cases} \quad (2.1)$$

式中, x_i 表示神经元的第 i 个输入; ω_i 则表示神经元的第 i 个输入权值; y 表示神经元的输出; T 表示神经元的阈值; n 为输入个数。

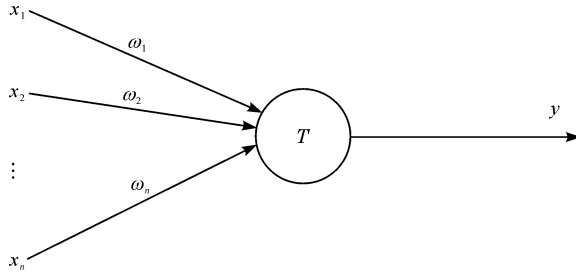


图 2.2 MP 模型

单个 MP 神经元模型可以实现与、或、与非、或非等二值逻辑运算(但不能实现异或运算)。另外,该模型曾说明了人工神经网络可通过简单的计算产生相当复杂的行为而引起极大的轰动,但它是一种静态神经元,即结构固定、权值无法调节,因此,缺乏一个关键性的要素,即学习能力。

2.1.2.2 一般神经元模型

由于 MP 模型过于简单,而且权值不能学习,因此,需要更复杂的灵活性更高的神经元模型。图 2.3 所示为一个具有 n 个输入的通用的神经元模型。与 MP 模型一样, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ 为神经元输入, $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^T$ 为可调的输入权值, θ 为偏移信号,用于建模神经元的兴奋阈值, $u(\cdot)$ 和 $f(\cdot)$ 分别表示神经元的基函数和激活函数。基函数 $u(\cdot)$ 是一个多输入单输出函数 $u = u(\mathbf{x}, \boldsymbol{\omega}, \theta)$, 激活函数 $f(\cdot)$ 的一般作用是对基函数输出 u 进行“挤压”: $y = f(\cdot)$, 即通过非线性函数 $f(\cdot)$ 将 u 变换到指定范围内。

下面介绍常用的基函数及激活函数的类型。

1) 基函数类型

(1) 线性函数。

绝大多数神经网络都采用这种基函数形式,如多层感知器、Hopfield 神经网络

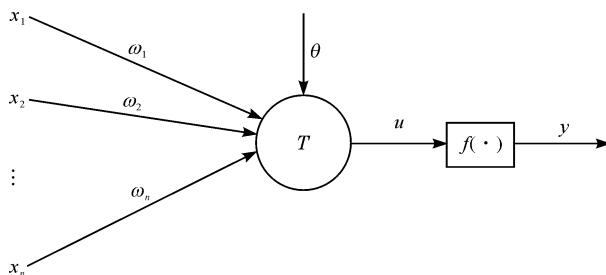


图 2.3 通用神经元模型

络等。采用线性函数时,基函数输出 u 为输入和阈值的加权和,即

$$u = \sum_{i=1}^n \omega_i x_i - \theta = \mathbf{x}^T \boldsymbol{\omega} - \theta \quad (2.2)$$

在多维空间中,该基函数形状是一个超平面。

(2) 距离函数。

此基函数的输出为

$$u = \sqrt{\sum_{i=1}^n (x_i - \omega_i)^2} = \|\mathbf{x} - \boldsymbol{\omega}\| \quad (2.3)$$

式中, $\boldsymbol{\omega}$ 常被称为基函数的中心。显然, u 表述输入矢量 \mathbf{x} 和权矢量 $\boldsymbol{\omega}$ 之间的欧氏距离。在多维空间中,该基函数形状是一个以 $\boldsymbol{\omega}$ 为球心的超球。RBF 主要用于 RBF 神经网络。

(3) 椭圆基函数。

此基函数的输出为

$$u = \sqrt{\sum_{i=1}^n c_i (x_i - \omega_i)^2} \quad (2.4)$$

在多维空间中,该基函数形状是一个椭球。

2) 激活函数类型

激活函数也称为神经元函数、挤压函数或活化函数,它是人工神经元中的一个重要概念,类似于生物神经元具有的非线性转移特性,其基本作用如下:控制输入对输出的激活作用;对输入、输出进行函数转换;将可能无限域的输入变换成指定的有限范围内的输出。

激活函数可以是线性的,也可以是非线性的。常用的激活函数有以下一些类型。

(1) 硬极限函数。

硬极限函数的表达式如下:

$$y = f(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0 \end{cases} \quad (2.5)$$

或

$$y = f(u) = \text{sgn}(u) = \begin{cases} 1, & u \geq 0 \\ -1, & u < 0 \end{cases} \quad (2.6)$$

式中, $\text{sgn}(\cdot)$ 为符号函数。式(2.5)的硬极限函数也叫单极限函数, 式(2.6)的硬极限函数也叫双极限函数。硬极限函数的曲线如图 2.4 所示。

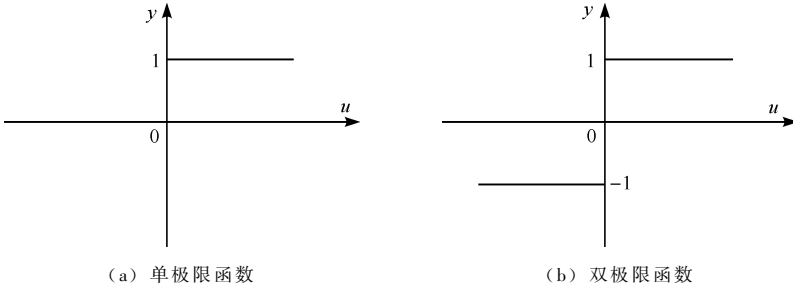


图 2.4 硬极限函数

(2) 线性函数。

线性函数的表达式如下:

$$y = f(u) = u \quad (2.7)$$

线性函数的曲线如图 2.5 所示, 该激活函数常用于实现函数逼近的神经网络的输出层神经元。

(3) 饱和线性函数。

饱和线性函数的表达式如下:

$$y = f(u) = \frac{1}{2}(|u+1| - |u-1|) \quad (2.8)$$

饱和线性函数的曲线如图 2.6 所示, 该激活函数也常用于分类问题。

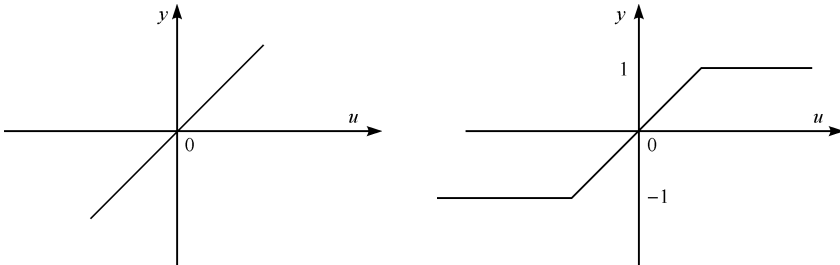


图 2.5 线性函数

图 2.6 饱和线性函数

(4) Sigmoid 函数。

Sigmoid 函数也叫 S 形函数,是目前构造人工神经网络中最常用、最重要的激活函数,无论神经网络用于分类、函数逼近或优化,都可以采用该函数。S 形函数是一个严格单增的光滑函数,并具有渐进特性。对数正切函数是 S 形函数中的一种形式,其函数形式为

$$y = f(u) = \frac{1}{1 + e^{-\lambda u}} \quad (2.9)$$

式中,参数 λ 称为 Sigmoid 函数的增益,是 S 形函数的斜率参数,通过改变此参数可以得到不同斜率的 S 形函数, λ 值越大,曲线越陡。对数正切函数也叫单极性 Sigmoid 函数,该函数是可微的,它的取值在 0 到 1 的范围内连续变化。图 2.7 给出了当 $\lambda=1$ 时对数正切函数的函数图形。

方程(2.9)定义的 S 形激活函数的值域为 0 到 1。有时,激活函数的值域需要从 $-1 \sim 1$ 变化,并且关于原点奇对称。为此,可以采用双曲正切 S 形激活函数(也叫双极性 Sigmoid 函数)。

$$y = f(u) = \tanh(\lambda u) = \frac{e^{\lambda u} - e^{-\lambda u}}{e^{\lambda u} + e^{-\lambda u}} \quad (2.10)$$

同样, λ 为该函数的斜率参数,通过改变此参数可以得到不同斜率的 S 形函数。图 2.8 给出了当 $\lambda=1$ 时双曲正切函数的函数图形。

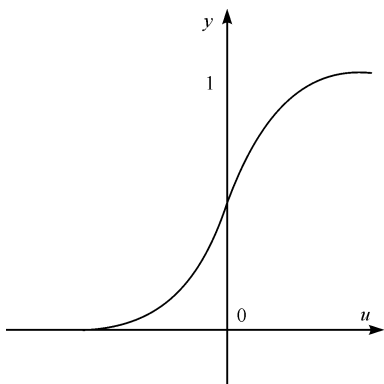


图 2.7 单极性 Sigmoid 函数

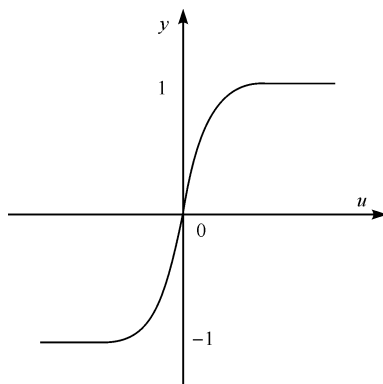


图 2.8 双极性 Sigmoid 函数

(5) 高斯函数。

高斯函数又称为钟形函数,也是一类重要的激活函数,常用于 RBF 神经网络,其表达式为

$$y = f(u) = \exp\left[-\frac{u^2}{\delta}\right] \quad (2.11)$$

式中,参数 δ 称为高斯函数的宽度或扩展常数,其值越大,函数曲线就越平坦,其值

越小,函数曲线就越陡峭。图 2.9 给出了当 $\delta=1$ 时高斯函数的函数图形。

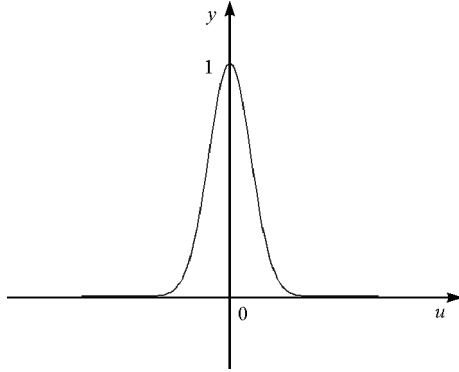


图 2.9 高斯函数

2.2 神经网络的训练和学习

生物之所以能适应环境,是因为生物神经系统具有从周围环境进行学习的能力。对于人工神经网络,学习能力也是其最为重要的特点。任何一个人工神经网络模型要实现某种功能的操作,就必须先对它进行训练,即让它学会它要做的事情,并把这些知识记忆(存储)在网络的权值中。所以,学习或训练的实质就是权值矩阵随外部激励作自适应变化。用数学式表示为

$$\frac{d\omega}{dt} \neq 0 \quad (2.12)$$

正是因为学习或训练的实质是变动加权,所以,学习和训练可以混用。但严格地说,训练通常是指调节网络权值的操作动作和过程,这个过程对网络来讲就是学习。网络被训练后,它每加入一组输入就会产生一组要求的输出。训练就是相继加入输入向量,并按照预定规则调节网络权值。在训练过程中,网络的各权值都收敛到一确定值,以便每个输入向量都会产生一个要求的输出向量。调节权值所遵循的预定规则就是训练算法。在人工神经网络中,一般有两类训练算法,一类是有指导的训练,一类是无指导的训练。

有指导学习也称有监督学习(supervised learning)。对有指导的训练算法,不但需要训练用的输入向量,还要求与之对应的表示所需输出的目标向量。输入向量与对应的目标向量称做一个训练对,即 (p_i, d_i) , $i=1, 2, \dots, N$, 其中, p_i 为样本输入, d_i 为样本输出(教师信号)。通常,训练一个网络需要很多训练对,这些训练对组成训练组。当加上一个输入向量时,要计算网络的实际输出,并同相应的目标向量比较,比较结果的误差用来按规定的算法改变加权。这样,把训练组中的