

人工神经网络与 盲信号处理

Artificial Neural Network and
Blind Signal Processing

杨行峻 郑君里 编著
Xingjun Yang Junli Zheng



清华大学出版社

77183
11

人工神经网络与 盲信号处理

Artificial Neural Network and
Blind Signal Processing

杨行峻 郑君里 编著
Xingjun Yang Junli Zheng

清华大学出版社

(京)新登字 158 号

内 容 简 介

21 世纪科技界面临的重大挑战之一是在不断扩大的广度和深度上用计算机来替代并且更好地完成人的各种智能工作。以人工神经网络、模糊逻辑与进化计算为核心的计算智能学正是为完成这一使命而构建的新学科之一。本书以人工神经网络为重点,其中包含多层前向神经网络、Hopfield 神经网络、自组织神经网络以及一个新的重要发展方向——盲信号处理;还包括模糊神经网络和遗传算法及其在人工神经网络中的应用等交叉学科内容。书中除了给出本学科主要的理论基础外还广泛介绍了这一领域最新研究方向、研究方法和成果、大量应用实例以及参考文献。

本书除了供大学研究生作为教材使用外,还可作为科技、经济、军事、生物和医学等各界人员从事研究和完成各项实际工作时自学和参考用书。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

人工神经网络与盲信号处理/杨行峻,郑君里编著. —北京:清华大学出版社,2002

ISBN 7-302-05880-6

I. 人… II. ①杨…②郑… III. ①人工神经元网络②信号处理—应用—人工神经元网络 IV. TP183

中国版本图书馆 CIP 数据核字(2002)第 071710 号

出 版 者: 清华大学出版社(北京清华大学学研大厦,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

责任编辑: 邹开颜

印 刷 者: 北京牛山世兴印刷厂

发 行 者: 新华书店总店北京发行所

开 本: 787×1092 1/16 印张: 25.5 字数: 587 千字

版 次: 2003 年 1 月第 1 版 2003 年 1 月第 1 次印刷

书 号: ISBN 7-302-05880-6/TP·3487

印 数: 0001~3000

定 价: 39.00 元

前 言

人工神经网络以及与其密切相关的模糊推理(模糊逻辑、模糊集)、进化算法(本书主要讨论其中的遗传算法)和盲信号处理等构成了一门新学科——计算智能学——的核心内容。这门新学科的主要目标是实现用计算机来替代并且更好地完成人的各种智能工作。这是 21 世纪科技界面临的重大挑战之一,它不但对于信息科学技术而且对于经济、军事、工业生产和生物医学等领域将产生非常深远的影响。本书第 1 章为绪论,其余章节以人工神经网络为重点,分 3 个部分来介绍这一课题。第 1 部分从第 2 章至第 4 章,包括 3 种主要的人工神经网络:前向多层神经网络(含递归神经网络)、自组织神经网络和 Hopfield 神经网络。其中第一种研究得最为广泛、深入且应用面最宽,迄今仍是最重要的发展方向之一。第二种的学习算法与构成思路和第一种不同,由于它在模式识别、聚类和数据采掘等领域中表现出独特的优势,近年来发展很快,受重视程度日益提高。第三种是最早提出并促进人工神经网络快速发展的主要几种网络之一,其主要用途是优化和联想记忆。虽然其发展过程有一些曲折,但是近年来它的一些不足之处已全部或部分解决,目前仍然是一个重要的研究方向。第 2 部分为第 5 章和第 6 章,包括模糊神经网络和遗传算法在人工神经网络中的应用两项内容。人工神经网络的优点是具有强学习能力而缺点是不容易纳入人的推理知识,模糊推理系统的优缺点正好与其相反。二者相结合构成的模糊神经网络可以取长补短,因而有很大优势,近年来发展很快,在信号处理、控制、机器人等领域中起重要作用。遗传算法是一种并行逐代优化算法,在解决很多复杂优化问题时表现出独特的优势。而人工神经网络的参数和结构学习正是一个复杂优化问题,所以二者的结合将使人工神经网络学习效率有很大提高。第 3 部分为第 7 章,是盲信号处理,其中主要包含盲解卷和盲分离两大部分内容。所谓盲是指对于被处理的信号没有或只有很少先验知识的条件下,实现多个相加混合信号的分离或卷积信号的解卷。在军事、生物医学、声学 and 地球物理等许多领域中都需要解决这类问题。这一课题是人工神经网络和统计信号处理两大学科相互结合的产物,近年来异军突起且成果斐然,已成为人工神经网络研究的一个新重点。

本书的写作可以追溯到 1989 年春季我们为清华大学电子工程等系博士生、硕士生开设的“人工神经网络”课程。1992 年在授课的基础上并结合我们自己的研究工作编写了一本教材——《人工神经网络》(北京:高等教育出版社,1992.9)。该书的出版受到了国内、外读者的欢迎。十年来这一领域的研究与应用有了极大的进展而且产生了许多新的研究方向。这样,编写一本新书显得十分必要,国内很多同行及各界涉及这一领域的人士也非常关心这项工作的进展。从 1998 年动笔,历经四年完成了本书的写作。在写作过程中,我们想提到两位先贤的学术思想对于我们的影响。一位是曾在我校长期执教的陈寅恪教授(1890—1969)。他毕生提倡学术研究的“独立之精神、自由之思想”,并且身体

力行,已成为近百年来中国学术思想界的楷模。另一位是英国科学哲学家卡尔·波普尔(Karl Popper,1902—1994),他的许多思想曾对人工神经网络的研究产生过深刻的影响。他的论述也对我们深有启发,特别是在科学研究和创新的过程中运用批评的精神、通过“证伪”引起科学进步的思想以及独立运用理智的必要性等,都值得我们反复思考。

关于本书,其初衷是作为一本博士生和硕士生的教材,但从完稿的篇幅和内容看已大大超出了这个目标,即使一门64学时的课程也很难将本书的全部内容讲完。因此在授课时只能讲授关键和重点部分,而较深入部分可作为自学或研究参考使用。本书的写作着眼于方便读者自学,即每一章的内容都是自给自足的,既包含每一研究课题的背景、发展历史、基本假设条件、研究思路、重要算法、重要结论、工程应用以及未来发展方向等,又包含有关各重要定理和算法的证明和推导。对于不给出证明的部分,则给出相应的参考文献。本书列出了五百余篇参考文献,虽然远不全面,但是关键文献以及关于未来发展方向论述的文献大体具备。

本书写作的分工是:

第1、2、3、4、5、7章由杨行峻撰写。第6章由郑君里撰写,陈文霞博士协助完成了第6章的部分编写工作。

本书的完成和出版得到了多方面的鼓励和帮助。这里要特别提出的是清华大学“985教材基金”对本书编写给予的资助。对此我们表示由衷的感谢。最后,对长期关心本书出版的各界同行深致谢忱,并恳请读者对书中不足之处批评指正。

作者

2002年8月于清华园

目 录

第 1 章 绪论	1
1.1 人脑与人的智能、人工神经网络与智能机器、人工智能与计算智能	1
1.2 ANN 的主要模型与研究途径	6
1.3 ANN 的学习	8
1.4 ANN 的应用	11
1.5 脉冲耦合神经网络和脑成像技术	14
1.6 ANN 与盲信号处理	18
1.7 本书的组织	19
参考文献	21
第 2 章 前向多层神经网络与递归神经网络	23
2.1 概述	23
2.2 线性函数 MLP	28
2.3 硬限幅函数 MLP	29
2.4 用 Sigmoid 函数和线性函数的 MLP	31
2.5 BP 算法的主要问题及其改进	35
2.6 RBF 网络	41
2.7 小波神经网络	44
2.8 MLFN 的前端信号处理	48
2.9 MLFN 的函数逼近能力	53
2.10 MLFN 推广能力的统计学习理论	55
2.11 提高 MLFN 推广能力的实用方法	60
2.12 MLFN 作为后验概率估值器	68
2.13 递归神经网络	69
2.14 MLFN 应用举例之一——在非线性动力系统中的应用	80
2.15 MLFN 应用举例之二——手写数字识别	90
2.16 MLFN 应用举例之三——语音识别	95
参考文献	105
第 3 章 自组织神经网络——SOM 和 ART	113
3.1 概述	113
3.2 SOM 的结构和自组织学习算法	114

3.3	SOM 自组织学习算法中的参数自适应	119
3.4	学习向量量化——LVQ 和 LVQ2	126
3.5	SOM 的应用举例	129
3.6	ART 的基本原理和算法实现框架	137
3.7	ART 的各种变型及其比较	140
3.8	ART 应用举例——在化学工业中的应用	147
	参考文献	149
第 4 章	Hopfield 神经网络	153
4.1	概述	153
4.2	HM 及其在 TSP 中的应用	156
4.3	采用约束平面及 HC 的 EHM	166
4.4	采用罚函数的 EHM	183
4.5	离散时间 HNN 与自联想记忆	201
4.6	双向联想记忆及其他联想记忆神经网络	220
	参考文献	228
第 5 章	模糊神经网络	234
5.1	概述	234
5.2	FNN 的结构和类型	235
5.3	实现函数映射的 FNN	239
5.4	基于 Fuzzy ART 的 FNN 和 Simpson 网络	255
5.5	实现聚类的 FNN	264
5.6	FNN 在非线性动力系统辨识与控制中的应用	272
5.7	FNN 用于时间序列预测及其在金融和财务等领域中的应用	288
	参考文献	297
第 6 章	遗传算法及其在人工神经网络中的应用	301
6.1	概述	301
6.2	基本的遗传算法	302
6.3	模式定理	305
6.4	遗传算法的收敛性能	307
6.5	遗传算法面临的问题及改进算法	311
6.6	遗传算法与人工神经网络的结合	318
	参考文献	324

第 7 章 盲信号处理	327
7.1 概述	327
7.2 源信号 pdf 描述、等价可分性、二阶和高阶相关的去除	331
7.3 ICA 的目标函数	334
7.4 ICA 的学习算法	339
7.5 ICA 解的稳定性	350
7.6 用 ICA 实现源分离时解的精确度	355
7.7 ICA 算法中信号源 pdf 的确定	358
7.8 盲信号抽取	362
7.9 盲解卷与盲均衡	365
7.10 DCA 算法	369
7.11 IFA 算法	378
7.12 ICA 和 IFA 的实际应用和待解决问题	392
参考文献	394
缩略语	397

第 1 章 绪论

1.1 人脑与人的智能、人工神经网络与智能机器、人工智能与计算智能

现代人的大脑重量平均值约为 1400g, 大脑内约含 1000 亿(10^{11})个神经元, 每个神经元与其他神经元之间约有 1000 个连接, 这样, 大脑内约有 10^6 亿(10^{14})个连接。人的智能行为就是由如此高度复杂的组织所产生的。浩瀚的宇宙中, 就复杂度而言, 可能只有包含数千亿颗星球的银河堪与人脑相比拟。因此人的智能行为是如此复杂也就毫不足奇了。

人脑的基本组成单位——神经元(neuron)——是一种特化的细胞, 其示意结构如图 1-1 所示。神经元表面有许多为引进输入信号的短突起, 称为树突(dendrite), 而输出信号的是一根长突起, 称为轴突(axon), 轴突周围包有髓鞘。树突的任何部位都可以与来自其他神经元的轴突末梢建立联系, 构成突触(synapse)。神经元的活动和神经元之间信号的传递是一个极复杂的生物化学——电过程。概言之, 一个神经元只能取两种状态——激发态(兴奋状态)或抑制态, 当神经元从抑制态转变为激发态时, 即沿自身的轴突送出一个脉冲信号, 此信号通过突触送到其他神经元。突触可分为兴奋型和抑制型两种且可能具有不同的强度, 这就是说虽然神经元送出的脉冲信号是单一的, 但随着突触的不同, 接收方收到的是强度各异的兴奋或抑制信号。当一个处于抑制态的神经元接收到的信号总和为兴奋性且超过某一阈值时, 即转变为兴奋态, 且在传送出自身的电信号以后逐渐减退为抑制态, 在一短段时间中此神经元对外来信号不再作出响应(这段时期称为不应期)。在 1.5 节中还要对此机理作阐述。此外, 大脑的工作具有模块化和层次化的特点,

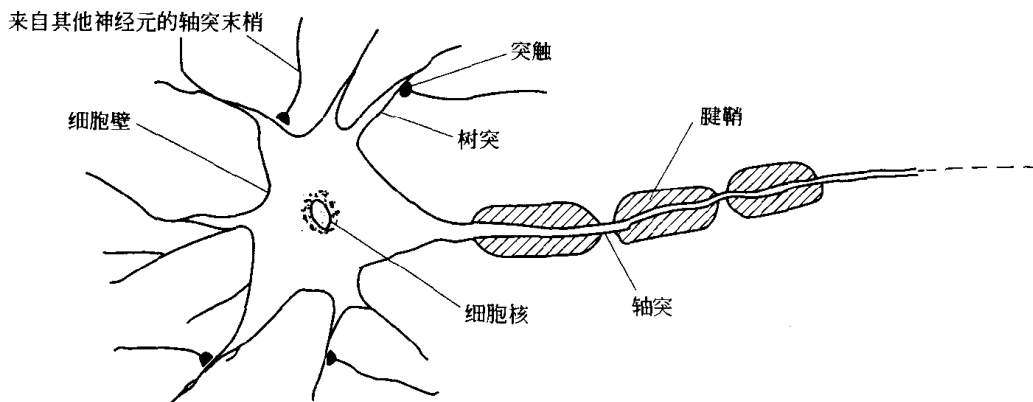


图 1-1 神经元结构示意图

即大脑皮层分成若干功能区,某个区域中互相协同作用的一群神经元组成具有某种特定功能的模块,例如对视觉信号作初步加工——检测出一帧图像中的边缘、线段及其方向等。对理解一帧图像而言,除了这种低层次的信息加工模块外,还有高层和更高层的模块对来自低层模块的信息作进一步加工(模块之间的信息交换是双向的,即既有“由底向上”的又有“由顶向下”的信息传送)。

什么是人的智能呢?对智能作出严格定义相当困难。下面只列出人的智能活动所涉及的一些领域,所列的这些智能行为之间往往有紧密关系,很难严格区分。此外,所列出的项目是极不完备的且每个项目还能分成若干更细的子项目,而且还可能涉及深层次的心理学、哲学和认知科学等问题。这里做的只是粗略介绍。

1. 感知和认知

感知是大脑通过各种感觉器官接受外界的声、光、触、嗅等信息。初级的认知是以感知信息为基础,辨识出不同事物并建立各特定事物及其特殊属性的概念。例如,每一头具体的牛都与另一头不同,但是可以建立关于牛这种动物的概念。这个概念包含世界上所有的牛和它们的共同属性,当然还可能包含更多深层次的信息。更高级的认知涉及各事物之间的关系以及抽象的概念,例如5这个数字可用于5头牛也可用于5个苹果。

2. 记忆

大脑不可能将所有接收到的信息全部存储起来,记忆存储与信息的加工(认知)和选择(集中注意、兴趣等)相关。记忆的另一方面是提取出所存的某一部分信息,这种信息的搜索和提取往往是联想式的。这就是在存储各种信息时必然还需存储它们之间的联系,而不是存储孤立的信息,这样才能进行联想式的记忆项搜索提取。记忆是大脑最重要的功能之一,它是其他智能活动的基础。

3. 学习与知识

学习是大脑最重要的功能之一。学习的作用是获取、存储(记忆)和使用知识。知识是一种结构性的信息,例如因果关系(知因求果或反过来知果求因)、空间关系、时间关系、推理关系以及根据事物的特征进行分类和模式识别等等。无论何种知识,大体上都可以表现为一个称为输入的变向量对一个称为输出的变向量之间的映射关系。例如,前者是因后者是果;前者是一个函数的变量后者是函数值;前者是一个类别或模式的特征向量后者是分类或模式识别的结果,等等。需要说明的是,分类的作用是根据某种事物的特征来判断其属于哪一种具有特定意义的类别,例如当一个人体细胞只有具备了某些特征后就被认为变成了癌细胞,反之则否。从这个意义而言,分类与模式识别是同义的。如果按产品的若干性能指标将其分为一级品、二级品等,则这可以看成纯分类问题。类别划分的另一种作用是,把具有若干相似特征的事物归属于一个类别,再赋予某种标号和名称,其目的不在于同一类型事物中区分出类别,而在于将具相似特征者聚为一类以便于形成概念、进行操作。例如,为了检索和利用,人们通常将内容相近的资料、文档、图书按类区分,分区储放。这一种类别划分纯依其内容相似程度,而前述的模式识别型的分类则取决于其外在特性,例如前文所举的是或不是癌细胞之例(在这种情况下,即使输入特征差异很大还可能归入一类,而输入特征差异不大也可能分为异类)。在科技、工程领域将模式识别(pattern recognition)和分类(classification)都归于这一型问题。而将纯依内容相似程度

进行的类别划分称为聚类(clustering),将一个函数的输入和输出之间的映射称为回归(regression)问题。

无论是人还是智能机器,学习都是一个关键问题,是重中之重。对学习提出的要求是探索一种学习算法,既有很好的推广性能又有很高的效率。在本节和本章乃至全书,将一再讨论这个问题。

4. 语言能力(听和说,读和写)及视觉功能能力(分辨物体——颜色、形状、边缘、纹理等,判断空间关系——距离、方向、深度等,以及边缘及线条检测等)

语言既是一种重要的信息交换工具又是知识表示和思维与推理的重要手段。而视觉信息又占了外界输入信息量的80%以上,所以大脑皮层中,涉及视觉信息处理者占了很大百分比。

5. 行动和动作

行动就是用腿脚将自身从一个地点移动到另一个地点,这涉及避免与各种静止或活动的障碍物相碰撞以及设计一条从起点到终点的路径。动作就是通过臂和手的协作将一物移至另一处或完成某种动作(扳开关、拧螺丝等)。智能机器人应该和人一样做到这一点。

6. 优化

优化就是根据某一准则制订一个目标函数(或称为价值函数),通过改变或选择各种可供选择的方案、参数、结构或计划,使该函数达到极大值(或极小值),即得到最大“收益”(或产生最小“损失”)。优化问题在人的日常生活和社会生活的方方面面(经济、工业技术、政治、军事……)无不经常出现并需要得到尽好的解决。

7. 预测

人在制订自己的行动计划时,无时无刻不用到预测,预测既包含对外界环境未来的估计,也包含对自身任何行为所会产生后果的估计,并且根据这种预测作出选择。在人类的社会生活中,小至一个工作单位大至一国的政府乃至国际组织、金融机构在制订政策和决定行动时都很强地依赖于预测和预报。

8. 计划、判断和决定

人在完成一定任务时总要做计划和做决定,这实际上是根据任务的要求在已有的知识中进行选择和组合。而选择的标准则取决于优化目标是否达到和对计划或决定执行后果的预测,即必须进行判断。

9. 自适应和鲁棒性

人所生活的环境决非一成不变,且有许多不确定因素,因此一个人学得再多也不可能知道自己的未来所将面临的所有课题。这样人必须随着环境的变化不断修正自己的应对策略,根据自身所做行动受到环境的奖或惩来学习新知识。在遇到从未接触的不定因素时能作出尽量好的抉择,即人的智能行为必须有鲁棒性(robust)和自适应性。

10. 博弈与对策

博弈论与对策论的研究对于揭示人的智能行为的特点和如何针对利益不同的对手(利益相反、利益相同或中性)在博局中的行为给出自身相应对策的策略都极富启发性。特别在经济、军事、社会等许多强竞争领域中,更是对智能行为研究的一种好工具。

11. 创新和发明

人的创造能力是智能的一个重要特点,即人可以从学得的知识中通过重新组合、选取和随机变化产生出从未学过的新知识——新的原理、新的机器……

12. 深层的智能因素

如意识、感情、意志、注意力、直觉、理解等。

在粗浅地介绍了人脑和人的智能行为后,现在讨论本书的主题——人工神经网络。众所周知,工业革命通过用各种动力机器(蒸汽机、汽油和柴油机、电动机……)代替人的体力劳动开创了一个新时代,使得社会的物质财富有了极大的增长,同时使人们摆脱了过于繁重的体力劳动。现在我们面临着信息时代,这个时代将通过用各种智能机器替代人的智力劳动,来创造一个精神和物质财富远比工业革命时代丰富且使人们过于繁重的智力劳动能够得到减轻的新时代。这个新时代的标志是计算机、包括互联网和移动通信在内的各种通信工具、信号处理(包括视频、音频、文字等)、自动控制和智能机器人等许多领域中取得的飞速进展。大规模集成电路的高速发展以及光盘存储容量的持续提高为这个时代提供了强有力的硬件支持,各式各样的信息如潮水般涌来。现在对于能够帮助我们处理各种信息的智能机器的需求十分迫切,例如互联网的快速发展及其多媒体业务和信息量的快速扩展对于网络的智能管理、智能接口(例如语音识别、图像和文字识别等)以及数据采掘(data mining, DM)的要求越来越高。再如大规模集成电路等现代高精尖产品的质量提高依赖于智能性质量控制、故障检测等。但是,现在所能提供软件的智能水平仍然非常低下。为了应对迫在眉睫的各种智能信息处理课题,20 世纪后半叶开展了多方面的研究与探索,人工神经网络就是在 20 世纪 80 年代初期涌现出来的。经过 20 年来的发展,其成就已蔚为可观。人工神经网络是从生理、心理和认知等各种不同视角出发,对人脑的结构和运行方式进行各种层次的借鉴,以便在计算机上实现具有人类智能行为特点的各层次功能。

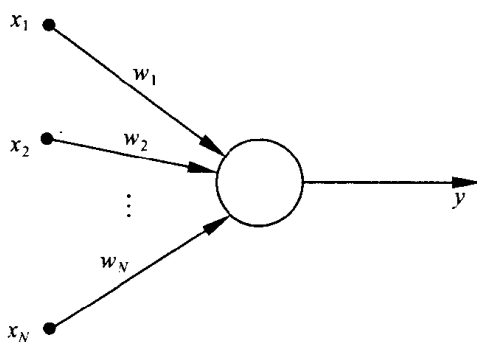


图 1-2 人工神经元

人工神经网络的研究从一开始就借助于神经元的生理结构模型。对于图 1-1 所示神经元的最简单而直接的模仿是图 1-2 所示的人工神经元, $x_1 \sim x_N$ 表示来自其他神经元的输入信号。 $w_1 \sim w_N$ 表示突触强度, y 表示本神经元的输出信号,上述各变量和参数皆取实数值。一种简单的输入-输出映射关系是

$$y = \text{sgn} \left[\sum_{i=1}^N w_i x_i \right] \quad (1-1)$$

其中 $\text{sgn}[\cdot]$ 表示取符号值。这就是 20 世纪 40 年代提出的 McCulloch-Pitts 模型^[1]。下文中

将介绍多种其他模型。基于一种特定的神经元模型,可以将很多个神经元按照不同方案组合成各种类型的人工神经网络。各种类型的网络具有如下的一些共同特点。

(1) 网络由同一种简单的非线性处理单元即神经元构成。

(2) 网络的构成和运行都按照分布-并行的方式。每一个神经元和每一个突触强度都承担着所有记忆内容的存储,而每一个记忆项的提取也涉及每个神经元和突触。这就是说人工神经网络的信息存储是分布的,根本找不到某特定信息的存储区;其运行是并行而非串行的,即每一时刻所有神经元同时运行而不是按时序依次进行运算。

(3) 网络中神经元的数量非常巨大,从而使得网络成为一个复杂系统。

(4) 网络有学习能力。它并非按已经编制好的程序来实现某种所需的功能,而是通过学习来获得此功能。

概言之,可以将人工神经网络描述为复杂的、分布-并行的、有学习能力的非线性动力系统。复杂性的概念在当代信息科学、生物学、经济学、物理学和宇宙科学中起重要作用^[2]。人工神经网络的分布-并行原理使其能用一般速度的处理单元完成极高速的运算,其学习能力和非线性可使之完成过去传统方法无法完成的若干任务。

在智能科学的研究中,与人工神经网络的研究并行且早开始十余年的一项研究是人工智能 AI(artificial intelligence)。AI 采用符号逻辑来构成推理和判断,它利用符号间的语义关系在归纳专家知识的基础上构成“IF...THEN”式的规则库。在解决一个具体智能问题时,采用各种启发式的搜索方案,从知识库(规则库)中寻求解答。其研究重点是如何表示知识,如何进行符号处理,如何进行搜索,如何搜集和归纳专家的知识等。归根到底,其思路是用精确的计算机程序将人类专家的知识归纳和表现出来,计算机用严格安排好的程序按推理方式解决问题,所以人们常称之为专家系统。30 多年来,AI 研究的主要结论之一是:智能需要知识!这无疑十分正确,但是 AI 在解决实际世界的各种问题时只取得了有限的成功,例如在医疗诊断、弈棋等方面。人工神经网络 ANN(artificial neural network)的研究方法及思路与 AI 形成了鲜明对照,特别在于它借鉴人脑解决智能问题时的生理和心理特点以及通过学习来解决问题。

近十年来,包括 ANN 在内的一个新学科出现并显示出强劲的生命力,这就是计算智能 CI(computational intelligence)或称为软计算(soft computation)。它包含以下三个相互独立却又相互紧密结合与渗透的子学科:

(1) ANN;

(2) 模糊推理系统 FIS(fuzzy inference system),或称为模糊集或模糊逻辑;

(3) 进化算法 EA(evolutionary algorithm),其中包含遗传算法 GA(genetic algorithm)、进化计算 EC(evolutionary computation)和进化程序 EP(evolutionary program)等大致相同而略有差异的算法。

其中 FIS 采用模糊集,给予语言一种近似人类思维的数学表示并用模糊逻辑进行推理。FIS 在表示人类知识和认知方面有很强的能力,但是缺乏学习能力。反之,ANN 有强学习能力却很难纳入人类知识。因此,二者结合所构成的模糊神经网络 FNN(fuzzy neural network)具有明显的优势。EA 是一种并行随机优化算法,它是通过借鉴达尔文的进化论学说构思出来的。在解决许多过去经典优化算法难以解决的高难度、复杂优化问题时,EA 已表

现出很多独特的优势。而 ANN 本身最关键的问题,即学习问题,就是一个高难度、复杂的优化问题。将 EA 用于 ANN 的学习,显然是一个十分理想的方案。不但如此,过去 ANN 的学习主要是针对参数的学习,而结构是通过尝试来决定的。采用 EA 则不但能进行参数学习而且可以进行结构学习,甚至可以进行学习规则的学习。后者在某些方面已接近于人类思维中的创造性范畴了。当今 ANN 的研究已与 FIS 和 EA 的研究分不开,这三项研究的学术会议常在 CI 的大题目下联合举行。文献[3]对于该领域的最新进展有 11 篇综述论文予以详细介绍。本书第 5 章专门讨论 FNN,第 6 章讨论 GA 及其在 ANN 学习中的应用。

1.2 ANN 的主要模型与研究途径

在 ANN 的研究中提出了很多模型,其差异主要表现在以下几方面。

1. 所取的研究途径

指借鉴人脑的哪一种生理或心理的运行特点以及其他相关学科(如物理学、信号处理、统计学等)的何种涉及记忆、学习、分类等的算法。

2. 网络中神经元取何种函数

例如,除上文中提过的 $\text{sgn}[\cdot]$ 函数外还可以取 Sigmoid 函数、径向基函数(RBF)以及更复杂的脉冲耦合神经元模型(见 1.5 节)等。

3. 网络的结构

这是指网络中各神经元是如何连接的。在全连接时,每一个神经元与网络中其他神经元之间都有连接。在部分连接时,一个神经元的输入信号只取自网络中一部分神经元。在全连接时,必然存在反馈,即信号存在一个闭合的传输路径。在部分连接时,则可能存在反馈,也可能不存在。

4. 运行方式

这涉及很多方面,例如运行的时间变量可取连续值或离散值。如果要更精确地模仿人脑,则需要采取十分复杂的运行方式(见 1.5 节)。

5. 学习算法

学习算法很多,概括地可分为有监督的(supervised)和无监督的(unsupervised)两种,后者又称为自组织的(self-organized)。这两种算法的差异在于,前者在学习时所赋予的学习样本既含所需完成输入-输出映射的输入,又含输出,而后者只含输入不含输出(输出由学习者通过学习予以确定)。

6. 应用

可以将 ANN 的应用粗划分为函数逼近(回归)、分类与模式识别、聚类、优化等,还可以更具体地进行划分其实际应用。

ANN 的主要模型介绍如下。

1. 多层前向神经网络 MLFN(multilayer feedforward neural network)

这是一种直至目前研究得最多且应用最广的 ANN,它采用多层局部连接结构、无反馈,神经元函数通常取 Sigmoid 函数或 RBF,一般按离散时间运行,采用有监督学习算法。其应用涵盖面很宽:函数逼近、分类与模式识别、系统辨识与控制、后验概率估计、主分量

分析 PCA(principle component analysis)……它也是本书研究的重点之一。

2. 递归神经网络 RNN(recurrent neural network)

这是一种局部连接或全连接、有反馈的网络,其学习、运行、神经元函数选择等方面与 MLFN 类似。主要用途是非线性动力系统的辨识、建模和控制。

3. 自组织神经网络(self-organized neural network)

这种网络的主要特点是学习算法为无监督的自组织算法,其主要功能是实现对输入特征向量的聚类且在此基础上用于完成函数逼近、分类及模式识别等映射。它的用途的一个明显例证是用发现知识的数据采掘。最著名的两种自组织神经网络是自组织映射 SOM(self-organized mapping) 网络 and 自适应谐振理论 ART(adaptive-resonance theory) 网络。前者又称 SOFM(self-organized feature mapping),是芬兰科学家 T. Kohonen 在 20 世纪 80 年代初提出的,目前这种网络在 DM、模式识别和信号处理等领域颇受重视。后者是美国科学家 S. Grossberg 等在 20 世纪 80 年代中期提出的,其特点在于汲取了人脑智能活动的许多心理特点,诸如集中注意、短期与长期记忆、记忆的弹性与刚性、学习与外界奖惩的关系等。

4. Hopfield 神经网络

也可简记为 HNN,这种网络是美国物理学家 J. J. Hopfield 在 20 世纪 80 年代初提出的。这是一种全连接反馈网络,其神经元函数为 $\text{sgn}[\cdot]$,其运行可按连续时间也可按离散时间进行。前者称为连续时间 HNN,主要用于解决各种优化问题。后者称为离散时间 HNN,主要用于联想记忆、信号的增强与恢复。Liapunov 能量函数的概念在这类网络的研究中起重要作用。

5. 模糊神经网络 FNN

FNN 是 ANN 与 FIS 的结合。一种方案是在 MLFN 或 RNN 中纳入 FIS,以使得人的知识能够以 FIS 的形式用于 ANN 的结构设计和参数的粗调整,而 ANN 的学习算法则用于参数细调整。这样就使结构设计和参数学习的效率有很大提高。其应用与 MLFN 及 RNN 相似,特别适用于非线性动力系统的辨识和控制。另一种方案是 FIS 与 SOM 相结合构成模糊聚类系统,与确定聚类系统相比,这种系统对聚类的形成更符合人的认知行为,因此在 DM 这一类应用中性能明显优越。

6. 脉冲耦合神经网络 PCNN(pulse coupled NN)

这种网络取更加精确的、更符合实测的生物脑神经元生理-电活动规律的人工神经元模型,从而在神经生理学智能模仿方面向前跨了一大步(见 1.5 节)。目前它主要用于视觉信号的初级处理,如图像分割等,已颇见成效。

7. 波尔兹曼机

这是一种全连接、反馈式 ANN,除了其运行规律是随机式的以外,这种网络的结构及用途都与 HNN 类似。其特点是神经元的输出以概率方式依从于输入且采取模拟退火方案使网络的运行从开始到结束的过程中随机性由大渐次变小,最终变成确定网络。这种网络与人脑的某些活动特点如睡眠、反学习、做梦等有些关连^[4]。

以上介绍的模型是不全面的。例如细胞神经网络(cellular NN),它主要从电路理论的角度出发讨论 ANN 的构成,本书就割爱了。再如 LPN(learning petri network)^[5],它

借鉴最近的脑科学知识以及大脑具有不同功能区的特点来构成 ANN。由于这一研究刚起步,故本书也不做讨论。这样的情况很多,就不一一列举了。

1.3 ANN 的学习

学习问题对于 ANN 是一个大题目。学习目的和学习算法各式各样,为了节省篇幅,这里只讨论完成输入特征向量至输出变量之间映射的学习。这实际上已涵括了人脑智能系统和在实际应用中要求 ANN 所完成任务的相当大一部分,诸如函数逼近、分类和模式识别以及聚类等。有关学习的问题正如有关知识的问题一样,方面很多,层次很多,不可能尽述。这里只择要予以介绍。

1.3.1 三类学习

前文已指出学习可粗分为有监督和无监督两大类。对于前者,在学习开始前,要提供若干对已知输入向量和相应输出变量构成的样本集(或称为训练集),可以认为此集由教师提供,所以称为有教师指导下的学习。其用途为函数逼近、分类和模式识别等。对于后者,学习前只提供若干已知输入向量构成的训练集。其用途为聚类。如做细分,有监督学习还可以分成普通有监督学习和增强学习(reinforcement learning)两类。二者的区别在于,对前者的输入-输出映射要求在训练集中规定得十分明确,而对于后者输入-输出映射的输出部分很难规定得十分明确,而只能根据总的待实现目标判断输出的某种变化具有正面效果还是负面效果,而不能给定明确输出值。例如弈棋者以当前棋局为输入向量,下一着棋的弈法为输出,待实现目标为赢棋,即属此种情况。

1.3.2 Hebb 学习律

D. O. Hebb 根据生物神经元的工作特点提出了他的著名学习律^[6]。这可以描述如下。设网络中有编号为 i 和 j 的两个神经元,它们的输出分别记为 x_i 和 x_j ,它们之间的突触强度(连接权)记为 w_{ij} 。再设有编号为 $m = 1, \dots, M$ 的 M 个训练样本,对于每个样本神经元 i, j 的相应输出记为 $x_i^{(m)}, x_j^{(m)}$,那么 w_{ij} 可用下式计算:

$$w_{ij} = \frac{1}{M} \sum_{m=1}^M x_i^{(m)} x_j^{(m)} \quad (1-2)$$

如果 x_i 或 x_j 只能取值 1(兴奋)或 -1(抑制),那么可以看到,如果在训练中大部分 $x_i^{(m)}$ 和 $x_j^{(m)}$ 取相同符号,则 w_{ij} 取校正之值,反之则取较负之值。各 w_{ij} 还可取一种自适应的方式予以修正。设按离散时间 t 赋予网络训练样本,在时刻 t 神经元 i, j 输出为 $x_i(t), x_j(t)$,那么 $t+1$ 时刻的 w_{ij} 可由 t 时刻求得如下:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha x_i(t) x_j(t) \quad (1-3)$$

其中 $0 < \alpha \ll 1$ 。与 Hebb 学习律对立的还有反 Hebb 学习律(anti-Hebb learning)。它是在(1-3)式中取 $\alpha < 0$ 或者其和式前加负号。反 Hebb 学习往往起遗忘或打乱作用,有些学习算法中借助其跳出常规的途径。

1.3.3 目标函数与最陡下降算法

学习问题是一个优化问题。在 ANN 的结构已取定的情况下,学习问题归结为求网络中连接各神经元的权 w_{ij} ,使得一个目标函数达到极小值。此目标函数以某种准则衡量 ANN 对训练集中各输入向量的实际输出与理想输出(在训练集中给定)之间的差异(有监督学习情况),或者训练集中各输入向量的聚类误差(自组织学习情况)。目标函数在统计学中称为经验风险函数(empirical risk function),如果 ANN 中所含的各个权用一个向量 ξ 表示,则此函数可表示为 $R_{\text{emp}}(\xi)$ 。一种求最优 ξ 使此函数达到极小的算法是,从一个随机初值 $\xi(0)$ 出发,按节拍 k 进行下列迭代计算:

$$\xi(k+1) = \xi(k) - \alpha \nabla_{\xi} R_{\text{emp}}(\xi) |_{\xi=\xi(k)}, \quad k = 0, 1, 2, \dots \quad (1-4)$$

这就是最陡下降算法,其中 $0 < \alpha \ll 1$,称为步幅。当 α 足够小时,每迭代一步, $R_{\text{emp}}(\cdot)$ 将下降或不变;当 k 足够大时, $R_{\text{emp}}(\cdot)$ 将收敛到一个极小值点。

对于 MLFN 而言,用最陡下降算法求最优 ξ 的尝试在早期研究中受到了挫折。这是因为无法计算 $R_{\text{emp}}(\cdot)$ 相对于各隐含层神经元输入权相应的偏微分,因而不能求得迭代计算中的梯度函数。20 世纪 80 年代前期提出的 BP(back propagation) 算法成功地解决了此问题。这是 ANN 研究蓬勃展开的重要动因之一。最陡下降算法及基于它建立的 BP 算法的优点是非常简单,缺点是因 α 必须非常小(否则将导致振荡而不收敛)而使学习速度极慢,另一个缺点是当目标函数存在多个局部极小值点时,如初值选择不当,会收敛到低质局部极小点。即便如此,BP 算法仍是目前使用最广的学习算法之一。

1.3.4 学习算法的性能

一种学习算法的性能优劣是指由之产生的 ANN 推广(generalization)性能的优劣。这里是指用训练集内数据所确定的 ANN 用于训练集外数据时,其误差若略有增加而差异不大,则推广性能优越;反之,若增加很多则推广性能劣。推广性能至关重要,若 ANN 的推广性能低劣,则其实用价值很低。一个 ANN 推广性能的优劣取决于以下因素。

(1) 待完成的映射任务的复杂度。这主要取决于输入特征向量的维数,维数越高越复杂。也取决于映射函数本身的复杂性。作粗略估计时,就以输入向量的维数估计复杂度。

(2) 训练集的规模。用集中样本的个数来衡量其规模的大小。

(3) ANN 的结构和规模。以 ANN 中所含神经元数以及连接权个数的多寡来衡量。

(4) 学习算法本身。

推广性能与这些因素的关系是,映射任务越复杂或训练集规模越小则推广性能越差;反之,则佳。ANN 的规模对于一定的复杂度和训练集规模而言,有一最佳适中值,过大过小都会削弱推广性能。学习算法本身的作用自然是不言而喻的。

1.3.5 学习算法的时间效率与 ANN 结构的空

一个学习算法的时间效率是指完成参数学习所耗费的计算机时的多寡。如果所耗机时随着输入特征向量维数的增加而按多项式关系增加,即称为高效率的;反之,如按指数