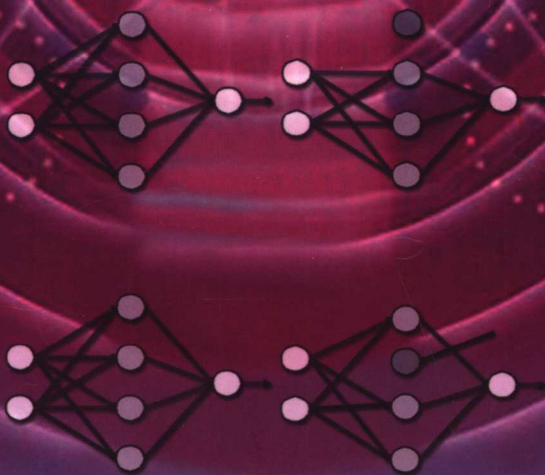


前馈神经网络

工作机理分析 与学习算法

李爱军 著



中国科学技术出版社

前馈神经网络
工作机理分析与学习算法

李爱军 著

中国科学技术出版社

· 北 京 ·

图书在版编目 (CIP) 数据

前馈神经网络工作机理分析与学习算法 / 李爱军著. —北京: 中国科学技术出版社, 2006. 6

ISBN 7-5046-4395-5

I. 前... II. 李... III. 前馈-人工神经网络-研究 IV. TP183

中国版本图书馆 CIP 数据核字 (2006) 第 066622 号

自 2006 年 4 月起本社图书封面均贴有防伪标志, 未贴防伪标志的为盗版图书。

责任编辑: 郑洪炜

封面设计: 李梦思

责任校对: 凌红霞

责任印制: 王 沛

中国科学技术出版社出版

北京市海淀区中关村南大街 16 号 邮政编码: 100081

电话: 010-62103210 传真: 010-62183872

<http://www.kjpbbooks.com.cn>

科学普及出版社发行部发行

北京长宁印刷有限公司印刷

*

开本: 850 毫米×1168 毫米 1/32 印张: 5.875 字数: 160 千字

2006 年 6 月第 1 版 2006 年 6 月第 1 次印刷

印数: 1—200 册 定价: 12.00 元

(凡购买本社的图书, 如有缺页、倒页、
脱页者, 本社发行部负责调换)

摘 要

人工神经网络 (Artificial Neural Networks, ANN) 是一个高复杂度的非线性系统, 虽然从形式上模拟了人脑的学习结构, 但由于所依赖的生物学理论基础尚不完善, 因此人工神经网络不仅功能上远远没有达到预期的接近于人脑学习能力的目标, 而且对于现有神经网络模型的工作机理也不明确, 使神经网络模型的研究和性能的改进也就变得越来越困难, 应用领域也受到一定的影响。

我们的研究以构建更有效的人工神经网络模型为目的, 以神经网络的工作机理分析为基本出发点, 采用非参数化的决策树 (Decision Tree) 与传统人工神经网络结合的方法, 研究神经网络的结构设计方法, 并进一步探讨了人工神经网络的增量学习算法, 主要包括如下几方面的工作。

1. 对前馈网络的工作机理进行分析

本书在给出了分类前馈网络和决策树的等价的形式化证明的基础上, 结合神经元的高维空间几何解释, 对前馈网络模型的工作机理作了合理的分析, 给出了如下的解释: 在分类时, 前馈网络通过学习正反例样本, 对样本空间按照所属类别进行划分。对于一个三层前馈网络, 其输入层起到形成决策界面 (分割) 的作用, 隐层起到形成决策域 (对决策界面 “与”) 的作用, 输出层则是对决策域进行综合 (对决策域 “或”) 的作用。并对影响神经网络泛化能力、增量学习能力的原因和因素作了直观

的分析。

2. 提出了神经网络结构设计的信息论方法

根据对分类前馈网络的分析,指出了前馈网络采用三层结构比较合理。针对离散和连续属性的分类问题,根据三个层次神经元的具体意义,提出了神经网络结构设计的信息论方法:基于决策树的神经网络(Decision Tree-based Neural Network, DTBNN)和基于熵的神经网络(Entropy-based Neural Network, EBNN)设计方法。DTBNN 是一种针对具有离散属性数据的分类问题,在熵网络(Entropy-net)基础上改进的,利用决策树确定神经网络结构、权值和阈值初始值的系统的神经网络设计方法。该方法提高了神经网络的学习速度和学习结果的稳定性,实验结果也表明 DTBNN 的比熵网络的收敛速度快,且识别率也有所增加。EBNN 则是针对具有连续属性数据的分类问题,把熵作为选择神经元的准则来构建神经网络结构的方法。EBNN 神经网络的结构趋于简单、易于理解,并对神经网络某些参数可以进行合理设置;同时,EBNN 采用分层、分类学习机制使神经网络的学习过程得以简化,解决了前馈网络的信任分配问题。EBNN 在相同的神经元数目的情况下其容量要比全连接网络的容量大,因此也比全连接网络有更好的泛化性能,且该方法一定收敛。实验结果也表明 EBNN 的有效性。

3. 基于集成系统的神经网络增量学习

本书采用基于神经网络集成系统的方法,实现前馈网络在有限的存储空间和计算复杂度下使神经网络学习新的知识时尽量保持知识结构的唯一性的增量学习能力。基于集成系统的增量学习方法以主动学习为基础,通过把新知识的样本作为重点

学习的目标, 结合神经网络集成系统的特性, 在对原有知识结构不加任何修改的基础上实现了知识的增量学习。在 LEARN++ 算法基础上, 该算法借助于 Boosting 技术中主动选择样本的过程解决了神经网络的不能学习只有正例样本和小样本问题, 并有效地选择了构成集成系统的个体网络, 使其增量学习的泛化性能比 LEARN++ 有很大提高。基于集成的增量学习算法增加的计算开销很少, 几乎可以忽略; 而所需的存储容量也在一个合理的范围内。基于集成系统的神经网络增量学习算法具有一定的普适性, 也可以用于其他学习模型。

关键词 前馈神经网络 决策树 熵 增量学习 集成

ABSTRACT

Artificial Neural Networks (ANN), a highly-complicated non-linear system, stimulates the learning structure of human brain in form. Because the biological theory that it relies on is not perfect, ANN fails not only to achieve the expected aim of approximating the learning ability of human brains, but also to explain the working mechanism of existing neural networks models. Therefore, the research of neural networks models and the improvement of their performance become more and more difficult, which affects their application.

In this thesis, we aim to construct a more effective artificial neural networks model. With the analysis of the working mechanism of the neural networks as a starting point, we adopt the method of combining the non-parametric Decision Tree and traditional artificial neural networks to investigate the structure design method. Furthermore, we discuss incremental learning algorithm of the neural networks. The following aspects are the details.

1. Analyze the working mechanism of the feedforward neural network..

First, the thesis provides a formalized proof of the equivalency between classified feedforward neural networks and Decision Tree. Next, the equivalent relation is combined with the explanation of high dimensional space geometry of neurons to

present a rational analysis of the working mechanism of the feedforward neural network. The following explanation is put forward: By studying positive and negative samples, the feedforward neural network partitions the sample spaces according to their categories. For a three-layer feedforwrd neural network, the inputting layer plays the role of forming a decision-making interface, the hidden layer plays the role of forming a decision region, and the output layer plays the role of summarizing. The causes and factors which affect the generalization and incremental learning ability of neural networks are also analyzed.

2. Put forward new methods of designing neural networks architecture based on information theory.

Based on the analysis of classified feedforward neural network, the thesis employs three-layer feedforwrd neural network as the research object and put forward new methods of designing neural networks architecture based on information theory. Taken into consideration of the classification of discrete and continuous attributes, neural network construction design methods—DTBNN (Decision Tree-Based Neural Networks) and EBNN (Entropy-Based Neural Networks) design methods—are proposed according to the specific meaning of the three-layer neurons. Dealing with the classification of data with discrete attribute, DTBNN, an improvement of the Entropy-net, uses decision tree to determine neural network structure, the weight value and the initial threshold value so as to improve the learning speed and the stability of the learning results of neural network. The experiments indicate that the convergence speed of DTBNN is faster than

ABSTRACT

Entropy-net and the recognition rate also increases. Dealing with the classification of data with continuous attribute, EBNN treats Entropy as a standard for choosing neurons to construct neural network models. EBNN neural network construction is much simpler and easier to understand. It can also set up some parameters of neural networks. In addition, the learning mechanism of separate layers and separate categories that EBNN adopts simplifies the learning process of neural networks and solves the credit assignment problem of feedforward neural network. With the same number of neurons, EBNN has a stronger capability than complete linking network. Therefore, it has better generalization ability and this method is sure to converge. The experiments results also prove the effect of EBNN.

3. Offer the incremental learning method of neural network based on ensemble system.

Based on the method of neural network ensemble system, the thesis shows that feedforward neural network with limited storage space and computational complexity can enable the neural network to maintain the uniqueness of knowledge structure when learning new knowledge, the so-called incremental learning ability. Based on ensemble system, the incremental learning has active learning as its foundation. It treats the sample of new knowledge as its learning focus, combines the features of neural network ensemble system, and attains incremental learning without any change of knowledge structure. With the aid of the process of positively choosing samples in Boosting technology, the calculation solves the problem of non-negative data and small sample of neural

network, and effectively chooses the individual network which constructs the ensemble system so as to possess a better generalization ability than LEARN++. The incremental learning calculation based on ensemble system increases a small amount of calculation, which can be neglected. The storage space needed is within a rational limit. This calculation has a general application, and can be used in other learning models.

Keyword: feedforward neural network, decision tree, entropy, incremental learning, ensemble

目 录

摘 要

ABSTRACT

第一章 概述.....	1
第一节 神经网络的产生与发展.....	1
一、神经网络的产生.....	1
二、神经网络的发展历史.....	3
第二节 神经网络研究内容.....	9
一、神经网络的研究内容.....	9
二、神经网络目前的研究热点和未来发展.....	12
第三节 本项研究的目的、意义和主要内容.....	15
一、本项研究的目的和意义.....	15
二、本项研究的主要内容.....	17
第四节 本书的组织安排.....	19
第二章 前馈神经网络的工作机理分析.....	20
第一节 引言.....	20
第二节 高维空间几何理论基础.....	26
一、基本概念.....	26
二、常用定理.....	30

第三节 神经网络的几何解释	32
一、神经元的几何意义	32
二、神经网络的高维空间几何意义	35
第四节 分类决策树与前馈网络的等价性	38
一、决策树	39
二、决策树的插值表示	43
三、前馈网络与决策树的等价性	47
第五节 分类前馈网络的决策树分析	54
一、前馈网络物理意义的决策树分析	54
二、神经网络学习和工作过程中常见问题的分析	60
第六节 小结	71
第三章 神经网络结构设计的信息论方法	73
第一节 神经网络结构设计	73
一、探索法	74
二、动态修改网络法	74
三、与符号系统相结合的结构设计方法	76
第二节 神经网络结构中的信息最大化	77
一、信息与熵	77
二、条件熵与互信息	80
三、神经网络中的信息理论	81
四、基于信息最大化的网络修剪	84
第三节 离散数据属性分类问题的神经网络结构设计	89
一、前馈网络与决策树	89

二、熵网络.....	92
三、基于决策树的神经网络构造方法DTBNN.....	94
四、实验分析与结论.....	98
第四节 基于熵准则的神经网络构造方法.....	103
一、简介.....	103
二、神经元选择的“熵”准则.....	106
三、基于熵准则的神经网络构造过程.....	110
四、多类问题的扩展.....	112
五、EBNN算法的主要性质.....	115
六、实验及结果分析.....	118
第五节 小结.....	122
第四章 基于集成系统的前馈网络增量学习.....	124
第一节 增量学习算法简介.....	124
一、增量学习的定义.....	125
二、神经网络增量学习.....	128
第二节 基于集成的增量学习.....	131
一、神经网络集成.....	131
二、AdaBoost的渐进学习能力.....	133
三、LEARN++增量学习算法.....	135
四、EBILNN增量学习.....	138
第三节 实验分析.....	145
一、人工实验数据的增量学习结果分析.....	145
二、Glass分类数据的增量学习性能.....	148

第四节 小结·····	151
第五章 研究工作总结与展望·····	153
第一节 研究工作的总结·····	153
一、前馈网络的工作机理分析·····	153
二、神经网络结构设计的信息论方法·····	154
三、基于集成系统的神经网络增量学习·····	155
第二节 神经网络技术的前景和进一步的研究工作·····	156
参考文献·····	159

第一章 概述

第一节 神经网络的产生与发展

人脑是由大量基本单元（称之为神经元）经过复杂的相互连接而形成的一种高度复杂的、非线性的、并行处理的信息处理系统，更重要的是它具有“认知”、“意识”和“感情”等高级功能，其性能要比现代计算机高得多。因此人们从模仿人脑智能的角度出发，探寻新的信息表示、存储和处理方式，设计全新的计算机处理结构模型，构造一种更接近人类智能的信息处理系统来解决实际工程和科学研究领域中传统的冯·诺伊曼计算机难以解决的问题。神经网络就是一种借用人脑规律的启迪，根据其原理，模仿设计求解问题的方法，它的中心问题是智能的认知和模拟。人工神经网络的研究首先突破了传统计算机的设计应用的思想，拓展了计算概念的内涵；其次，人工神经网络的研究有助于加深对思维及智能的认识，对认知和智力的本质的基础研究提供更有效的方法。因此人工神经网络的研究必将大大促进科学进步，并在人类生活的各个领域引起巨大变化，具有重要的理论和实际意义。

一、神经网络的产生

20 世纪 40 年代初，神经生物学家 McCulloch 与青年数学

家 Pitts 合作, 从人脑信息处理观点出发, 采用数理模型的方法研究了脑细胞的动作和结构及其生物神经元的一些基本生理特性, 提出了第一个神经计算模型, 即神经元的阈值元件模型, 简称 MP 模型 [McCu 1943]。MP 模型由一些节点及节点与节点之间相互联系构成一个简单神经网络模型——逻辑网络, 来模拟大脑功能。构成逻辑网络的节点具有很强的并行计算能力, 为神经计算提供了可能性, 从而开创了神经网络研究的历史。

实际上在此之前数学家 Turing 建立了通用计算机的抽象模型 [Turi 1936], 他和 Post 都证明了一个重要定理: 原则上存在着一种“万能自动机”, 它能识别任何别的自动机能够识别的符号 [Turi 1939, Post 1936]。Turing 机理论为带有存储程序的计算机的形式程序语言的发明提供了理论框架。Turing 机和逻辑神经网络之间或多或少的等价性进一步证实了大脑和计算机之间具有类似性的推测。

1949 年神经生物学家 Hebb 对大脑神经细胞、学习与条件反射作了大胆的假设, 也就是沿用至今的 Hebb 学习规则 [Hebb 1949]。其基本思想为: 假设大脑的学习过程是在神经元的突触上作微妙的变化, 这种变化是人脑学习和记忆的基础。神经元的突触由变化到稳定的过程导致了大脑自组织, 形成细胞集合 (几千个神经元的子集合), 其中循环神经冲动会自我强化, 并继续循环。任何一个神经元属于多个细胞集合, 可以说, 细胞集合是大脑思维信息的基本器件。Hebb 学习规则给出了突触调节模型, 描述了分布记忆模式。与人工智能的符号主义模拟系统 (比如规则推理系统) 对应, 以突触连接特点被称为连接主义 (connectionist) 模拟系统。Hebb 对神经网络的发展起到了重大的推动作用, 尽管初始的 Hebb 学习规则模型是被动学习过程, 只适用于正交矢量的情况, 但其思想对神经网络的研究有

极其深远的影响。学者们在此基础上经过多年的研究,通过把突触的变化与突触前后电位相关联,对 Hebb 学习规则作了变形和扩充,使之称为神经网络研究中经典的模型,至今仍然被人们引证。

1954 年生理学家 Eccles 提出了真实突触的分流模型,并通过突触的电生理实验得到证实,为神经网络模拟突触的功能提供了原型和生理学的证据[Ecc1 1954]。

20 世纪 50 年代初,神经网络理论具备了初步模拟实验的条件。Rochester, Holland 与 IBM 公司的研究人员合作,通过网络吸取经验来调节突触连接强度,以这种方式模拟 Hebb 的学习规则,出现了许多突现现象,几乎有大脑的处理风格,取得了成功的实验结果。

至此,生物学、神经生理学、神经生物学及其他学科的重大发现和发展为神经网络的快速发展奠定了理论基础,由此拉开了神经网络发展的序幕。

二、神经网络的发展历史

人工神经网络自产生以来,经过学者们不断的研究,形成了一系列的基于多学科的神经网络理论,建立了丰富的神经网络学习算法,并得到了广泛的应用。其主要的发展大约为四个阶段。

1. 第一阶段(1958~1968):神经网络研究的第一次高潮

1958 年计算机科学家 Rosenblatt 基于 MP 模型,增加了学习机制,推广了 MP 模型[Rose 1958]。他提出的感知器模型,