第 27 卷 第 6 期
2001 年 11 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 27, No. 6
Nov., 2001

综述与评论

神经网络的泛化理论和泛化方法

魏海坤　　徐嗣鑫　　宋文忠

（东南大学自动化研究所　南京　210096）

（E-mail：hkwei@seu.edu.cn）

摘　要　泛化能力是多层前向网最重要的性能，泛化问题已成为目前神经网络领域的研究热点. 文中综述了神经网络泛化理论和泛化方法的研究成果. 对泛化理论，重点讲述神经网络的结构复杂性和样本复杂性对泛化能力的影响；对泛化方法，则在介绍每种泛化方法的同时，尽量指出该方法与相应泛化理论的内在联系. 最后对泛化理论和泛化方法的研究前景作了展望.

关键词　神经网络，泛化能力，泛化理论，泛化方法.

GENERALIZATION THEORY AND GENERALIZATION METHODS FOR NEURAL NETWORKS

WEI Hai-Kun　XU Si-Xin　SONG Wen-Zhong

(*Research Institute of Automation, Southeast University, Nanjing*　210096)

(E-mail: hkwei@seu.edu.cn)

Abstract　Generalization ability is the most important performance of a feed-forward neural network, and the problem of generalization has been widely studied recently among the neural network community. Research on this subject can be divided into two fields: generalization theory discusses the factors that affect the generalization ability, while generalization methods try to find algorithms for improved performance. This survey reviewed the main results on generalization research, and tried to point out the relationship between generalization theory and corresponding generalization methods. A prospect on generalization research was also given in the last part of this paper.

Key words　Neural networks, generalization ability, generalization theory, generalization methods.

1　引言

多层前向网是指拓扑结构为有向无环图的前向网络，包括 MLP，BP 网、RBF 网等，是实际中应用最多的神经网络. 但是，多层前向网还存在许多理论上的问题有待解决，如泛化

收稿日期　1999-07-27　　收修改稿日期　2000-03-20

能力、结构设计、算法的收敛速度、局部最小点等,其中泛化能力是人们最关心的问题.

多层前向网的泛化能力是指学习后的神经网络对测试样本或工作样本作出正确反应的能力.所以,没有泛化能力的神经网络没有任何使用价值.正因为其重要性,泛化问题已成为近年来国际上十分关注的理论问题.这一问题也引起了国内一些学者的注意,如张鸿宾[1]讨论了多种情况下为保证多层前向网的泛化能力所需的样本数问题;阎平凡等人[2,3]在分析了多层前向网的泛化能力与结构复杂性和样本复杂性关系的同时,也介绍了一些神经网络的结构选择方法.

多层前向网的泛化问题是指

1) 哪些因素影响神经网络的泛化能力,它们是如何影响神经网络泛化能力的;

2) 在样本来源可靠的情况下,对于固定结构的神经网络,需要多少训练样本才能使神经网络达到给定的泛化能力;

3) 如何选择神经网络结构以保证网络的泛化能力,而影响神经网络泛化能力的其它因素可通过学习算法来加以避免?

上述问题中,前两个问题属于泛化理论的范畴,第三个问题属于泛化方法的范畴.对于第一个问题,人们已经发现许多影响泛化能力的因素,这些因素包括神经网络的结构复杂性、训练样本的数量和质量、初始权值、学习时间、目标规则的复杂性、对目标规则的先验知识等.但除了网络结构和训练样本数对泛化能力的影响(即第二个问题)已有一些定量的结果外,其余因素对泛化能力的影响还只有定性的解释.由于影响神经网络泛化能力的主要因素是神经网络的结构复杂性和样本复杂性,所以第二个问题才是泛化理论的核心问题.

第三个问题,即泛化方法问题,也是目前神经网络领域的研究热点之一.泛化方法与泛化理论密切相关,对于几乎每种影响泛化能力的因素,都已提出了相应的解决方案以改进神经网络的泛化能力.但目前泛化方法更主要的是集中在神经网络结构设计和正则化方法方面.

本文综述了泛化理论和泛化方法的研究成果.对泛化理论,我们在讨论各因素对泛化能力影响的同时,重点讲述神经网络的结构复杂性和样本复杂性对泛化能力的影响,即对第二个问题作较深入的讨论;对泛化方法,则在介绍每种泛化方法的同时,尽量指出该方法与相应泛化理论的内在联系.在本文最后,我们还对当前泛化理论和泛化方法研究中存在的一些问题作了探讨,并对其研究前景作了展望.

2　神经网络的泛化理论

尽管影响神经网络泛化能力的因素很多,但在神经网络结构设计时,唯一的信息通常就是一定数目的训练样本.因此,人们更关心神经网络的泛化能力与网络结构复杂性和样本数之间的定量关系,即引言中的第二个问题.下面先重点介绍结构复杂性和样本复杂性对神经网络泛化能力的影响,再简单介绍其它因素对泛化能力的影响.

2.1　结构复杂性和样本复杂性对神经网络泛化能力的影响

神经网络的结构复杂性是指神经网络的规模或容量,对线性阈值神经网络来说是指神经网络函数类的 VC 维数,对函数逼近神经网络来说则可用权参数和隐节点数目来衡量.样本复杂性(sample complexity)是指训练某一固定结构神经网络所需的样本数.

结构复杂性和样本复杂性对泛化能力的影响问题在泛化理论中得到了最多的研究,也获得了许多定量的成果.所研究的神经网络类型也涵盖了多种最常见的前向网络,如线性阈值神经网络和函数逼近神经网络(包括 RBF 网络).

2.1.1 线性阈值神经网络

线性阈值神经网络是指激活函数为线性阈值型的任意多层前向网络(包括 MLP),是被研究得最早的多层前向网.泛化理论最重要的成果之一,便是对单输出线性阈值神经网络,得出了泛化能力与网络结构复杂性、训练样本数和学习精度之间的关系.

线性阈值神经网络的泛化理论欲解决以下问题:假定训练样本和工作样本都取自某一不变但可以是任意的分布,网络中的计算节点数和自由参数个数分别为 N 和 W,对 l 个训练样本的允许学习误差为 ε,欲以一定概率保证学习后的神经网络对工作样本的分类正确率至少为 $1-\varepsilon$,则训练样本数 l 应取多大? 这一提法沿用 Valiant[4]的 PAC 学习框架.应该指出,在上述框架下,以学习后的神经网络对工作样本正确分类的概率来衡量神经网络的泛化能力.解决上述问题的数学工具是 Vapnik 和 Chervonenkis 等人[5,6]的经验风险最小与期望风险最小之间关系的理论.

关于 Vapnik 等人的理论及相关的 VC 维数、成长函数等概念,张鸿宾[1]有扼要的介绍.

对线性阈值神经网络, Baum 和 Haussler[7]得到了以下重要结论:假使线性阈值神经网络有N个计算节点和W个自由参数,$0<\varepsilon\leqslant 1/8$为逼近误差,且训练样本数$m\geqslant 0(W/\varepsilon\log W/\varepsilon)$,如果神经网络对至少 $1-\varepsilon/2$ 的训练样本能正确分类,则可以相信,该神经网络至少能正确分类 $1-\varepsilon$ 的未来工作样本.由此得到了为保证固定结构神经网络的泛化能力所需的训练样本数上界.对类似网络,张鸿宾[1]也给出了多种情况下为保证神经网络泛化能力所需要的样本数.

对单隐层全连接前向网络, Baum 与 Hauslle[7]还指出,如果训练样本数小于 $\Omega(W/\varepsilon)$,则即使神经网络对训练样本能全部正确学习,也不能保证学习后的神经网络有好的泛化能力,从而给出了样本复杂性下界.

由于计算上面的样本复杂性边界时考虑了最不利的情况(如独立于样本分布),因此许多人都试图通过增加新的限制条件,以缩小样本复杂性边界[8~10].考虑到神经网络的 VC 维数对样本复杂性的巨大影响,另一个改进样本复杂性边界的方法是获得更好的 VC 维数边界.但是,尽管对神经网络函数类的 VC 维数获取已有一些结论[11~13],但总的说来,VC 维数的获得依然是一个困难的课题;其次,有些人发现,由于 VC 维数只是神经网络复杂性的一种粗略测度[14,15],因此,Bartlett[15]提出了粗分(fat-shattering)维数的定义,并以此代替 VC 维数计算泛化误差,得到了比 Baum 等人更好的结果,但所得结果离真正实用仍有距离.

应该指出,上述理论依然有着重要的意义:它通过研究神经网络泛化能力与结构参数之间的定量关系,首次证实了神经网络结构设计的最简原则.它给我们以这样的信息:要使一个神经网络达到给定的泛化能力,必须使神经网络的结构复杂性与训练样本数匹配.我们的选择只能是,要么增加训练样本,要么减小神经网络规模.

2.1.2 函数逼近神经网络

这里的函数逼近神经网络是指输入输出均取实值、隐单元取 Sigmoid 函数或 RBF 函数,而输出节点采用线性函数的前向网络,该网络常用于函数逼近.在下面的讨论中,假定该神经网络含 d 个输入节点、n 个隐节点、一个线性输出节点、训练样本数为 N.

Moody[16]研究了期望测试集误差(泛化误差)与期望训练集误差之间的关系,得到了关系式$\langle \varepsilon_{test}(\lambda)\rangle_{\xi\xi'} \approx \langle \varepsilon_{train}(\lambda)\rangle_{\xi} + 2\sigma_{eff}^2\, p_{eff}(\lambda)/N$,其中$\lambda$为正则化参数,$p_{eff}(\lambda)$和$\sigma_{eff}^2$分别为有效参数个数及输出变量的有效噪声方差. $p_{eff}(\lambda)$并不等于神经网络中自由参数的个数p,它与模型偏差、非线性程度、正则化参数及正则化项的形式有关.上述重要结论被称为Moody准则,该准则证实了实值神经网络结构设计的最简原则:对已达到给定训练精度的神经网络,其有效参数越少,泛化能力就越好,从而为设计最小复杂性神经网络的合理性提供了理论基础.

对单隐层且隐节点取Sigmoid函数的函数逼近神经网络,假定网络函数为$f_{n,N}$,目标函数为f,Barron[17]以f和$f_{n,N}$之间的期望偏差$err(f_{n,N})=E(\|f-f_{n,N}\|)$来衡量神经网络的泛化能力,并给出了学习后神经网络的泛化误差同样本数和隐节点数之间的关系$err(f_{n,N}) \leqslant O(C_f^2/n)+O(nd/N)\log N$,其中$C_f=\int |\omega|_1 |\tilde{f}(\omega)|\mathrm{d}\omega$,$|\omega|_1$为$\omega$的$l_1$范数,$\tilde{f}(\omega)$为$f$的Fourier变换. $err(f_{n,N})$上界的第一项称为逼近误差(偏差),第二项称为估计误差(方差).由偏差和方差的关系可见,随着隐节点数n的增加,偏差将逐步减小,而方差将逐步增大,好的泛化能力取决于两者的协调.这也同German等人[18]的结论是一致的.

对RBF网,Niyogo等人[19]也得到了与Barron[17]类似的泛化误差分解,对任意$0<\delta<1$,有$\|f_0-\hat{f}_{n,N}\|^2_{L^2(P)} \leqslant O\left(\frac{1}{n}\right)+O\left(\left[\frac{nd\ln(nN)-\ln\delta}{N}\right]^{1/2}\right)$,其中$f_0$为目标函数,$\hat{f}_{n,N}$为神经网络函数.

2.2　其它因素对泛化能力的影响

除了神经网络的结构复杂性和样本复杂性之外,还有许多其它因素将影响神经网络的泛化能力,这些因素中目标规则的复杂性是不可控的,因此下面简单介绍样本质量、对目标函数的先验知识、初始权值及学习时间对神经网络泛化能力的影响.对它们的研究目前大都停留在定性讨论的阶段.

2.2.1　样本质量

样本质量指训练样本分布反映总体分布的程度,或者说整个训练样本集所提供的信息量.尽管样本质量对神经网络的泛化能力有相当大的影响,但定量分析样本质量对泛化能力的影响却是一个非常困难的课题. 但Partridge[20]对用于分类的三层BP网的研究发现,训练集对泛化能力的影响甚至超过网络结构(隐节点数)对泛化能力的影响.

采用主动学习(active learning)[21]机制(选择采样),改进训练样本质量,也是改善神经网络泛化能力的一个重要方法.

2.2.2　先验知识

一般来说,给定一组样本,欲求得样本后面的"目标规则",其解将有无穷多个.而先验知识实际上是对所学习的目标规则所做的一些合理的假定.先验知识一般有两种表现方式:一种是Abu-Mostafa[22,23]所提出的提示(hint);另一种是假定神经网络权值服从某种先验分布[24,25].后一种方式在神经网络分析中引入贝叶斯方法,因而更易为人接受.

使用先验知识,相当于对模型的参数维数加以限制,从而使模型更靠近"目标规则"所在的区域.事实上,加入先验知识相当于减小了神经网络函数类的VC维数[26,27],或者说增加了神经网络的训练样本[26].

正则化[28~30]是一种非常有效的神经网络泛化方法,该方法中损失函数的一般形式为

$C(w)=S(w)+\xi * R(w)$，其中 $S(w)$一般取误差平方和，$R(w)$是正则化项，ξ 为正则化系数. 常用的正则化项的形式与神经网络权值的先验分布的对应关系如下[25]：

$$R(w)=\frac{1}{2}\sum_{j=1}^{N}w_j^2 \quad \text{(Gauss 分布)},$$

$$R(w)=\sum_{j=1}^{N}|w_j| \quad \text{(Laplace 分布)},$$

$$R(w)=\frac{1}{\alpha}\sum_{j=1}^{N}\log(1+\alpha^2 w_j^2) \quad \text{(Cauchy 分布)}.$$

上式中 N 都表示权参数的数目，α 为一中等大小的常数. 应该指出，Weigend 等人[29]以 $w_j^2/1+w_j^2$代替 $\log(1+\alpha^2 w_j^2)$，实际上同 Cauchy 分布是类似的.

如果在学习过程中最小化上述损失函数，则神经网络的冗余权值将随着学习的进行逐步衰减到零附近，因而可以被“剪除”，这一特性常被称为剪枝特性. 通常认为，Laplace 先验的剪枝特性优于 Gauss 先验[25].

2.2.3 初始权值

许多人发现，BP 算法对初始权值极为敏感，在只有初始权值不同的情况下进行训练，将得到不同泛化性能的神经网络[31,32].

对采用线性隐节点的两层线性网络来说，当使用梯度下降算法时，Polycarpou 等人[33]已经证明，权值的最终解仅仅是解曲面上最靠近初始权值的点. 对以 Sigmoid 函数为激活函数的神经网络，Atiya 等人[34]指出，最终权值将收敛到目标函数的某个局部最小点，且该最小点最接近于初始状态时的目标函数值. 在过参数的情况下，目标函数的最小点也将不再是一个点，而是一个误差为零的解曲面，此时解的情况与线性网络的情形类似. 该文还指出，用较小的初始权值进行训练，最终将得到低复杂性的神经网络，因而小的初始权值对泛化性能的影响类似于权衰减法中使用大的正则化系数.

2.2.4 学习时间

神经网络的学习时间指神经网络的训练次数. 过多的训练无疑会增加神经网络的训练时间，但更重要的是会使神经网络拟合数据中的噪声信号，产生所谓的过学习(over-learning)，从而影响神经网络的泛化能力.

小川英光等[35]认为，学习和泛化的评价基准不一样，是过学习产生的原因. Baldi[36]和 Chauvin[37]对线性网络的训练和泛化动态进行了研究，发现神经网络的泛化误差非常依赖于初始权值，并存在泛化误差的最小点. Wang 等[38,39]对单隐层神经网络训练的动态过程进行分析后发现，泛化过程可分为三个阶段：在第一阶段，泛化误差单调下降；第二阶段的泛化动态较为复杂，但在这一阶段，泛化误差将达到最小点；在第三阶段，泛化误差又将单调上升. 他们还发现，最佳的泛化能力出现在训练误差的全局最小点出现之前，并给出了最佳泛化点出现的时间范围，从而从理论上证明了在神经网络训练过程中，存在最优的停止时间. 上述理论也说明，只要训练时间合适，较大的神经网络也会有好的泛化能力. 这就是用最优停止法设计神经网络的主要思想.

3 神经网络的泛化方法

神经网络的泛化方法中，研究最多的是神经网络的结构设计方法，它包括剪枝算法、构

造算法及进化算法等.除了结构设计,其余泛化方法还有主动学习、最优停止、在数据中插入噪声、表决网及提示学习方法等.

3.1　结构设计方法

由 2.1 节的讨论可知,无论是线性阈值神经网络,还是函数逼近神经网络,给定一组训练样本,都存在同样本复杂性匹配的最小结构神经网络,该结构下的神经网络将有最好的泛化能力.这一原则同奥卡姆剃刀(Ocam's razor)原则也是一致的,即在所有达到给定学习精度的神经网络中,结构越简单,泛化能力就越好.因此,当神经网络的结构复杂性与样本复杂性协调时,神经网络就会有较好的泛化能力.

主要的结构设计方法有剪枝方法[40]、构造方法[41]、进化方法[42]和信息论方法[43,44]等,其基本思路都是通过调整神经网络的权值或隐节点数目,实现结构复杂性与样本复杂性的最佳匹配.鉴于篇幅,且这些方法大都已有综述[45],这里不再深入讨论.

3.2　主动学习

由于训练样本质量对神经网络的泛化能力有极大影响,在每次采样代价较大时采用主动学习方法,是改进神经网络泛化能力的一个重要方法.

主动学习通过对输入区域加以限制,有目的地在冗余信息较少的输入区域进行采样,从而提高了整个训练样本集的质量,改善了神经网络的泛化能力.目前的主动学习机制大部分用于分类或概念学习,且一般通过"询问"(query)的方式实现[21,46]:在输入定义域内按某种概率取一点 x,判断该点是否位于不确定区,如果不位于不确定区,则抛弃该点;否则"询问"该点输出 y(进行一次采样),并把(x,y)加入样本集进行训练,直至采到足够的样本.主动学习也可用于函数逼近,Mackay[47]讨论了贝叶斯框架下候选样本输入点信息量的几个测度,可用于函数逼近问题的选择采样.

3.3　在样本输入中添加随机噪声

在样本输入中添加随机噪声,也是改善神经网络泛化能力的一种有效方法[48].在噪声方差较小时,Bishop[49]已经证明,在样本输入中添加噪声,等价于神经网络结构设计的正则化方法,而正则化系数则与噪声方差有关.事实上,当训练数据被循环地作为网络的输入时,由于每次添加的噪声不同,结果是一方面相当于增加了训练样本的数量,另一方面迫使神经网络不能精确地拟合训练数据,从而使噪声起到平滑作用[50],防止了过拟合.

3.4　表决网

由 2.2.3 节可知,选择不同的初始权值,将使神经网络的泛化能力体现出一定的随机性.利用这一特性也可以改善神经网络的泛化能力,表决网(或称多神经网络)便是这种思路的体现.该方法先训练一组只有初始权值不同的子网,然后通过各子网"表决"的方式得到学习系统的输出. Sarker[51]和 Lincoln 等[52]用普通 BP 算法训练多个相同结构的子网,并把各子网的输出综合,作为整个系统的输出,整个系统的泛化性能明显优于单个 BP 子网.

对模式分类问题, Sarker[53]指出,当单个子网正确分类的概率大于 0.5 时,随着表决系统中子网数目的增加,整个表决系统的泛化能力也将增加.对函数逼近问题,上田修功和中野良平[54]也都定量分析了表决系统的泛化误差同学习样本特性及每个子网特性之间的关系.表决网能改进泛化能力的原因,也可以从算法复杂性角度加以解释[55]:使用表决系统降低了整个学习系统的复杂性,从而提高了表决系统的泛化能力. Partridge 和 Yates[56]还给出了一种设计最优表决系统的方法.

3.5 基于先验知识的泛化方法

前面已经介绍过,对目标概念的先验知识可以表示成神经网络参数先验分布的形式,并由此导出了结构设计的正则化方法. 先验知识也可以通过在学习过程中嵌入提示[22,23]来表示. 提示能成功应用的关键是如何在学习过程中插入提示. Abu-Mostafa[57]建议把提示转化为虚拟样本(virtual samples),然后以虚拟样本和普通样本作为共同的训练样本集,完成神经网络的训练. 在证券市场预测时使用对称性提示[57],以及在手写数字识别时使用最小Hamming 距离作为提示[58],都得到了较好的效果.

先验知识也可以以其余方式嵌入学习. 如 Jean 和 Wang[59]在目标函数中增加一个权值光滑性限制项,来反映神经网络相邻输入之间的空间相关性,也得到了很好的效果.

3.6 最优停止法

考虑在适当的时间停止学习,即当神经网络的泛化误差达到最小时停止学习,也是改进神经网络泛化能力的重要方法.

Sjoberg 等[60]指出,最优停止方法也是一种隐式的正则化方法. Cataltepe 等人[61]则认为,最优停止方法所设计的神经网络之所以有好的泛化能力,是因为最优停止法倾向于设计低复杂性的神经网络.

用最优停止法设计神经网络的关键是确定学习算法在何时停止学习. 为了得到最优停止点,可以把所有样本数据分为训练数据和测试数据,并当测试误差达到最小时停止网络的训练. 该方法实际上就是最简单的交叉测试(cross-validation,简称 CV)方法.

应该指出,当训练样本数相对于神经网络规模较小时,简单 CV 方法并不适用[62],此时应使用改进的 CV 方法,如留一 CV 法(leave-one-out CV 法)和 v 重 CV 法(v-fold CV 法).

Amari 等人[63]和 Kearns[64]还研究了 CV 方法中测试样本数占总样本数比例对神经网络泛化能力的影响.

4 讨论

近年来,神经网络泛化理论和泛化方法的研究取得了许多成果. 但是,无论是泛化理论还是泛化方法,都依然存在许多问题待解决.

1) Baum 和 Haussler[7]给出的是最不利情况下为保证神经网络的泛化能力所需的样本数. 如何根据应用问题的背景或先验知识增加限制条件,降低样本复杂性边界,依然是今后泛化理论的研究内容之一. 而如何计算各种神经网络函数类的 VC 维数,也有待于进一步研究.

2) 人们更关心函数逼近神经网络的样本复杂性问题,即训练实值输出神经网络所需的样本数. Haussler[65]把 PAC 模型推广到函数逼近神经网络,并对实值输出函数类引入了伪维数(pseudo dimension)等概念,来衡量函数逼近神经网络的复杂性,为解决这一问题提供了一条途径,引起了一些人的兴趣.

3) 最近, Wolpert[66,67]提出了 NFL(no free lunch)定理,该定理指出:如果以 OTS(off-training-set)误差作为泛化能力测度,且目标函数的先验分布是均匀分布时,任何两种算法都应该有相同的泛化性能. 这意味着,如果不指定某算法所适用的目标函数的先验分布,那

么声称该算法优于别的算法是没有意义的,因为不存在各个场合普遍"好"的算法.所以,寻找各种神经网络泛化方法背后隐含的先验分布,具有极大的理论和应用价值.

4) 对泛化方法来说,考虑到神经网络结构对泛化能力的巨大影响,结构设计,尤其是最小结构设计,依然是神经网络泛化方法的主流.另外,我们已经讨论过,先验知识对神经网络泛化能力有极大的影响,而神经网络最终总是要用于实际问题的.所以,如何把对实际问题的先验知识嵌入神经网络学习,是改进神经网络泛化的最有效途径之一.

参 考 文 献

1 张鸿宾. 训练多层网络的样本数问题. 自动化学报, 1993, **19**(1):71～77

2 阎平凡. 人工神经网络的容量、学习与计算复杂性. 电子学报, 1995, **23**(4):63～67

3 张乃尧,阎平凡,李衍达. 神经网络与模糊控制. 北京:清华大学出版社, 1998

4 Valiant L G. A theory of learnable. *Comm. of the ACM*, 1984, **27**(11):1134～1142

5 Vapnik V. Estimation of Dependcies Based on Emperical Data. New York: Springer-Verlag, 1982

6 Vapnık V, Chervonenkis A. On the uniform convergence of relative frequencies of events to their probabilities. *Theory Prob. Appl.*, 1971, **16**(2):264～280

7 Baum E B, Haussler D. What Size Net Gives Valid Generalization?. NIPS 1, 1989, San Mateo, CA:81～90

8 Turmon M J, Fine T L. Sample Size Requirements for Feedforward Neural Networks Classifiers. In:*IEEE* 1993 Intern. Sympos. Inform. Theory, 432

9 Kowalczyk A, Ferra H. Generalızatıon in Feedforword Networks. NIPS 7, 1995, Cambridge, MA:215～222

10 Turmon M J, Fine T L. Sample Size Requirements for Feedforward Neural Networks. NIPS 7, 1995, Cambridge, MA:327～334

11 Koiran P, Sontag E D. Neural Networks with Quadratic VC-dimension. NIPS 8, 1996, Cambridge, MA:197～203

12 Mass W. Neural nets with superlınear VC-dimension. *Neural Computation*, 1994, (6):877～884

13 Vapnik V, Levin E, LeCun Y. Measuring the VC-dimension of a learning machine. *Neural Computation*, 1994, (6):851～876

14 Kowalczyk A, Ferra H. MLP Can Provably Generalize Much Better Than VC-bound Indicate. NIPS 9, 1995, Cambridge, MA:190～196

15 Bartlett P L. For Valid Generalization, the Size of the Weight Is More Important Than the Size of the Network. NIPS 9, 1995, Cambridge, MA:134～140

16 Moody J E. The Effective Number of Parameters: An Analysis of Generalizatıon and Regularization in Nonlinear Learning System. NIPS 4, 1992, San Mateo, CA:847～854

17 Barron A R. Approximatıon and estımation bounds for artificial neural networks. *Machine Learning*, 1994, (14): 115～133

18 Geman S. Neural networks and bias/variance dilemma. *Neural Computation*, 1992, (4):1～58

19 Niyogo P, Girosi F. On the relatıonshıp between generalization error, hypothesis complexity, and sample complexity for radial basis function. *Neural Computation*, 1996, (8):819～842

20 Partridge D. Network generalizatıon dıfferences quantıfied. *Neural Networks*, 1996, **9**(2):263～271

21 Cohn D. Improving generalization with active learning. *Machine Learning*, 1994, (15):201～221

22 Abu-Mostafa Y S. Learning from hints in neural networks. *Journal of Complexity*, 1990, (6):192～198

23 Abu-Mostafa Y S. A Method of Learning from Hints. NIPS 5, 1993, San Mateo, CA:73～80

24 Mackay D J C. Bayesian interpolation. *Neural Computation*, 1992, **4**(3):415～447

25 Williams P M. Bayesian regularızation and pruning using a laplace prior. *Neural Computation*, 1995, (7):117～143

26 Abu-Mostafa Y S. Hints and the VC-dimention. *Neural Computation*, 1993, (5):278～288

27 Solla S A. Capacity control in classifiers for pattern recognition. In:Kung S Y, Fallside F, Sorenson J, Kamm C A Eds. Neural Networks for Signal Processing II-1992, IEEE Workshop, Piscataway, NJ:255～266

28 Poggio T. Networks for approximation and learning. *Proceeding of the IEEE*, 1990, **78**(9):1481～1497

29 Weigend A S, Rumelhart D E, Huberman B A. Generalization by Weight-Elimination with Application to Forcasting, NIPS 3, 1991, San Mateo, CA:875～882

30 Girosi F, Jones M, Poggio T. Regularization theory and neural networks architecture. *Neural Computation*, 1995, (7):219～269

31 Kolen J F, PollackJ B. Back Propagation is Sensitive to Initial Conditions. NIPS 3, 1991, San Mateo, CA: 860～867

32 Schmidt W F, Raudys S, Kraaijveld M A *et al*. Initialization, back-propagation and generalization of feed-forward classifiers. In: Proc. IEEE Int. Conf. Neural Networks, 1993. 598～604

33 Polycarpou M, Ioannou P. Learning and convergence analysis of neural-type structured networks. *IEEE Trans. Neural Networks*, 1992, (3):39～50

34 Atiya A, Ji C. How initial conditions affect generalization porformance in large networks. *IEEE Trans. Neural Networks*, 1997, (8):448～451

35 小川英光,山崎一孝. 过学习の理论. 信学论, 1993, **J76-D-Ⅱ**(6):1280～1288

36 Baldi P. Temporal evolution of generalization during learning in linear networks. *Neural Computation*, 1991, (3): 589～603

37 Chauvin Y. Generalization Dynamics in LMS Trained Linear Networks. NIPS 3, 1991, San Mateo, CA:890～896

38 Wang C, Venkatesh S S. Optimal Stopping and Effective Machine Complexity in Learning. NIPS 6, 1994, San Mateo, CA:303～310

39 Wang C, Venkatesh S S. Temporal Dynamics of Generalization in Neural Networks. NIPS 7, 1995, Cambridge, MA:263～270

40 Reed R. Pruning algorithms: A survey. *IEEE Trans. Neural Networks*, 1993, (4):740～747

41 Kwok T Y, Yeung D Y. Constructive algorithms for structure learning in feedforward neural networks for regressing problems. *IEEE Trans. Neural Networks*, 1997, (8):630～645

42 Maniezzo V. Genetic evolution of the topology and weight distribution of neural networks. *IEEE Trans. Neural Networks*, 1994, (5):39～53

43 Murata N, Yoshizawa S, Amari S. Network information criterion-determining the number of hdden units for an artificial neural network model. *IEEE Trans. Neural Networks*, 1997, **5**(6):865～872

44 栗田喜多夫. 情报量基准による3层ニユ～ラルネットの隐れ层のユニットの决定法. 信学论, 1990, **J73-D-Ⅱ**(11): 1872～1878

45 何述东,瞿坦,黄献青,黄心汉. 多层前向神经网络结构的研究进展. 控制理论与应用, 1998, **15**(3):313～319

46 Baum E B. Neural networks algorithm that learn in polynomial time from examples and queries. *IEEE Trans. Neural Networks*, 1991, **2**(3):5～19

47 Mackay D J C. Information-based objective function for active data selection. *Neural Computation*, 1992, (4):590～604

48 Sietsma J, Dow R J F. Creating artificial neural networks that generalization. *Neural Networks*, 1991, (4):67～79

49 Bishop C M. Training with noise is equivalent to Tikhonov regularization. *Neural Computation*, 1995, (7):108～116

50 An G. The effect of adding noise during backpropagation training on a generalization performance. *Neural Computation*, 1996, (8):643～671

51 Sarker D. Improving generalization through multiple output unit. In: Proc. World Congr. Neural Networks, 1993

52 Lincoln W L, Skrzypek J. Synergy of Clustering Multiple Back Propagation Networks. NIPS 2, San Mateo, CA: 1990. 650～657

53 Sarker D. Randomness in generalization ability: A source to improve it. *IEEE Trans. Neural Networks*, 1996, (7):676～685

54 上田修功,中野良平. アンサンブル学习における泛化误差解析. 信学论, 1997, **J80-D-Ⅱ**(9):2512～2521

55 Pearlmutter B A, Rosenfield R. Chaitin-Kolmogorov Complexity and Generalization in Neural Networks. NIPS 3, 1991, San Mateo, CA:925～931

56 Partridge D, Yates W B. Engineering multivertion neural-net systems. *Neural Computation*, 1996, (8):869～893

57 Abu-Mostafa Y S. Financial application of learning from hints. NIPS 7, 1995, Ambridge, MA:411～418

58 Al-Mashouq K A, Reed I S. Including hints in training neural networks. *Neural Computation*, 1991, (3):418～427

59 Jean J S N, Wang J. Weight smoothing to improve network generalization. *IEEE Trans. Neural Networks*, 1994, **5**(5):752～763

60 Sjoberg J, Ljung L. Overtraining, regularization and searching for a minimum, with application to neural networks. *International Journal of Control*, 1995, **62**(6):1391～1407

61 Cataltepe Z, Abu-Mostafa Y S, Magdon-Ismail M. No free lunch for earlystopping. *Neural Computation*, 1999, (11):995～1009

62 Muller K R, Finke M, Murata N *et al*. A numerical study onlearning curves in stochastic multilayer feedforward networks. *Neural Computation*, 1996, (8):1085～1106

63 Amari S, Murata N, Muller K R *et al*. Asymptotic statistical theory of overtraining and cross-validation. *IEEE Trans. Neural Networks*, 1997, **8**(5):985～996

64 Kearns M. A bound on the error of cross validation using the approximation and estimation rates, with consequences for the training-test split. *Neural Computation*, 1997, (9):1143～1161

65 Haussler D. Decision theoretic generalization of the PAC model for neural net and other learning application. *Information and Computation*, 1992, (100):78～150

66 Wolpert D H. The lack of a priori distinctions between learning algorithms. *Neural Computation*, 1996, (8):1341～1390

67 Wolpert D H. The existence of a priori distinctions between learning algorithms. *Neural Computation*, 1996, (8):1391～1420

魏海坤　博士后. 研究方向为神经网络的泛化理论和泛化方法.

徐嗣鑫　教授. 研究领域为模糊控制、神经网络及 MIS 系统.

宋文忠　教授、博士生导师. 研究方向为 DEDS、过程辨识和控制.

第 12 届中国过程控制年会在沈阳召开

岳　恒

（东北大学自动化研究中心　沈阳　110004）

第 12 届中国过程控制年会于 2001 年 8 月 5 日～8 日在辽宁省沈阳市举行，本届年会由中国自动化学会过程控制专业委员会和中国有色金属学会计算机学术委员会共同主办，东北大学自动化研究中心和中南大学信息科学与工程学院共同承办.

本届会议共有来自中国大陆、香港、英国、日本、澳大利亚等地的 170 余位学者和专家参加. 开幕式上，辽宁省科技厅领导和沈阳市科委的领导以及东北大学领导发表了热情洋溢的讲话，向大会的成功举办表示了热烈的祝贺，对来自国内外的学者和专家们的到来表示了热烈的欢迎. 大会程序委员会双主席之一东北大学柴天佑教授致开幕辞.

本届年会共安排了 5 场大会特邀报告，报告题目和报告人分别是：自动化学科现状与发展的讨论（吴澄教授，中国工程院院士、国家“863”高技术计划自动化领域首席科学家）；“十五”“863”先进制造与自动化领域战略发展的思考（孙家广教授，中国工程院院士、国家“十五”“863”高技术计划先进制造与自动化领域

（下转第 849 页）