国 防 科 技 大 学 学 报
第20卷 第4期 JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY Vol.20 No.4 1998

神经网络的函数逼近理论*

李明国 郁文贤

(国防科技大学电子技术系 长沙 410073)

摘 要 分析了将函数逼近理论与方法引入神经网络研究的必要性；从经典函数逼近与统计分析两方面详细地讨论了多层前馈网(MLP)逼近能力分析的基本方法及结论；分析了正则理论观点下的径向基函数网络(RBF)的逼近能力；讨论了RBF网与多层前馈网在最佳逼近特性上的差异。文末指出了神经网络函数逼近的发展方向。

关键词 神经网络，函数逼近，正则理论，推广能力

分类号 TP18

On Neural Networks Founction Approximation

Li Mingguo Yu Wenxian

(Department of Electronics Engineering, NUDT, Changsha, 410073)

Abstract In this paper, an analysis of the necessity to introduce classical function approximation theory and methods into the research of ANN is presented with a brief introduction of its contents. Then, the main results about the approximation capability of MLP network and the basic analysis methods are detailed from the two aspects of classical function appoximantion and statistical analysis. The approximation capability of RBF ntework is analyzed under the view of Regularity Theory, and the difference of the best approximation property between the RBF and MLP network is revealed.

Key words artificial neural networks, function approximation, regularization theory, generalization property

神经网络走向实用化的一个很大的障碍是系统能力验证(system validation)研究的落后，从而导致在实际应用时对网络的性能没有一个数量上的把握。目前，多层前馈网络由于具有对非线性现象的强有力的刻画与建模能力而得到广泛应用。该类网络可以通过例子学习到隐含在训练样本中的模式总体分布信息，在模式识别中可用以估计后验概率密度，从而得到最优BAYES判别边界[1]。在系统辨识及控制中可用以对一般非线性函数乃至定义在Banach空间上的非线性泛函与非线性算子进行一致最优逼近[2]。经训练的BP(反向传播网络)网可以基于少量的样本对具有混沌特性的Mackey-Glass延迟差分方程与著名的LORENTZ方程作精确的逼近[3]。在上述广泛应用中，前述障碍变得逾加明显，迫切需要解决。神经网络函数逼近理论是解决上述问题的一条可行途径。

本文所介绍的神经网络函数逼近理论包括两方面的内容：(1)基于经典函数逼近论的理论与方法发展一种非线性分析方法，对前馈网络这样一个非线性系统进行分析，以指导网络的构造并对构造与训练完成的网络的映射能力作出一个理论上的分析，从而推进神经网络的实用化进程。(2)基于前馈网络特殊的非线性映射能力，发展一种新的函数逼近方法

1 概述

神经网络本质上是一个信息的非线性变换系统。设一个三层网的输入层，隐含层及输出层的节点数

* 1998年3月2日收稿
第一作者：李明国，男，1972年生，博士生

分别为 n，k，p，第 j 个隐层节点的传递函数为 $\phi_j(\ )$，则 p 个输出分别为:

$$y_i = \sum_{j=1}^{k} C_{ji}\phi_j(W_j^T X + \theta_j), i = 1, 2, \cdots p \tag{1}$$

显然，若将式(1)中 ϕ_j 看作逼近的基函数，则式(1)定义了一个函数逼近结构。而且，式(1)中参数 C_{ji}，W_j，θ_j 都可调，而调整 W_j，θ_j 就使得用于逼近的基函数相应调整。所以，相比一般逼近方法，神经网络所定义的函数逼近结构的优越性就在于它不但是一个对逼近系数寻优的过程，而且是一个对逼近基函数组自适应寻优的过程。当然，由于 ϕ_j 本身形式的限制，上述基函数的寻优是在一定的范围内进行的。但是，随着网络层数的增加，叠加的结果使基函数寻优的自由度增加了。所以，网络的非线性建模能力随着层数的增加而快速增长，这是一般逼近方法所不能比拟的。ϕ_j 的作用可从两方面看：一方面，它是逼近基函数；另一方面，在式(1)中它又是神经元的活化函数，所以其形式有一些不同于一般基函数的特点。神经网络函数逼近的另一个优势在于，一旦选定了网络结构，那么各逼近参数将通过统一的训练算法求得。这使得算法的适应性好，不依赖于数据。而由于其分布并行计算的特点又使其实时性好，这一点在神经网络的 VLSI 阵列实现中表现得尤为明显[4]。

由上可见，神经网络函数逼近提供了一种不同于传统理论的方法。这样，神经网络函数逼近的研究将前馈网络的研究与函数逼近的研究有机结合起来。从理论上研究神经网络的非线性逼近能力就是研究以前馈网络为代表的一类网络结构所定义的映射关系究竟对哪些非线性映射具有逼近能力，逼近的阶及精度是怎样的，与经典函数逼近相比有哪些特点、优点，能否建立与经典函数逼近中 Weierstrass 第一定理、Chebyshev 定理、Borel 定理以及 Jackson 定理[5]相应的结果。

神经网络逼近能力的研究与 ANN(人工神经网络)几十年曲折的发展密切相关。1987年，Hecht 等指出了 Kolmogrov1957年证明的多变量连续函数表示定理[6,7]与多层前馈网非线性逼近能力之间的关系[8]，这是第一次将前馈网络的映射能力与逼近理论相联系。虽然后来 Poggio 指出这两者实际上是无关的[9]，但是很快许多学者严格地证明了具有 S 型隐层函数的三层 BP 网可以任意精度一致逼近任意紧集上的连续函数。

Rumelhart&Mccleland 的 BP 网成功之处可归结到一点，就是引入了具有非线性传递函数的隐层单元。BP 网的各种变型以及其它各种函数逼近网络的引出也无不基于此原则。实际上，1991年，Kreinovich 得到了如下的一个结果：在非常微弱的限制下，在 ANN 的隐层引入任意的非线性都可使网络具有任意逼近连续函数的能力[10]。这一结果对函数逼近网络的实现具有十分重要的意义。在上述定理的指导下，就可以寻找多种函数充当隐层函数，使能量函数地貌简单，训练算法收敛快，从而有效解决 BP 算法的问题。子波神经网络的研究就是一个较为成功的例子[11]。

由于 ANN 在系统辨识以及复杂的模式分类问题中的应用，导致了关于 ANN 对于定义在 Banach 空间上的非线性泛函与非线性算子的逼近能力的研究，并得到了一系列具有指导意义的结果[2,12]。Tomaso Piggio 根据他与 Marr 所建立的计算机视觉中的正则理论建立了正则网络[13]。该网络包容了包括 RBF 网在内的一大类三层前向网络，并与模式识别中的 Parzen 窗技术和势函数理论密切相关，同时也启发人们到 BAYES 判决的框架下去讨论各种逼近方法的统一的统计解释。

神经网络泛化能力的研究也是神经网络函数逼近能力研究的另一个不可缺少的方面。这方面内容本文未涉及。

2 Mc-P 及 MLP 网逼近能力分析

2.1 多层 Mc-P 网

1943年 McCulloch&Pits 提出的 Mc-P 模型就已在网络中引入了简单的非线性 Heaviside 阶跃函数，由 Kreinovich[10]的结果知，Mc-P 网有可能具有较强的函数逼近能力。

Edward K. Blum 详细讨论了基于 Mc-P 模型的多层前传网的逼近能力[14]。

定义1 (简单函数) 设 E 是可测集，$f(x)$ 在 E 上只取有限多个实数值 c_1，c_2，…，c_n；且 $E(f=c_1)$，$E(f=c_2)$，…，$E(f=c_n)$ 均可测，则 f 称为 E 上的简单函数。

下述实变函数中的事实表明，简单函数空间在定义于同一有界可测集 E 上的 L^p 空间中稠密。

引理**1** 设 $f\in L^p$，则 $\forall\ \epsilon>0$，$\exists$ 有界可测函数 g，$s.\ t.\ \ f-g\ {}_p<\epsilon$。

引理**2** 设 E 是有界可测集，g 是 E 上的有界可测函数，则 $\forall\ \epsilon>0$，$\exists$ 简单函数 $\varphi(x)$，$s.\ t.\ \ g-\varphi_p<\epsilon$。

Blum 得到了如下引理：

引理**3** 令 f：$[a,\ b]\to R$ 是一简单函数，则 f 可用一个具有 Mc-P 隐层单元和线性输出单元的三层（输入也看做一层）网络来实现。

证明 因 $f\ (x)$ 是 $[a,\ b]$ 上的简单函数，故可将区间 $[a,\ b]$ 做如下划分：

$$a=b_0<b_1<\cdots<b_{n+1}=b,$$

也即将 $[a,\ b]$ 划分为如下半闭区间 $(b_i,\ b_{i+1}]$，$i=\overline{1,\ n}$，在这些区间上满足

$$f(x)=f_i,b_i<x<b_{i+1},f(a)=f_0,\text{则显然有：}$$

$$f(x)=\sum_{i=0}^{n}w_iH(x-b_i)+f_0H(x-b_{-1}),\text{这里},b_{-1}<b_0,\text{且}\ w_i=f_{i+1}-f_i$$

证毕。

由上述三引理易得如下定理[15]：

定理**1** (Hornik-Stinchcombe-White) 令 $f\in C\ [S]$，其中 S 是 R^n 上的紧集，则 $\forall\ \epsilon>0$，$\exists$ 一个具有 Mc-P 隐层单元和线性输出单元的三层网络，使其逼近误差按无穷范数 $_\infty$ 小于 ϵ。

进一步可有如下结果：

定理**2** 令 $f\in L^2[S]$，$S\in R^n$ 为紧集，则 $\forall\ \epsilon>0$，$\exists$ 一个三层 Mc-P 网 $s.\ t.$ 可按 $_2$ 范数使逼近误差小于 ϵ。

可见，Mc-P 网在紧集上的连续函数空间与 L^2 空间中稠密。如果不考虑算法的复杂性，那么，Mc-P 网是能一致最优逼近连续函数空间的最简单网络。

下述定理说明上述 Mc-P 网对于给定的 f：$[0,\ 1]^n\to R$，对于给定的逼近误差，所需 Mc-P 神经元的个数。

定理**3** 令 f：$[0,\ 1]^n\to R$，则为达到逼近精度 ϵ，所需 Mc-P 神经元的个数为 $m^n+2n\ (m-1)$。其中 m 如下确定：

若 f 满足，当 $x-y<\epsilon$ 时，$f\ (x)-f\ (y)<\omega(\epsilon)$，则 $m=[1/\omega(\epsilon)]$。

若 f 的 Lipschitz 指数为 L，则 $m\geq L/\epsilon$。

用 BP 网逼近的阶为 $O\ (n)$，而用 Mc-P 网逼近的阶为 $O\ (m^n)$。可见，Mc-P 网是用网络规模的复杂性代替了隐层单元的复杂性。

2.2 三层 MLP 网逼近能力分析与隐层传递函数的选择

2.2.1 函数逼近能力分析

对一个具有任意层的 MLP 网，若其各层传递函数都取为 $\sigma(\ )$，则其输入输出关系可表达为：

$$F(W,x)=\sigma(\sum_n w_n\sigma(\sum_j v_j\sigma(\cdots\sigma(\sum_i u_ix_i)\cdots)))$$

现在考虑一个三层 MLP 网，假设输出为线性单元并假设输出为一元的，则其所定义的函数映射关系的集合为：

$$\Lambda=\{f\in C[U]\ f(\bar{x})=\sum_{i=1}^{n}c_i\sigma(\bar{x}\ \overline{w_i}+\theta_i),U\in R^n,\overline{w_i}\in R^n,c_i,\theta_i\in R,n\in N\}$$

现在的问题归结为在 σ 满足怎样的条件下，Λ 在 $C\ (U)$ 中稠密。从 Λ 的表达式可见，三层 MLP 网所定义的映射关系实际上是一组仿射基函数所构成的线性空间。这与经典函数逼近论中的多项式基不同，而包含了子波基。

Hornik，Stinchcombe，White 以及 Cybenko 等人证明了如下三层网络逼近能力的定理[15,16]：

定理**4** 设 φ 为有界单调递增连续函数，$K\subset R^n$ 为紧集，固定层数 $k=3$，则对任何连续映射 f：$K\to R^n$ 可被一 k 层（$k-2$ 个隐层）网络一致逼近。该网络隐层传递函数为 φ，输入与输出为线性单元。

神经网络是通过对简单的非线性函数的复合完成这一映射的，只要经过少数几次复合即可得到极复杂的函数功能。

证明的途径大致有以下两种：一是利用经典函数逼近中的 Stone-Weierstrass 定理；二是利用积分变换。Yoshifusa Ito 利用了逆 Radon 变换[17]，Seimiyake 利用了 n 维 Fourier 变换[18]，然后再将上述积分变换化为有限和。

关于三层 MLP 网对 L^2 空间函数的逼近，有如下定理：

定理5 $\forall \epsilon > 0$，$f \in L^2[a, b]$，$\exists$ 一个三层 MLP 网，使逼近误差按 L^2 范数小于 ϵ。

上述逼近定理的证明都是存在性的，并没有给出构造网络的准则，特别是没有给出网络的规模(隐层单元的数目)与逼近误差间的关系。也就是说，已经建立了神经网络逼近的 Weierstrass 第一定理，但对应的 Jackson 定理是否存在还是个问题。

2.2.2 隐层传递函数的选择

隐层传递函数的选择首先决定着网络的逼近能力，其次决定着由此所导出的网络训练算法的复杂性。要找到一种隐层函数使网络逼近能力强、容易实现，同时又能导出性能优异的算法就需要对什么样的函数能充当隐层函数作深入研究。

Yoshifusa Ito 讨论了利用 Radon 变换来构造逼近网络的方法[17]，讨论了隐层函数为阶跃函数与 Sigmoid 型函数（关于 Sigmoid 型函数的定义见后面定义2）的逼近网络，证明其对连续函数的一致逼近能力。这就包容了前述 Mc-P 网及三层 MLP 网。

Horinik 证明[15]，任意的非常值有界连续函数都可充当隐层函数。Mhaskar&Michelli 证明，只要对连续函数在无穷远处的幅度做适当限制，任意非多项式函数都可充当隐层函数。Tian Ping Chen&Hong Chen 得到了如下具有一般意义的结果[2,12]。

定义2 若函数 $\varphi: R^1 \to R^1$ 满足：

$$\varphi(x) \to 0, x \to -\infty$$
$$\varphi(x) \to 1, x \to +\infty$$

则称 φ 为 Sigmoid 型函数。一般还对 Sigmoid 函数附加单调连续性的要求。

定义3 （Tauber-Wienner 函数）如果函数 $g: R^1 \to R^1$（连续或不连续）满足：所有如下的线性组合 $\sum_{i=1}^{N} c_i g(\lambda_i x + \theta_i)$，$\lambda_i$，$\theta_i$，$c_i \in R$，$i = 1, 2, \cdots, N$ 在 $C[a, b]$ 中稠密，则称 g 是一个 Tauber-Wienner 函数，Tauber-Wienner 函数的全体记为 (TW)。

定理6 如果 φ 是一个有界 Sigmoid 型函数，则 $\varphi \in (TW)$。

定理7 设 $K \subset R^n$ 为紧集，$U \subset C(K)$ 为紧集，$f \in C(K)$，$g \in (TW)$，则 $\forall \epsilon > 0$，$\exists N > 0$，$\theta_i \in R$，$W_i \in R^n$ 及常数 $c_i(f)$，$i = 1, 2, \cdots, N$，s.t. $\left| f(x) - \sum_{i=1}^{N} c_i(f) g(W_i x + \theta_i) \right| < \epsilon$。

定理7表明，一个函数 g 能充当网络隐层函数的充分条件是 $g \in (TW)$。而 (TW) 包含了 Sigmoid 函数类，从而大大扩展了可充当隐层函数的函数类范围。

如果在寻找最优权 W 的同时优化基函数 g，就要对 $E(W, g, \bar{x})$ 作双重变分。这是一个极为困难的问题。一条比较现实的途径是研究 (TW) 类中性态较好的几类函数，并导出相应的训练算法。

3 正则理论观点下的 RBF 网及其逼近能力分析

3.1 正则理论与 RBF 网

计算机视觉与模式识别中的许多问题都是所谓病态（ill-posed）问题。如 Marr&Piggio 的正则理论认为，视觉的初期阶段是光学成象的逆过程。由于在把三维世界投影为二维图像时丢失了大量信息，因而这个逆过程不存在唯一解，只有附加一些自然的约束才能有明确的输出。在模式识别中，对几个不同类模式所获得的数据往往是相对于其维数高度稀疏且带噪的，由这些数据本身并不足以给出各模式分布的可靠的估计，因而必须利用先验知识进行约束。实际上，正则化方法是以吉洪诺夫为代表的数学家，

为解决如病态方程组、弗雷德霍姆第一积分方程求解及函数的解析延拓等数学问题而提出来的。吉洪诺夫于1963年在抽象算子形式下给出了正则化方法, 其基本思想是构造一个连续的算子(正则化算子) 去逼近不连续算子。Poggio 后来将凡能把任何一种不适定问题转变为适定问题的方法统称为正则化方法, 并且在这方面作了许多工作。

Poggio 的正则化理论的主要思想是引入适当的先验知识以限制容许解的集合。

设 $S=\{(x_i, y_i)\in R^n\times R, i=1, \cdots, N\}$ 是一个已知数据集合, 现在拟用函数 f 逼近它。正则化方法就是确定 f 使如下泛函达最小: $H[f]=\sum_{i=1}^{N}[y_i-f(x_i)]^2+\lambda\|Pf\|^2$

式中, P 是约束算子, 它的构造要体现有关解的先验知识。所以, 它取决于待求解的具体问题的性质。Poggio 通过 Bayes 理论讨论了正则化理论的合理之处以及算子 P 是怎样体现着先验知识。

泛函 H 的最小化导出如下的 Euler-Lagrange 方程:

$$\hat{P}Pf(x)=\frac{1}{\lambda}\sum_{i=1}^{N}[y_i-f(x_i)]\delta(x-x_i)$$

$\hat{P}$ 为 P 的伴随算子。设算子 $\hat{P}P$ 的 Green 函数为 $G(x, y)$, 即

$\hat{P}PG(x,y)=\delta(x-x_i)$ 则 $f(x)$ 可解得为:

$$f(x)=\frac{1}{\lambda}\sum_{i=1}^{N}[y_i-f(x_i)]G(x-x_i)$$

若算子 P 具有平移不变性, 则 G 是一个径向基函数 (RBF), $G(x, x_i)=G(\|x-x_i\|)$, 则:

$$f(x)=\sum_{i=1}^{N}c_iG(\|x-x_i\|)$$

由此可构成广义径向基函数网络 (GRBF)。

上述讨论中没有考虑 Green 函数的线性项与常数项。若 $G(x)$ 取为 Gauss 函数, 就得到我们熟知的 RBF 网。

3.2 RBF 网的逼近能力

由上述正则理论导出的 RBF 网也是一类特殊的三层 MLP 网络, 也与起源于80年代 Powell 对所谓径向基函数插值的研究有关。设其非线性变换函数为 $\sigma(\ )$。则上述 RBF 网所表达的函数映射关系为:

$$\Delta=\{f\in C(U)\mid f(\bar{x})=\sum_{i=1}^{M}v_i\sigma(\|\bar{x}-\bar{c}_i\|\lambda_i), U, c_i\in R^n, v_i, \lambda_i\in R, M\in N\}$$

F. Girosi&T. Poggio 利用 Stone 形式的 Weierstrass 定理证明了一类 RBF 网具有与 MLP 网同样的逼近能力[19]。

定理8 (Stone) 设 $X\subset R^n$ 为紧集, 若 A 是 $C(X)$ 的一个子代数, 且 A 满足以下两条:

(1) $f(x)=1\in A$

(2) $\forall x\neq y\ x, y\in X, \exists f\in A\ s.t.\ f(x)\neq f(y)$

则 A 在 $C(X)$ 中稠密。

对于一维的情形, 若取 $\sigma(x)=\exp(-(x-c_i)^2\lambda)$, 容易验证 Stone 定理的条件是满足的, 则可知由此得到的集合 Δ 在 $C(X)$ 中稠密。这就是我们熟知的高斯径向基函数网络。

Chen Tian Ping 等证明[12], $\forall g\in C(R^1)\cap S(R^1)$, 只要它不是一个偶多项式, 就有 Δ 在 $C(U)$ 中稠密。这个结果对定义在紧集上的非线性连续算子同样成立。

3.3 与 MLP 网的比较

上面讨论了 MLP 网及 RBF 网对定义在紧集上的连续函数的逼近能力, 结论是两者的逼近能力是相近的。然而, F. Girosi&T. Poggio 指出, 当隐层单元数固定以后, Λ 与 Δ 有本质的不同[19]。具体说来就是, 若隐层单元数 $m\geq 2$ 则 Λ 不具有最佳逼近特性, 而对任意 $m\geq 1$ Δ 都具有最佳逼近特性, 且在某些特定范数下最佳逼近元是唯一的。这可简要说明如下。

对于 Λ，讨论 $\sigma(x)=\frac{1}{1+e^{-x}}$ 的情形。Λ 具有最佳逼近特性的一个必要条件是 Λ 是一个闭集。考虑如下函数：

$$f_\delta(\bar{x})=\frac{1}{\delta}\left(\frac{1}{1+e^{-[\bar{w}\bar{x}+\theta]}}-\frac{1}{1+e^{-[\bar{w}\bar{x}+(\theta+\delta)]}}\right)$$

显然，$\forall m\geq 2$，$f_\delta(\bar{x})\in\Lambda$，但是，

$$\lim_{\delta\to 0}f_\delta(\bar{x})\equiv g(\bar{x})=\frac{1}{2(1+\cosh[\bar{w}\bar{x}+\theta])}$$

$\forall m\geq 2$，$g(\bar{x})\notin\Lambda$，故 Λ 有一个不属于它的聚点，因而 Λ 非闭集。故 Λ 不具有最佳逼近特性。

MLP 网与 RBF 网在最佳逼近特性上的这一不同对于神经网络的实际实现有怎样的影响还不是很清楚。

4 MLP 网对 BAYES 判别边界的逼近

一般模式分类问题更适合在统计理论的框架下讨论。如果将 MLP 网的输入看做具有一定概率分布的随机矢量，则前传网可看做一个随机变换系统。那么，通过适当的监督学习算法的训练，它是否能学习到模式类之间的 BAYES 边界，条件是什么？

文献 [1]、[20] 证明，若神经网络训练采用 MSE（Mean Squared Error）目标函数且收敛到全局极小点，则网络产生基于样本分布的以 Hamming 测度为损失函数的 BAYES 判别边界。下面，较为形式化地阐述一下这个问题。

给定如下统计判别问题（Θ，X，A，P，ρ）。$\Theta=\{\theta\}$ 代表状态空间，$X=\{\bar{x}\}$ 代表观测空间，$A=\{\alpha\}$ 代表决策空间，P 代表 $A\times X$ 上的先验概率分布，ρ 是 $\Theta\times A\to R^+$ 上的一般损失函数。令 φ：$X\to A$ 表示判决函数，$\gamma[\varphi]$ 为相应风险。则

$$\gamma[\varphi]=\sum_{\theta\in\Theta}\sum_{x\in X}P(\theta,\bar{x})\rho(\theta,\varphi(\bar{x}))$$

按 BAYES 准则使风险 $\gamma(\varphi)$ 最小化的判别函数如下确定

$$\varphi(\bar{x})=\arg\min_{\alpha\in A}\sum_{\theta\in\Theta}W(\theta\,\bar{x})\rho(\theta,\alpha),\ \forall\bar{x}\in X$$

其中 $W(\theta\,\bar{x})$ 代表给定 $\bar{x}$ 后 θ 的后验概率分布。

假设 MLP 网按如下误差函数进行训练：

$$E=\frac{1}{m}\sum_{i=1}^{m}f(\theta^i,Z(\bar{x},w))$$

这里，f 是 MLP 网的目标函数。$Z(\bar{x},w)$ 代表网络输出，w 为权值。每一 $\alpha\in A$ 都用一 $|A|$ 维的矢量表示为 $t_\alpha=(0,\cdots,\underset{\alpha th}{1},0,\cdots,0)$ 则有如下结果[20, 21]：

定理**9** 若 MLP 网的目标函数取为 $f(\theta,Z)=\sum_{\alpha\in A}(\rho_{\max}-\rho(\theta,\alpha))\|t_\alpha-Z(\bar{x},w)\|^2$，若 MLP 网的结构使其有足够的泛函逼近能力，且算法保证其收敛到全局极小点，则该网络经训练后可产生使经验风险最小的 BAYES 判别边界。

文献 [20] 还在该框架下讨论了 MLP 网的泛化能力。可见，将前传网的逼近能力的分析纳入统计分析的框架不但可以更清楚地看清其信息处理的实质，而且还可以更深入地分析其用于模式分类的品质因素。

5 结语

从数学上来看，神经网络函数逼近代表了非线性函数的一类仿射展开，其特殊优点是仿射基的选取自由度大，展开系数可由统一的训练算法获得，相对于传统算法稳定性好。可以说，神经网络用于函数逼近为函数逼近理论的研究开辟了一个崭新的方向。值得一提的是，模糊神经网络的函数逼近也有类似的理论，并且具有自己的特色[22~25]，是一个值得注意的研究方向。

参考文献

1 Wan E. Neural Network Classicification: A Bayesian Interpretation. IEEE Trans. N. N, 1990, 1: 303

2 TianPing Cher&Hong Chen. Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamical System. IEEE Trans. 1995

3 Ergezinger S. &Thomsen E. An Acceleated Learning Algorithm for Multilayer Perceptrons: Optimization Layer by Layer. IEEE Trans. 1995

4 贡三元. VLSI 阵列处理. 南京: 东南大学出版社, 1993

5 徐利治. 函数逼近的理论与方法. 上海科技出版社, 1983

6 Kolmogrov A N. On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition. Dokl. Akad. Nauk, SSR. English Translation, 1963

7 洛伦茨. 函数逼近论. 上海科技出版社, 1981

8 Hecht_Nielsen R. Kolmogrov's Mapping Neural Network Existence Theorem. IEEE First International Conference on Neural Networks

9 Girosi F. &Poggio T. Representation Properties of networks: Kolmogrov's theorem is irrelevent, Neural Computation, 1989, 1: 465

10 Kreinorich V Y. Arbitrary Nonlineartity Is Sufficient to Represent all Functions by Neural Networks: A Theorem. N. N. 1991, 4: 381

11 Qinghua Zhang&Albert Benveniste. Wavelet Network. IEEE Trans. N. N1992, 3: 889

12 TianPing Chen&Hong Chen. Approximation Capability to Functions of Several Variables, Nonlinear Functionals, and Operators by Radial Basis Runction Neural Networks. IEEE Trans. N. N., 1995

13 Tomaso Poggio&Federico Girosi. Networks for Approxication and learning. Procs. IEEE, 1990

14 Edward K Blum. Approximation Theory and Feedforward Networks. N. N, 1991, 4: 511

15 Hormik K, Stincheomke M. and White H. Multilayer Feedforward Networks are Universal Approximators. 1989, 2: 359

16 Cybenko. Approximation by superposition of a sigmoidal function. Math. Control Systems Signals, 1989

17 Yoshifusa Ito. Representtation of Functions by Superposition of a Step or Sigmoid Function and Their Application to Neural Networks Theory. 1991, 4: 385

18 Bumpei Irie&Sei Miyake. Capabilities of Three_layered Perceptrons. 2nd IJCNN, IEEE, 1988

19 Girosi F&Poggio T. Neural Networks and the Best Approximation Property. Bol. Cyber. 1990, 63: 169

20 Levin E. et al. A Statistical Approach to Learning and Generalization in Layered Neural networks. Procs. of IEEE, 1990, 78: 1568

21 Amari et al. Four Ytpes of Learning Curves Neural computation. 1992, 4: 605

22 Bart Kosko. Fuzzy System as Universal Approximators. Procs. of IEEE Fuzz- 92. 1992

23 Kosko B. Fuzzy Function Approximation. IEEE Int. Conf. N. N. 1992, 1: 209

24 Kosko B. Neural Networks and Fuzzy systems. Prentice Hall

25 Gail A Carpenter & Stphen Grossberg. Fuzzy ART MAP: A Neural Network Architecture for Incermental Supervised Learingin of Analog Multidimensional Maps. IEEE Trans. N. N. 1992, 3: 699～713