系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
29(5) (2009, 5), 670–676

距离空间中插值神经网络的误差估计*

徐士英 曹飞龙

(中国计量学院理学院信息与数学科学系，杭州 310018)

摘要 研究距离空间中的神经网络插值与逼近问题．首先引进一类广义的激活函数，用比较简洁的方法讨论距离空间中插值神经网络的存在性，然后给出插值神经网络逼近连续函数的误差估计．

关键词 神经网络，插值，逼近，误差估计．

MR(2000) 主题分类号 41A05, 41A63

1 引 言

设 (X,ρ) 是一个距离空间， ρ 是其距离， $S=\{x_0,x_1,\cdots,x_n\}$ 是 X 中的 $n+1$ 个互异的点 (X 中的插值节点), $\{y_i:i=0,1,\cdots,n\}\subset R$, 我们称

$$\{(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)\} \tag{1}$$

为一插值样本．若 $f:X\to R$, 且满足 $f(x_i)=y_i(i=0,1,\cdots,n)$, 那么我们称 f 为插值样本 (1) 的一个插值泛函．如果存在前向神经网络 $N(x)$ 满足条件

$$N(x_j)=y_j,\quad j=0,1,\cdots,n,$$

则说 $N(x)$ 是样本 (1) 的一个精确插值网络. 如果对任意给定正数 ε, 存在前向神经网络 $N_a(x)$, 使得

$$|N_a(x_j)-y_j|<\varepsilon,\quad j=0,1,\cdots,n,$$

则称 $N_a(x)$ 是关于样本 (1) 的一个近似插值网络.

一个 R 上定义的有界函数 σ 被称为 Sigmoidal 函数，如果它满足条件

$$\lim_{x\to-\infty}\sigma(x)=0,\quad \lim_{x\to+\infty}\sigma(x)=1.$$

近年来，神经网络逼近与插值问题一直是神经网络理论与应用的研究热点之一，倍受国内外学者的关注 (参见 [1–8]). Shrivatava 与 Dasgupta[9] 在激活函数 $\phi(x)$ 为对数 Sigmoidal 函数 $S(x)=\frac{1}{1+\mathrm{e}^{-x}}$ 时，给出了精确插值网络存在性的代数证明． Ito 与 Saito[10] 证明了当激活函数 ϕ 是非减的 Sigmoidal 函数时，精确插值网络是存在的．然而，具体求解精确插值

* 国家自然科学基金 (60873206) 与浙江省自然科学基金 (Y7080235) 资助课题.
收稿日期：2007-05-15, 收到修改稿日期：2008-10-08.

网络的计算量是比较大的．于是，人们转向寻求近似插值网络的研究， Sontag 在 [11] 中曾讨论过这个问题．最近， Llanas 与 Sainz[12] 在激活函数为非减的 Sigmoidal 函数的条件下，给出三层前向精确插值网络存在性的代数证明，文献 [13] 对于一般的 Sigmoidal 激活函数和 d 维 Euclid 空间中的插值样本，分别构造了精确插值和近似插值的单隐层前向神经网络，并分别估计它们对连续函数的逼近误差，指出神经网络插值与一般代数多项式插值的本质差异．文献 [7] 在一般的距离空间中研究网络的插值与逼近问题：采用不同于文献 [14] 的方法，通过引进一个新的激活函数 $g_j(x)$, 在距离空间中构造精确插值与近似插值网络，并用之逼近距离空间的连续泛函．

假定 K 是 X 中的紧集, 对 $\delta>0$, 记 $s(x,\delta)=\{y:y\in X,\rho(y,x)<\delta\}$, 由 $\bigcup\limits_{x\in K}s(x,\delta)\supset K$ 及 K 的紧性知，对于 K 存在有限覆盖，即有 $\{x_0,x_1,\cdots,x_n\}\subset K$ 满足

$$\bigcup_{i=0}^{n}s(x_i,\delta)\supset K.$$

令 $\varphi(t):[0,+\infty)\to R$, 单调减少，且满足 $\varphi(0)=1$, $\lim\limits_{t\to+\infty}\varphi(t)=0$. 对于 $A>0$, 以及节点 $\{x_0,x_1,\cdots,x_n\}$, 记

$$g_i(x,A)=\frac{\varphi(A\rho(x,x_i))}{\sum\limits_{j=0}^{n}\varphi(A\rho(x,x_j))},\quad i=0,1,\cdots,n,\quad x\in X.$$

作关于 $g_j(x,A)$ 的线性组合

$$N(x)=\sum_{j=0}^{n}c_jg_j(x,A), \tag{2}$$

则 $N(x)$ 可以理解为一个 4 层前向神经网络的模型：第 1 层为输入层，输入为 $x(x\in X)$, 第 2 层为一个预处理层，将 x 变为 $\rho(x,x_j)$, 计算输入向量与插值结点的距离，且作为第 3 层的输入，第 3 层具有 $n+1$ 个神经元，第 j 个神经元的激活函数为 $g_j(x,A)$, 第 4 层为输出层，输出量为 $N(x)$.

当 $\varphi(t)=\mathrm{e}^{-t}$ 时，得到的神经网络即为文献 [15] 中构造的神经网络，文献 [7] 讨论了此时插值神经网络的存在性及其在 $C(K)$ 中的稠密性，本文对满足上述条件的一般 $\varphi(t)$ 进行讨论，得到插值神经网络的存在性，并给出误差估计，即证得下列定理．

定理 1.1 设 (X,ρ) 是距离空间， K 是 X 的紧子集， $\{x_0,x_1,\cdots,x_n\}\subset K$ 为插值节点， $d=\min\limits_{1\le i,j\le n,i\ne j}\rho(x_i,x_j)$, A^* 满足 $0<\varphi(A^*d)<\frac{1}{n}$, 则对定义在 $[0,+\infty)\to R$ 上单调减少且满足 $\varphi(0)=1$, $\lim\limits_{t\to+\infty}\varphi(t)=0$ 的函数 $\varphi(t)$, $A>A^*$ 及 $f\in C(K)$, 形如 (2) 的精确插值神经网络存在．

定理 1.2 设 (X,ρ) 是距离空间， K 是 X 的紧子集， $S=\{x_0,x_1,\cdots,x_n\}\subset K$ 为插值节点，且 $\bigcup\limits_{i=0}^{n}s(x_i,\delta)\supset K$, $d=\min\limits_{1\le i,j\le n,i\ne j}\rho(x_i,x_j)$, A^* 满足 $0<\varphi(A^*d)<\frac{1}{2n}$, 则对定义在 $[0,+\infty)\to R$ 上单调减少且满足 $\varphi(0)=1$, $\lim\limits_{t\to+\infty}\varphi(t)=0$ 的函数 $\varphi(t)$, $A>A^*$, 以 S 为插值节点形如 (2) 的插值神经网络 $N(x)$ 对 $f(x)\in C(K)$, 有如下误差估计

$$\|N(x)-f(x)\|_\infty\le\frac{2n\varphi(Ad)}{1-2n\varphi(Ad)}\|f\|_\infty+2n\|f(x)\|_\infty\frac{\varphi(2A\delta)}{\varphi(A\delta)}+\omega(f,2\delta),$$

其中 $\|f\|_\infty=\max\limits_{0\le i\le n}|f(x_i)|$, $\|f(x)\|_\infty=\max\limits_{x\in K}|f(x)|$, 而

$$\omega(f,\delta)=\sup_{\substack{x,y\in K,\\ \rho(x,y)\le\delta}}|f(x)-f(y)|$$

是函数 f 的连续模.

由定理 1.2 可知, 当激活函数 $\varphi(t)$ 满足 $\lim\limits_{t\to+\infty}\frac{\varphi(2t)}{\varphi(t)}=0$ 时, 对 $f(x)\in C(K)$ 及任意的 $\varepsilon>0$, 存在插值节点 $\{x_i\}_{i=0}^n$ 及 A^*, 当 $A>A^*$ 时, 对应的插值神经网络满足 $\|N(x)-f(x)\|<\varepsilon$. 即插值神经网络在 $C(K)$ 中稠.

事实上, $\forall\varepsilon>0$, 因为 $f(x)$ 为紧集 K 上的连续泛函, 所以存在 $\delta>0$, 使得 $\omega(f,2\delta)<\frac{\varepsilon}{3}$. 对此 δ, 由于 K 紧, 故 K 中存在有限 δ 网 $\{x_i\}_{i=0}^n$, 即 $\bigcup\limits_{i=0}^n s(x_i,\delta)\supset K$. 取 $\{x_i\}_{i=0}^n$ 为插值节点, 则 n,d,δ 确定, 当 $A\to+\infty$ 时, 定理 1.2 估计式中第一项、第二项均趋向于 0, 故存在 A^*, 当 $A>A^*$ 时, 有

$$\frac{2n\varphi(Ad)}{1-2n\varphi(Ad)}\|f\|_\infty<\frac{\varepsilon}{3},\quad 2n\|f(x)\|_\infty\frac{\varphi(2A\delta)}{\varphi(A\delta)}<\frac{\varepsilon}{3}.$$

从而 $\|N(x)-f(x)\|<\varepsilon$.

特别地, 当 $\varphi(t)=\mathrm{e}^{-t}$ 时, 我们有

推论 1.1 当 $A>\frac{\ln 2n}{d}$ 时, 精确插值神经网络存在, 且有误差估计

$$\|N(x)-f(x)\|_\infty\le\frac{2n\mathrm{e}^{-Ad}}{1-2n\mathrm{e}^{-Ad}}\|f\|_\infty+2n\mathrm{e}^{-A\delta}\|f(x)\|_\infty+\omega(f,2\delta).$$

2 定理的证明

定理 1.1 的证明 只须证明线性方程组

$$N(x_i)=\sum_{i=0}^n c_i\frac{\varphi(A\rho(x_i,x_j))}{\sum\limits_{j=0}^n\varphi(A\rho(x_i,x_j))}=f(x_i),\quad i=0,1,2,\cdots,n$$

有解, 即只须证对应系数矩阵

$$\boldsymbol{M}=\begin{bmatrix} m_{00} & m_{01} & \cdots & m_{0n}\\ m_{10} & m_{11} & \cdots & m_{1n}\\ \vdots & \vdots & \cdots & \vdots\\ m_{n0} & m_{n1} & \cdots & m_{nn}\end{bmatrix}$$

$$=\begin{bmatrix} \frac{1}{1+\sum\limits_{i=1}^n\varphi(A\rho(x_0,x_i))} & \frac{\varphi(A\rho(x_0,x_1))}{1+\sum\limits_{i=1}^n\varphi(A\rho(x_0,x_i))} & \cdots & \frac{\varphi(A\rho(x_0,x_n))}{1+\sum\limits_{i=1}^n\varphi(A\rho(x_0,x_i))}\\ \frac{\varphi(A\rho(x_1,x_0))}{1+\sum\limits_{i=0,i\ne1}^n\varphi(A\rho(x_1,x_i))} & \frac{1}{1+\sum\limits_{i=0,i\ne1}^n\varphi(A\rho(x_1,x_i))} & \cdots & \frac{\varphi(A\rho(x_1,x_n))}{1+\sum\limits_{i=0,i\ne1}^n\varphi(A\rho(x_1,x_i))}\\ \vdots & \vdots & \cdots & \vdots\\ \frac{\varphi(A\rho(x_n,x_0))}{1+\sum\limits_{i=0}^{n-1}\varphi(A\rho(x_n,x_i))} & \frac{\varphi(A\rho(x_n,x_1))}{1+\sum\limits_{i=0}^{n-1}\varphi(A\rho(x_n,x_i))} & \cdots & \frac{1}{1+\sum\limits_{i=0}^{n-1}\varphi(A\rho(x_n,x_i))}\end{bmatrix}$$

可逆.

由于

$$m_{ii}=\frac{1}{1+\sum\limits_{j=0,j\neq i}^{n}\varphi(A\rho(x_j,x_i))}\geq\frac{1}{1+n\varphi(Ad)}\geq\frac{1}{1+n\varphi(A^*d)}>\frac{1}{2},$$

$$\sum_{j=0,j\neq i}^{n}m_{ij}=\frac{\sum\limits_{j\neq i}\varphi(A\rho(x_j,x_i))}{1+\sum\limits_{j\neq i}\varphi(A\rho(x_j,x_i))}=1-m_{ii}<\frac{1}{2},$$

所以，$m_{ii}-\sum\limits_{j\neq i}m_{ij}>0$, 即矩阵 $\boldsymbol{M}$ 对角占优，于是 $\boldsymbol{M}$ 可逆，即线性方程组有解，从而，插值神经网络存在.

定理 1.2 的证明 为行文方便，记

$$g_i(x,A)=\frac{\varphi(A\rho(x,x_i))}{\sum\limits_{j=0}^{n}\varphi(A\rho(x,x_j))},\quad f_i=f(x_i),$$

则对 $x\in K$, 有

$$\begin{aligned}N(x)-f(x)&=\sum_{i=0}^{n}c_ig_i(x,A)-f(x)=\sum_{i=0}^{n}(c_i-f(x))g_i(x,A)\\&=\sum_{i=0}^{n}(c_i-f_i)g_i(x,A)+\sum_{i=0}^{n}(f_i-f(x))g_i(x,A)\\&=\sum_{i=0}^{n}(c_i-f_i)g_i(x,A)+\sum_{x_i\notin s(x,2\delta)}(f_i-f(x))g_i(x,A)\\&\quad+\sum_{x_i\in s(x,2\delta)}(f_i-f(x))g_i(x,A)=I_1+I_2+I_3.\end{aligned}$$

设

$$\boldsymbol{f}=(f_0,f_1,\cdots,f_n)^{\mathrm{T}},\quad \boldsymbol{C}=(c_0,c_1,\cdots,c_n)^{\mathrm{T}},$$

则由

$$\boldsymbol{MC}=\boldsymbol{f}\quad 及\quad \boldsymbol{If}=\boldsymbol{f},$$

有

$$\boldsymbol{C}-\boldsymbol{f}=(\boldsymbol{M}^{-1}-\boldsymbol{I})\boldsymbol{f}.$$

注意到

$$|m_{ii}-1|=|1-g_i(x_i,A)|=\frac{\sum\limits_{j\neq i}\varphi(A\rho(x_j,x_i))}{1+\sum\limits_{j\neq i}\varphi(A\rho(x_j,x_i))}\leq n\varphi(Ad)$$

与

$$|m_{ij}|=\frac{\varphi(A\rho(x_j,x_i))}{1+\sum\limits_{j\neq i}\varphi(A\rho(x_j,x_i))}\leq\varphi(Ad),\quad j\neq i.$$

所以

$$\|\boldsymbol{M}-\boldsymbol{I}\|_\infty=\max_{0\le i\le n}\Big(\sum_{j\ne i}|m_{ij}|+|m_{ii}-1|\Big)\le 2n\varphi(Ad)\le 2n\varphi(A^*d)<1.$$

故利用

$$\boldsymbol{M}^{-1}=[\boldsymbol{I}+\boldsymbol{M}-\boldsymbol{I}]^{-1}=\boldsymbol{I}+(\boldsymbol{I}-\boldsymbol{M})+(\boldsymbol{I}-\boldsymbol{M})^2+\cdots+(\boldsymbol{I}-\boldsymbol{M})^n+\cdots,$$

得到

$$\begin{aligned}\|\boldsymbol{M}^{-1}-\boldsymbol{I}\|_\infty&\le\|\boldsymbol{I}-\boldsymbol{M}\|_\infty+\|\boldsymbol{I}-\boldsymbol{M}\|_\infty^2+\cdots+\|\boldsymbol{I}-\boldsymbol{M}\|_\infty^n+\cdots\\&=\frac{\|\boldsymbol{I}-\boldsymbol{M}\|_\infty}{1-\|\boldsymbol{I}-\boldsymbol{M}\|_\infty}\le\frac{2n\varphi(Ad)}{1-2n\varphi(Ad)},\end{aligned}$$

所以

$$\|\boldsymbol{C}-\boldsymbol{f}\|_\infty\le\|\boldsymbol{M}^{-1}-\boldsymbol{I}\|_\infty\|\boldsymbol{f}\|_\infty\le\frac{2n\varphi(Ad)}{1-2n\varphi(Ad)}\|\boldsymbol{f}\|_\infty.$$

从而

$$|I_1|\le\sum_{i=0}^n|c_i-f_i|g_i(x,A)\le\|\boldsymbol{C}-\boldsymbol{f}\|_\infty\le\frac{2n\varphi(Ad)}{1-2n\varphi(Ad)}\|\boldsymbol{f}\|_\infty.$$

注意到当 $x_i\notin s(x,2\delta)$ 时，由 $\bigcup\limits_{i=0}^n s(x_i,\delta)\supset K$ 知，存在 i_0 使 $x\in s(x_{i_0},\delta)$, 且有

$$g_i(x,A)=\frac{\varphi(A\rho(x,x_i))}{\sum\limits_{j=0}^n\varphi(A\rho(x,x_j))}\le\frac{\varphi(A2\delta)}{\varphi(A\rho(x,x_{i_0}))}\le\frac{\varphi(2A\delta)}{\varphi(A\delta)},$$

所以

$$|I_2|\le\sum_{x_i\notin s(x,2\delta)}|f_i-f(x)|g_i(x,A)\le 2\|f(x)\|_\infty\sum_{x_i\notin s(x,2\delta)}g_i(x,A)\le 2n\|f(x)\|_\infty\frac{\varphi(2A\delta)}{\varphi(A\delta)},$$

$$|I_3|\le\sum_{x_i\in s(x,2\delta)}|f_i-f(x)|g_i(x,A)\le\omega(f,2\delta)\sum_{x_i\in s(x,2\delta)}g_i(x,A)\le\omega(f,2\delta).$$

综上，有

$$|N(x)-f(x)|\le\frac{2n\varphi(Ad)}{1-2n\varphi(Ad)}\|f\|_\infty+2n\frac{\varphi(2A\delta)}{\varphi(A\delta)}\|f(x)\|_\infty+\omega(f,2\delta).$$

由于 x 在 K 中任取，故

$$\|N(x)-f(x)\|_\infty\le\frac{2n\varphi(Ad)}{1-2n\varphi(Ad)}\|f\|_\infty+2n\frac{\varphi(2A\delta)}{\varphi(A\delta)}\|f(x)\|_\infty+\omega(f,\delta).$$

定理 1.2 证毕.

3 有限维插值

设 $X=R^m$, 取

$$\rho(x,y)=\|x-y\|_2=\Big(\sum_{i=1}^{m}(x_i-y_i)^2\Big)^{\frac{1}{2}},\quad K=\overbrace{[0,1]\times[0,1]\times\cdots\times[0,1]}^{m\text{ 个}}.$$

将 m 维单位正方体每边 n 等分，共分成 n^m 个小 m 维正方体，以这些小正方体的中心为插值节点，即 $S=\{x_0,x_1,\cdots,x_{n^m-1}\}$. 由于

$$\bigcup_{i=0}^{n^m-1}s\Big(x_i,\frac{\sqrt{m}}{2n}+\varepsilon\Big)\supset K,$$

此时有 $\delta=\frac{\sqrt{m}}{2n}+\varepsilon$ 及 $d=\frac{1}{n}$. 应用定理 1.2, 并假设 $\varphi(t)$ 右连续，令 $\varepsilon\to 0$, 可得

推论 2.2 当 $A>A^*$ 时，其中 A^* 满足 $\varphi\big(\frac{A^*}{n}\big)\le\frac{1}{2n}$, 对上述插值神经网络有误差估计

$$\|N(x)-f(x)\|_\infty\le\frac{2n^m\varphi\big(\frac{A}{n}\big)}{1-2n^m\varphi\big(\frac{A}{n}\big)}\|f\|_\infty+2n^m\frac{\varphi\big(\frac{\sqrt{m}A}{n}\big)}{\varphi\big(\frac{\sqrt{m}A}{2n}\big)}\|f(x)\|_\infty+\omega\Big(f,\frac{\sqrt{m}}{n}\Big).$$

当 $\varphi(t)=\mathrm{e}^{-t}, A>n\ln 2n$ 时，有误差估计

$$\|N(x)-f(x)\|_\infty\le\frac{2n^m\mathrm{e}^{-\frac{A}{n}}}{1-2n^m\mathrm{e}^{-\frac{A}{n}}}\|f\|_\infty+2n^m\mathrm{e}^{-\frac{\sqrt{m}A}{2n}}\|f(x)\|_\infty+\omega\Big(f,\frac{\sqrt{m}}{n}\Big).$$

若对 $X=R^m$, 取

$$\rho(x,y)=\|x-y\|_\infty=\max_{1\le i\le m}|x_i-y_i|,$$

并且 K 与插值节点的取法与推论 2.2 相同，则有 $\delta=\frac{1}{2n}+\varepsilon, d=\frac{1}{n}$, 可得插值神经网络的误差估计

$$\|N(x)-f(x)\|_\infty\le\frac{2n^m\varphi\big(\frac{A}{n}\big)}{1-2n^m\varphi\big(\frac{A}{n}\big)}\|f\|_\infty+2n^m\|f(x)\|_\infty\frac{\varphi\big(\frac{A}{n}\big)}{\varphi\big(\frac{A}{2n}\big)}+\omega\Big(f,\frac{1}{n}\Big).$$

参 考 文 献

[1] Cybenko G. Approximation by superpositions of a single function. *Mathematics of Control, Signals and Systems*, 1989, **2**: 303–314.

[2] Chen T P, Chen H, Liu R W. Approximation capability in $C(R^n)$ by multilayer feedforward networks and related problems. *IEEE Trans. Neural Networks*, 1995, **6**: 25–30.

[3] 陈天平. 神经网络及其在系统识别应用中的逼近问题. 中国科学 (A 辑), 1994, **24**: 1–7.

[4] 陈天平. 用 Sigmodial 函数的叠合逼近 Hilbert 空间中的连续泛函. 科学通报, 1992, **37**(13): 1167–1169.
[5] 蒋传海, 陈天平. $W_2^m(R^n)$ 中单个函数的平移和伸缩组合的稠密性. 数学学报, 1999, **42**(3): 495–500.
[6] 曹飞龙, 徐宗本. 神经网络的本质逼近阶. 中国科学 (E), 2004, **34**(4): 361–373.
[7] 曹飞龙, 张永全. 距离空间中的神经网络插值与逼近. 数学学报, 2008, **51**(1): 91–98.
[8] Xu Z B, Cao F L. Simultaneous L^p-approximation order for neural networks. *Neural Networks*, 2005, **18**: 914–923.
[9] Shrivatava Y, Dasgupta S. Neural networks for exact matching of functions on a discrete domain. Proceedings of the 29th IEEE Conference on Decision and Control, Honolulu, 1990, 1719–1724.
[10] Ito Y, Saito K. Superposition of linearly independent functions and finite mappings by neural networks. *Math. Sci.*, 1996, **21**: 27–33.
[11] Sontag E D. Feedforward nets for interpolation and classification. *J. Comp. Syst. Sci.*, 1992, **45**: 20–48.
[12] 谢庭藩, 曹飞龙. 关于插值神经网络的构造性. 自然科学进展. 2008, **18**(3): 334–340.
[14] Llanas B, Sainz F J. Constructive approximate interpolation by neural networks. *J. Comput. Applied Math.*, 2006, **188**: 283–308.
[15] Courrieu P. Two methods for encoding clusters. *Neural Networks*, 2001, **14**: 175–183.

ESTIMATIONS OF ERROR FOR INTERPOLATION NEURAL NETWORKS IN DISTANCE SPACES

XU Shiying CAO Feilong

(*Department of Information and Mathematics Sciences, China Jiliang University, Hangzhou* 310018)

Abstract This paper deals with the interpolation and approximation of neural networks in distance spaces. First a class of generalized activation function and is introduced a simple method is used to prove the existence of interpolation networks in the distance spaces. And then the estimates of error for the interpolation neural networks approximating continuous function is given.

Key words Neural networks, interpolation, approximation, estimations of error.