第51卷第1期 数学学报中文版 Vol.51, No.1
2008年1月 ACTA MATHEMATICA SINICA, Chinese Series Jan., 2008

文章编号: 0583-1431(2008)01-0091-08 **文献标识码: A**

距离空间中的神经网络插值与逼近

曹飞龙 张永全

中国计量学院信息与数学科学系 杭州 310018
E-mail: flcao@263.net

摘 要 已有的关于插值神经网络的研究大多是在欧氏空间中进行的, 但实际应用中的许多问题往往需要用非欧氏尺度进行度量. 本文研究一般距离空间中的神经网络插值与逼近问题, 即先在距离空间中构造新的插值网络, 然后在此基础上构造近似插值网络, 最后研究近似插值网络对连续泛函的逼近.

关键词 神经网络; 插值; 连续泛函; 逼近
MR(2000) 主题分类 41A05, 41A63
中图分类 O174.41

Interpolation and Approximation by Neural Networks in Distance Space

Fei Long CAO Yong Quan ZHANG

Department of Information and Mathematics Sciences, China Jiliang University, Hangzhou 310018*, P. R. China*
E-mail: flcao@263.net

Abstract There have been many studies on interpolation by neural networks in Euclidean spaces. However, there are plentiful concrete problems which must be measured by using non-Euclidean metrics. This paper deals with the interpolation and approximation of neural networks in general non-Euclidean spaces. That is, first construct interpolation networks in general spaces, and then construct approximate interpolation networks, finally, study the approximation of continuous functional with approximating network.

Keywords neural networks; interpolation; continuous functional; approximation
MR(2000) Subject Classification 41A05, 41A63
中图分类 O174.41

1 引言

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

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

收稿日期: 2006-05-12; 接受日期: 2007-07-06
基金项目: 国家自然科学基金资助项目 (60473034); 浙江省教育厅科研重点项目基金 (20060543)
通讯作者: 曹飞龙, 杭州市下沙高教园区学源街中国计量学院信息与数学科学系

为一插值样本. 若 $f: X \to \mathbb{R}$, 且满足 $f(x_i) = y_i$ $(i = 0, 1, \ldots, n)$, 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 $N_e(x)$ 满足条件

$$N_e(x_j) = y_j, \quad j = 0, 1, \ldots, n,$$

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

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

则称 $N_a(x)$ 是关于样本 (1) 的一个近似插值网络. 一个 $\mathbb{R}$ 上定义的有界函数 σ 被称为 Sigmoidal 函数, 如果它满足条件

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

由于神经网络在众多应用中的许多问题均可以转化为神经网络的逼近问题, 所以, 近年来, 人们一直关注神经网络逼近问题的研究 (参见文 [1–11]). 另一方面, 关于神经网络的插值问题一直是神经网络理论与应用的研究热点之一. Shrivatava 与 Dasgupta [12] 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 [13] 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 在文 [14] 中曾讨论过这个问题. 最近, Llanas 与 Sainz [15] 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, Xie 与 Cao [16] 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. 但已有的关于网络插值的研究大多是在欧氏空间 $\mathbb{R}^d$ 中进行的, 尽管欧氏空间是一类常用的距离空间, 但在实际中, 我们经常会遇到一些非欧氏空间的神经网络插值与逼近问题, 并且目标函数往往是以泛函形式给出的. 因此, 我们有必要在一般的距离空间中研究神经网络的插值与逼近问题. 本文采用不同于文 [15] 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中的连续泛函.

对于距离空间 (X, ρ) 中的节点 $S = \{x_0, x_1, \ldots, x_n\}$, $A > 0$, 令

$$g_j(x, A) = \frac{\mathrm{e}^{-A\rho(x, x_j)}}{\sum_{i=0}^{n} \mathrm{e}^{-A\rho(x, x_i)}}, \quad j = 0, 1, \ldots, n, \quad x \in X.$$

作关于 $g_j(x, A)$ 的线性组合

$$N(x) = \sum_{j=0}^{n} c_j g_j(x, A),$$

则 $N(x)$ 可以理解为一个四层前向神经网络的模型: 第一层为输入层, 输入为 $x (x \in X)$, 第二层为一个预处理层, 将 x 变为 $\rho(x, x_j)$, 计算输入向量与插值结点的值, 且作为第三层的输入, 第三层具有 $n+1$ 个神经元, 第 j 个神经元的激活函数为 $g_j(x, A)$, 第四层为输出层, 输出量为 $N(x)$.

2 插值网络的构造

令 K 为 X 上的紧集, $\{x_0, x_1, \ldots, x_n\} \subset K$, 我们试图从集合 $N_{n+1} = \{N(x) : N(x) = c_j g_j(x, A), c_j \in \mathbb{R}\}$ 中找一网络 $N(x)$, 对于插值样本 (1), 满足

$$N(x_i) = y_i, \quad (i = 0, 1, \ldots, n), \tag{2}$$

即寻找出一函数

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

使得

$$\sum_{j=0}^{n}c_jg_j(x_i,A)=y_i,\quad i=0,1,\dots,n,$$

或者

$$\mathbf{MC}=\mathbf{Y}, \tag{3}$$

其中

$$\mathbf{M}=\begin{bmatrix} g_0(x_0,A) & g_1(x_0,A) & \cdots & g_n(x_0,A)\\ g_0(x_1,A) & g_1(x_1,A) & \cdots & g_n(x_1,A)\\ \vdots & \vdots & \cdots & \vdots\\ g_0(x_n,A) & g_1(x_n,A) & \cdots & g_n(x_n,A)\end{bmatrix}$$

$$=\begin{bmatrix} \dfrac{1}{1+\sum_{i=1}^{n}\mathrm{e}^{-A\rho(x_0,x_i)}} & \dfrac{\mathrm{e}^{-A\rho(x_0,x_1)}}{1+\sum_{i=1}^{n}\mathrm{e}^{-A\rho(x_0,x_i)}} & \cdots & \dfrac{\mathrm{e}^{-A\rho(x_0,x_n)}}{1+\sum_{i=1}^{n}\mathrm{e}^{-A\rho(x_0,x_i)}}\\ \dfrac{\mathrm{e}^{-A\rho(x_1,x_0)}}{1+\sum_{i=0,\,i\neq1}^{n}\mathrm{e}^{-A\rho(x_1,x_i)}} & \dfrac{1}{1+\sum_{i=0,\,i\neq1}^{n}\mathrm{e}^{-A\rho(x_1,x_i)}} & \cdots & \dfrac{\mathrm{e}^{-A\rho(x_1,x_n)}}{1+\sum_{i=0,\,i\neq1}^{n}\mathrm{e}^{-A\rho(x_1,x_i)}}\\ \vdots & \vdots & \cdots & \vdots\\ \dfrac{\mathrm{e}^{-A\rho(x_n,x_0)}}{1+\sum_{i=0}^{n-1}\mathrm{e}^{-A\rho(x_n,x_i)}} & \dfrac{\mathrm{e}^{-A\rho(x_n,x_1)}}{1+\sum_{i=0}^{n-1}\mathrm{e}^{-A\rho(x_n,x_i)}} & \cdots & \dfrac{1}{1+\sum_{i=0}^{n-1}\mathrm{e}^{-A\rho(x_n,x_i)}}\end{bmatrix}$$

以及

$$\mathbf{C}=[c_0,c_1,\dots,c_n]^{\mathrm{T}},\quad \mathbf{Y}=[y_0,y_1,\dots,y_n]^{\mathrm{T}}.$$

定义 1[17] 若 $n+1$ 阶矩阵 $\mathbf{S}=(a_{ij})_{i,j=0}^{n}$ 满足

$$|a_{ii}|>\sum_{j\neq i}|a_{ij}|\quad (i=0,1,\dots,n),$$

则称 $\mathbf{S}$ 为强对角占优.

引理 1[17] 如果矩阵 $\mathbf{S}$ 强对角占优, 则 $\mathbf{S}$ 是非奇异的.

定理 1 存在常数 A^*, 当 $A>A^*$ 时, 矩阵 $\mathbf{M}$ 是可逆的.

证明 观察上述矩阵 $\mathbf{M}$ 发现：矩阵 $\mathbf{M}$ 中的每个元素都是非负的, 且矩阵的每一行元素之和为 1, 若要使矩阵满足强对角占优条件, 只需

$$g_{jj}=\frac{1}{1+\sum_{i=0,\,i\neq j}^{n}\mathrm{e}^{-A\rho(x_j,x_i)}}>\frac{1}{2},$$

即满足

$$\sum_{i=0,\,i\neq j}^{n}\mathrm{e}^{-A\rho(x_j,x_i)}<1.$$

令

$$t=\min_{i\neq j}\{\rho(x_i,x_j)\},\quad j=0,1,\dots,n.$$

又由于 $\{x_j\}_{j=0}^n$ 中各元素互不相同, 则 $t>0$, 从而有

$$\sum_{i=0,\, i\neq j}^{n} \mathrm{e}^{-A\rho(x_j,x_i)} \leq n\mathrm{e}^{-At}.$$

如果 $n\mathrm{e}^{-At}<1$, 即 $A>\frac{\ln n}{t}$. 此时取 $A^*=\frac{\ln n}{t}$, 即 $A>A^*$, 那么矩阵 $\mathbf{M}$ 为强对角占优, 再由引理 1 可知矩阵 $\mathbf{M}$ 是可逆的.

由上述定理可知: 方程组 (3) 的系数矩阵是可逆的, 再根据可兰姆法则可知, (3) 中关于 $(c_0, c_1, \ldots, c_n)$ 的线性方程组的解是唯一的. 为此, 我们可解得该线性方程组的解为 $(c_0^*, c_1^*, \ldots, c_n^*)$. 因此, 对于传递函数 $g_j(x,A)$ 满足插值条件 (3) 的插值网络是存在的.

3 近似插值网络及其与精确插值网络的关系

令 $S=\{x_0,x_1,\ldots,x_n\}\subset K$, 其中 K 为 X 的紧集. 设 f 为 K 上的连续泛函, 记 $f(x_i)=f_i$ $(i=0,1,\ldots,n)$, 对于插值样本 (1), 我们称网络

$$N_a(x)=\sum_{j=0}^{n} f_j g_j(x,A)$$

为之近似插值网络.

为进一步说明近似插值网络, 我们首先介绍一些定义和一个引理. 在 $n+1$ 维复欧氏空间 $\mathbb{C}^{n+1}$ 上, 定义如下范数

$$\|z\|_1=|z_0|+|z_1|+\cdots+|z_n|,$$

而 $|z|=\sqrt{z\bar{z}}$ ($\bar{z}$ 为 z 的共轭).

设 $\mathbf{B}=(b_{ij})_{i,j=0}^n$ 是 $n+1$ 阶复矩阵, 定义范数

$$\|\mathbf{B}\|_1=\max_{i=0,1,\ldots,n}\sum_{j=0}^{n}|b_{ij}|.$$

引理 2[15] 设 $\mathbf{B}, \mathbf{B}+\mathbf{\Delta B}$ 为复可逆矩阵, $\mathbf{Z}, \mathbf{Z}+\mathbf{\Delta Z}$ 是 $n+1$ 维列向量, 如果

$$\mathbf{BZ}=\mathbf{b}, \quad (\mathbf{B}+\mathbf{\Delta B})(\mathbf{Z}+\mathbf{\Delta Z})=\mathbf{b},$$

那么有

$$\frac{\|\mathbf{\Delta Z}\|_1}{\|\mathbf{Z}\|_1}\leq \mathrm{Cond}_1(\mathbf{B})\frac{\|\mathbf{\Delta B}\|_1}{\|\mathbf{B}\|_1}\left(\frac{1}{1-\mathrm{Cond}_1(\mathbf{B})\frac{\|\mathbf{\Delta B}\|_1}{\|\mathbf{B}\|_1}}\right),$$

其中 $\mathrm{Cond}_1(\mathbf{B})=\|\mathbf{B}\|_1\|\mathbf{B}^{-1}\|_1$ 为矩阵 $\mathbf{B}$ 的条件数.

定理 2 对于插值样本 (1), 令 $t=\min_{i\neq j,\,0\leq i,\,j\leq n}\{\rho(x_i,x_j)\}$, 则存在一正实数 A_1, 使得当 $A>A_1$ 时

$$|N_e(x,A)-N_a(x,A)|<\frac{2n\mathrm{e}^{-At}}{1-2n\mathrm{e}^{-At}}\sum_{j=0}^{n}|f_j|,$$

对所有 $x\in K$ 皆成立.

证明 首先考虑 $A>A^*$ ($A^*=\frac{\ln n}{t}$ 见上一节). 此时, 插值网络为

$$N_e(x,A)=\sum_{j=0}^{n} c_j^* g_j(x,A),$$

近似插值网络为

$$N_a(x,A)=\sum_{j=0}^{n}c_jg_j(x,A).$$

根据插值条件 (2), 可获得如下的向量形式

$$\mathbf{MC}^*=\mathbf{f},\quad \mathbf{UC}=\mathbf{f},$$

其中

$$\mathbf{U}=\begin{bmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&\cdots&\vdots\\0&0&\cdots&1\end{bmatrix},\ \ \mathbf{C}^*=[c_0^*,c_1^*,\dots,c_n^*]^{\mathrm{T}},\ \ \mathbf{C}=[c_0,c_1,\dots,c_n]^{\mathrm{T}},\ \ \mathbf{f}=[f_0,f_1,\dots,f_n]^{\mathrm{T}}.$$

令 $\mathbf{B}=\mathbf{U},\ \mathbf{Z}=\mathbf{C},\ \mathbf{M}=\mathbf{B}+\mathbf{\Delta B},\ \mathbf{Z}+\mathbf{\Delta Z}=\mathbf{C}^*,\ \mathbf{b}=\mathbf{f}$, 由引理 2 可得

$$\frac{\|\mathbf{C}^*-\mathbf{C}\|_1}{\|\mathbf{C}\|_1}\le\frac{\|\mathbf{U}^{-1}\|_1\|\mathbf{M}-\mathbf{U}\|_1}{1-\|\mathbf{U}^{-1}\|_1\|\mathbf{M}-\mathbf{U}\|_1}.$$

要对上述不等式右边进行估计, 只需要估计 $\|\mathbf{U}^{-1}\|_1\|\mathbf{M}-\mathbf{U}\|_1$. 注意到 $\mathbf{M}=(g_{ij})_{i,j=0}^n$ 中各元素的具体表示以及 $t=\min_{i\neq j}\{\rho(x_i,x_j)\}>0$, 我们有：对于 $A>A^*$,

$$\begin{aligned}|m_{jj}-1|=|g_{jj}-1|&=\left|\frac{1}{1+\sum_{i=0,\,i\neq j}^{n}\mathrm{e}^{-A\rho(x_j,x_i)}}-1\right|\\&=\frac{\sum_{i=0,\,i\neq j}^{n}\mathrm{e}^{-A\rho(x_j,x_i)}}{1+\sum_{i=0,\,i\neq j}^{n}\mathrm{e}^{-A\rho(x_j,x_i)}}\le\sum_{i=0,\,i\neq j}^{n}\mathrm{e}^{-A\rho(x_j,x_i)}\le n\mathrm{e}^{-At}\quad(j=0,\dots,n).\end{aligned}$$

对于 $i\neq j$ 时, 有

$$|m_{ij}|=\frac{\mathrm{e}^{-A\rho(x_j,x_i)}}{1+\sum_{i=0,\,i\neq j}^{n}\mathrm{e}^{-A\rho(x_j,x_i)}}\le\mathrm{e}^{-A\rho(x_j,x_i)}\le\mathrm{e}^{-At},$$

所以

$$\|\mathbf{M}-\mathbf{U}\|_1<n\mathrm{e}^{-At}+n\mathrm{e}^{-At}=2n\mathrm{e}^{-At}.$$

而 $\|\mathbf{U}^{-1}\|_1=1$, 故 $\|\mathbf{U}^{-1}\|_1\|\mathbf{M}-\mathbf{U}\|_1<2n\mathrm{e}^{-At}$.

由于存在数 A', 使得当 $A>A'$ 时, $2n\mathrm{e}^{-At}<1$, 又由于函数 $g(x)=\frac{x}{1-x}$ 在 $(-\infty,1)$ 上为严格单调增加函数, 所以

$$\|\mathbf{C}-\mathbf{C}^*\|_1<\frac{2n\mathrm{e}^{-At}}{1-2n\mathrm{e}^{-At}}\|\mathbf{C}\|_1=\frac{2n\mathrm{e}^{-At}}{1-2n\mathrm{e}^{-At}}\|\mathbf{f}\|_1,$$

而

$$|N_e(x,A)-N_a(x,A)|=\left|\sum_{j=0}^{n}(c_j^*-c_j)g_j(x,A)\right|\le\sum_{j=0}^{n}|c_j^*-c_j|=\|\mathbf{C}^*-\mathbf{C}\|_1.$$

因此, 若 $A>A_1=\max\{A',A^*\}$, 则有

$$|N_e(x,A)-N_a(x,A)|<\frac{2n\mathrm{e}^{-At}}{1-2n\mathrm{e}^{-At}}\sum_{j=0}^{n}|f_j|.$$

4 近似插值网络逼近

在这部分中, 考虑用近似插值网络逼近 (X,ρ) 中紧集上定义的连续泛函问题, 为此, 我们先介绍一些相关概念.

定义 1[18] 设 A, B 是距离空间 (X, ρ) 中的点集, $\gamma > 0$ 为给定的, 如果对于 A 中的任意一点 x, 必存在 B 中的一点 x', 使 $\rho(x, x') < \gamma$, 则称 B 是 A 的一个 γ- 网.

在定义 1 的基础上我们可以定义全有界集, 即

定义 2[18] 设 A 是距离空间 (X, ρ) 中的点集, 如果对于任给的 $\gamma > 0$, A 总存在有限的 γ- 网, 则称 A 是全有界的.

根据定义 2 的全有界的定义, 我们可以获得如下关于全有界集的三条性质:

(a) 任何全有界集必有界.

(b) 如果 A 为全有界集, 则对任给的 $\gamma > 0$, 我们可以取 A 的一个有限子集作为 A 的 γ- 网.

(c) 如果距离空间 (X, ρ) 的子集 A 是紧集, 则 A 必是全有界的.

在插值结点组 $\{(x_0, f_0), \ldots, (x_n, f_n)\}$ 中, 我们令

$$f_j = f(x_j), \quad j = x_0, x_1, \ldots, x_n$$

并且, $A = A(n)$, 即 A 是依赖于 n 的, 由此, 我们可定义如下的近似插值网络

$$N_a(x, A) = \sum_{j=0}^{n} f_j \frac{\mathrm{e}^{-A(n)\rho(x, x_j)}}{\sum_{i=0}^{n} \mathrm{e}^{-A(n)\rho(x, x_i)}} = \sum_{j=0}^{n} f(x_j) \frac{\mathrm{e}^{-A(n)\rho(x, x_j)}}{\sum_{i=0}^{n} \mathrm{e}^{-A(n)\rho(x, x_i)}}.$$

定理 3 设 $f(x)$ 是紧集 $K \subseteq X$ 上的连续泛函, 对于任意的 $\varepsilon > 0$, 存在 $A(n)$ 及自然数 N, 使得

$$|f(x) - N_a(x, A(n))| \leq \varepsilon$$

对所有的 $n > N$ 与 $x \in K$ 皆成立.

证明 由于泛函 $f(x)$ 在紧集 K 上连续, 则在 K 上是一致连续的. 因此, 对于任意 $\varepsilon > 0$, 存在 $\gamma^* > 0$ 的, 使得对任意的 $x, y \in K$, 当 $\rho(x, y) < \gamma^*$ 时, 有

$$|f(x) - f(y)| \leq \frac{\varepsilon}{2}.$$

又由全有界集的性质推知紧集 K 是全有界的, 所以, 对于任意的 $\gamma > 0$ $(\gamma < \gamma^*)$, 在 K 中可以找出有限个点 $x_0, x_1, \ldots, x_n$ 所构成的集合作为紧集 K 的一个 γ- 网, 即

$$K \subseteq \bigcup_{i=0}^{n} U(x_i, \gamma),$$

其中 $U(x_i, \gamma)$ 是以 x_i 为中心, 以 r 为半径的开球. 对于 $x \in K$, 则 x 必在 K 的 γ- 网中某些元素的 γ 邻域中, 即

$$x \in U(x_{J_k}, \gamma), \quad k = 1, \ldots, m,$$

且

$$B = \{x_{J_1}, \ldots, x_{J_m}\} \subseteq \{x_0, \ldots, x_n\},$$

其中 $0 \leq m \leq n$, 并且满足

$$\rho(x, x_{J_k}) < \min_{0 \leq i \leq n, x_i \notin B} \rho(x, x_i), \quad k = 1, \ldots, m. \tag{4}$$

令 $M=\max_{x\in K}|f(x)|$. 由此得到

$$\begin{aligned}
&|f(x)-N_a(x,A(n))|\\
&=\left|f(x)-\sum_{j=0}^{n}f(x_j)\frac{\mathrm{e}^{-A(n)\rho(x,x_j)}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}\right|\\
&=\left|\sum_{j=0}^{n}(f(x)-f(x_j))\frac{\mathrm{e}^{-A(n)\rho(x,x_j)}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}\right|\\
&\le|f(x)-f(x_{J_1})|\frac{\mathrm{e}^{-A(n)\rho(x,x_{J_1})}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}+\cdots+|f(x)-f(x_{J_m})|\frac{\mathrm{e}^{-A(n)\rho(x,x_{J_m})}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}\\
&\quad+\sum_{j=0,\,x_j\notin B}^{n}|f(x)-f(x_j)|\frac{\mathrm{e}^{-A(n)\rho(x,x_j)}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}\\
&\le|f(x)-f(x_{J_1})|\frac{\mathrm{e}^{-A(n)\rho(x,x_{J_1})}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}+\cdots+|f(x)-f(x_{J_m})|\frac{\mathrm{e}^{-A(n)\rho(x,x_{J_m})}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}\\
&\quad+\sum_{j=0,\,x_j\notin B}^{n}\frac{2M}{1+\sum_{i=0,i\ne j}^{n}\mathrm{e}^{-A(n)(\rho(x,x_i)-\rho(x,x_j))}}\\
&\le|f(x)-f(x_{J_1})|\frac{\mathrm{e}^{-A(n)\rho(x,x_{J_1})}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}+\cdots+|f(x)-f(x_{J_m})|\frac{\mathrm{e}^{-A(n)\rho(x,x_{J_m})}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}\\
&\quad+\sum_{j=0,\,x_j\notin B}^{n}2M\mathrm{e}^{-A(n)(\rho(x,x_j)-\rho(x,x_{J_k}))}=I_1+I_2\quad(x_{J_k}\in B),
\end{aligned}$$

其中

$$I_1=|f(x)-f(x_{J_1})|\frac{\mathrm{e}^{-A(n)\rho(x,x_{J_1})}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}+\cdots+|f(x)-f(x_{J_m})|\frac{\mathrm{e}^{-A(n)\rho(x,x_{J_m})}}{\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}.$$

因为 $\gamma<\gamma^*$, 并且当 $x\in U(x_{J_i},\gamma)$, $i=1,\dots,m$ 时, 有

$$|f(x)-f(x_{J_i})|\le\frac{\varepsilon}{2},$$

所以

$$I_1\le\frac{\varepsilon\mathrm{e}^{-A(n)\rho(x,x_{J_1})}}{2\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}+\cdots+\frac{\varepsilon\mathrm{e}^{-A(n)\rho(x,x_{J_k})}}{2\sum_{i=0}^{n}\mathrm{e}^{-A(n)\rho(x,x_i)}}\le\frac{\varepsilon}{2}.$$

对于 I_2 来说, 当 $x_j\notin B$, 由 (4) 知, $\rho(x,x_j)>\rho(x,x_{J_k})$, $k=1,\dots,m$. 所以 $A(n)(\rho(x,x_j)-\rho(x,x_J))>0$. 因此, 对于上述 $\varepsilon>0$, 存在函数 $A(n)>0$, 使得

$$\sum_{j=0,\,x_j\notin B}^{n}\mathrm{e}^{-A(n)(\rho(x,x_j)-\rho(x,x_J))}\le\frac{\varepsilon}{4M}.$$

所以

$$I_2\le 2M\times\frac{\varepsilon}{4M}=\frac{\varepsilon}{2}.$$

综合以上两种情况可得

$$|f(x)-N_a(x,A(n))|\le\varepsilon.$$

定理证毕.

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] Funahashi K., On the approximate realization of continuous mapping by neural networks, *Neural Networks*, 1989, **2**: 183–192.

[3] Chen T. P., Chen H., Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks, *IEEE Trans. Neural Networks*, 1995, **6**: 904–910.

[4] 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.

[5] Chen T. P., Neural networks and approximation problem in application of systems recognition, *Science in China, Ser. A,* 1994, **24**(A): 1–7.

[6] Chen T. P., Approximation continuous function with superposition of Sigmodial function in Hilbert space, *Chinese Science Bulletin*, 1992, **37**(13): 1167–1169.

[7] Jiang C. H., Wu H. B., Li C. M., Approximation of multivariate periodic function and degree of approximation, *Acta Mathematica Sinica, Chinese Series*, 1999, **42**(2): 263–270.

[8] Jiang C. H., Chen T. P., Denseness of superposition of one function with its translation and Dilation in Sobolev space $W_2^m(R^n)$, *Acta Mathematica Sinica, Chinese Series*, 1999, **42**(3): 495–500.

[9] Cao F. L., Zhang Y. Q., Zhang W. G., Neural networks with single hidden layer and the best polynomial approximation, *Acta Mathematica Sinica, Chinese Series*, 2007, **50**(2): 385–392.

[10] Cao F. L., Xu Z. B., The essential order of approximation for neural networks, *Science in China, Ser. E,* 2004, **34**(4): 361–373.

[11] Xu Z. B., Cao F. L., Simultaneous L^p-approximation order for neural networks, Neural Networks, 2005, **18**: 914–923.

[12] Shrivatava Y., Dasgupta S., Neural networks for exact matching of functions on a discrete domain, in Proceedings of the 29th IEEE Conference on Decision and Control, Honolulu, 1990, 1719–1724.

[13] Ito Y., Saito K., Superposition of linearly independent functions and finite mappings by neural networks, *Math. Sci.*, 1996, **21**: 27–33.

[14] Sontag E. D., Feedforward nets for interpolation and classification, *J. Comp. Syst. Sci.*, 1992, **45**: 20–48.

[15] Llanas B., Sainz F. J., Constructive approximate interpolation by neural networks, *J. Comput. Applied Math.*, 2006, **188**: 283–308.

[16] Xie T. F., Cao F. L., The constructivity on interpolation neural networks, to appear.

[17] Cheng Y. P., Matrix theory, Northwestern Polytechnical University Publishers, 2000.

[18] Zheng W. X., Wang S. W., Essentials of real variable function and functional analysis, Shanghai: Renmin Jiaoyu Publishers, 1980.