高校应用数学学报
APPLIED MATHEMATICS A
JOURNAL OF CHINESE UNIVERSITIES
1999年 第14卷 第1期 Vol.14 No.1 1999

与神经网络有关的逼近问题

蒋传海 李南 徐永强

摘要 本文来自神经网络表示能力问题的研究，主要讨论单个函数满足什么条件其所有伸缩和平移的线性组合在某类函数空间中稠密.本文结果对于神经网络理论研究具有重要意义.

关键词 稠密性，Fourier变换，广义函数.
分类号 (中图) O174.41; (1991MR)41A20.

APPROXIMATION PROBLEM CONCERNED WITH NEURAL NETWORKS

Jiang Chuanhai
(Dept.of Information,Shanghai University of Finance and Economics,Shanghai 200433)

Li Nan
(Anhui University,Hefei 230039)

Xu Yongqiang
(Institute of Mathematics,Fudan University,Shanghai 200433)

Abstract This paper results from the research work of representation capability of neural networks,which discusses what kinds of conditions are imposed on single function such that it's linear combinations with it's translations and dilations are dense in some function space . The results of this paper are of great importance in the study of neural networks theory.

Keywords Denseness,Fourier Transform, Generalized Function.
Subject Classification (CL)O174.41;(1991MR)41A20.

1 引 言

与研究神经网络的表示能力密切相关，Pinkus[1]讨论了以下问题：设$g\in C(R^n)$, 置$\mu_g=\overline{\mathrm{span}}\{g(Ax-b)\}$,这里A是非奇异对角矩阵，$b\in R^n$；$\mu_g$称为由g生成的$C(R^n)$中最小平移伸缩不变(TDI)空间.在紧集上一致收敛的拓扑意义下，［1］得到了$\mu_g=C(R^n)$时g所要满足的充要条件.用广义函数Fourier分析的方法，［2］得到了与［1］中相类似的结果，特别证明了只要$g(x)\in L^2(R^n)\backslash\{0\}$，就有$\mu_g=L^2(R^n)$.本文继续［1,2］的讨论，研究：1) n元函数g(x)需满足何种条件所有有限线性组合$\sum_{i=1}^{N}g(A_ix-b_i)$在$L^p(K)$中稠密;2) 分别对于C(K)及$L^p(K)$中紧集内的函数给出更强逼近结果.

以下介绍本文所用的一些符号和广义函数理论中一些熟知的结果.K：n 维欧氏空间R^n中任一紧集.C(K)：K上所有连续函数组成的空间，范数$|f|:=\max_{x\in K}|f(x)|$. $L^p(K)$:K 上满足$\|f\|_p=(\int_K|f(x)|^p dx)^{\frac{1}{p}}<\infty$的所有函数空间. $\mathscr{D}(\mathbf{R}^n)$:$R^n$上支集紧的所有无穷可微函数组成的空间，$\mathscr{D}'(\mathbf{R}^n)$表示$\mathscr{D}(\mathbf{R}^n)$的对偶空间.$S(R^n)$:$R^n$上的Schwartz速降函数空间，$S'(R^n)$为$S(R^n)$的对偶.

若$\varphi\in L^1(R^n)$,$\mu\in D'(R^n)$,它们的Fourier变换定义如下：

$$\hat{\phi}(t)=\int_{\mathbf{R}^n}\phi(x)e^{-ix\cdot t}\mathrm{d}x,\quad t\in\mathbf{R}^n,\hat{\mu}(\psi)=\mu(\hat{\psi}),\quad \psi\in S(\mathbf{R}^n).\tag{1.1}$$

下面是本文要用到的两个广义函数理论中熟知的结果.

(1) [3] 设u, v, w∈$\mathscr{D}'(\mathbf{R}^n)$且至少有两个支集紧，则卷积

$$u*v=v*u,\quad u*(v*w)=(u*v)*w.\tag{1.2}$$

(2) 若u∈S′ (R^n), φ∈S(R^n), 则

$$\widehat{u*v}=\hat{\phi}\hat{u}.\tag{1.3}$$

另外还记R*={x∈R^n:存在 x_i=0}, 这里x_i表示x的第i个分量, R$^\#$=R^n\R*, A表示$\mathscr{A}$阶非奇异实对角阵全体.

2 稠密性定理

定理2.1 设$g(x)\in L^p_{loc}(\mathbf{R}^n)\bigcap S'(\mathbf{R}^n),1\leqslant p<\infty$，若存在$t^0$∈R$^\#$使得$t^0$∈supp$\hat{\mathbf{g}}$(在广义函数意义下)，则所有有限线性组合$\sum_{i=1}^{N}c_ig(A_ix-b_i)$在 $L^p(K)$中稠密. 这里c_i∈R^1, A_i∈A, b_i∈R_n, i=1, …, N.

我们需要下述容易证明的引理.

引理2.2 [4] μ∈S′ (R^n), φ∈S(R^n), 若在开集U中φ≠0而φ· μ=0, 则在U中μ等于零.

定理2.1的证明 假设有限线性组合 $\sum_{i=1}^{N}c_ig(A_ix+b_i)$中不稠，根据Hahn－Banach延拓定理和Riesz表示定理，存在非零h(x)∈L^q(K)$\left(\frac{1}{p}+\frac{1}{q}=1\right)$使得对所有A∈$\mathscr{A}$, t∈R^n,

$$\int_K g(Ax-t)h(x)\mathrm{d}x=0.\tag{2.1}$$

若记$g^A(x)$=g(Ax)并在K外h(x)=0. 则(2.1)可改写为

$$\int_{\mathbf{R}^n}g^A(x-t)h(x)\mathrm{d}x=0.\tag{2.2}$$

对任意k(x)∈$\mathscr{D}(\mathbf{R}^n)$, 记f(x)=p*h(x), 易知f(x)∈$\mathscr{D}(\mathbf{R}^n)$, 由(1.2)得到

$$\int_{\mathbf{R}^n}g^A(x-t)f(x)\mathrm{d}x=0.\tag{2.3}$$

结合(1.3)我们有

$$\hat{f}\,\overline{\hat{g}^A}=0.\tag{2.4}$$

下面证明对任意$t\in R^{\#}$，存在$A\in\mathscr{A}$，使得$t\in \operatorname{supp}\hat{g}^A$. 设N是t的任一开邻域，令$A=\operatorname{diag}\left(\frac{t_1^0}{t_1},\cdots,\frac{t_n^0}{t_n}\right)$，这里$t^0_i$，$t_i$分别是$t^0$, t的第i个分量，则$N_0:=AN$是$t^0$的开邻域. 由定理的条件，则存在$\varphi\in\mathscr{D}(\mathbf{R}^n)$且$\operatorname{supp}\varphi\subset N_0$, 使得

$$\langle g,\hat{\phi}\rangle=\langle\hat{g},\phi\rangle\neq 0. \qquad (2.5)$$

令$\Phi(x)=\varphi(Ax)$, 则$\Phi(x)\in\mathscr{D}(\mathbf{R}^n)$且$\operatorname{supp}\Phi\ \hat{g}\ N$. 那么

$$\hat{g}^A(\Phi)=g^A(\hat{\Phi})=\int_{\mathbf{R}^n}g(Ax)\int_{\mathbf{R}^n}\phi(u)e^{-iAx\cdot u}|A|^n\mathrm{d}u\mathrm{d}x=\langle g,\hat{\phi}\rangle\neq 0. \qquad (2.6)$$

因此对任意$t\in R^{\#}$, 存在$A\in\mathscr{A}$使得$t\in\operatorname{supp}\hat{g}^A$. 由(2.3)和引理(2.2)知道$\hat{f}(t)=0$. Fourier变换的唯一性和$k(x)\in D(R^n)$的任意性保证$h(x)=0$. 此与$h(x)\neq 0$矛盾. 定理2.1证毕.

推论2.3　若$g(x)\in L^p(R^n)\setminus\{0\}$, $1\leqslant p<\infty$, 则所有有限线性组合$\sum_{i=1}^{N}c_ig(A_ix-b_i)$在$L^p(K)$稠密.

证　R^n的零测度R^*是p-薄集(参考［5］)，若$g(x)\in L^p(R^n)$且$g(x)\neq 0$，则必有一点$t^0\in R^{\#}$使得$t^0\in\operatorname{supp}\hat{g}$(在广义函数意义下)，定理2.2保证了推论2.3的结果.

3　强逼近结果

定理3.1［1］　$g(x)\in C(R^n)$, 则所有有限线性组合$\sum_{i=1}^{N}c_ig(A_ix-b_i)$在C(K)中稠密的充要条件是g(x)非下述形式的函数:

$$\sum_{i=1}^{n}\sum_{j=0}^{m_i}c_{ij}x_i^jg_{ij}(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n), \qquad (3.1)$$

这里$g_{ij}\in C(R^{n-1})$, $c_i\in R^1$, $A_i\in\mathscr{A}$, $b_i\in R^n$, $i=1,\cdots,N$.

C(K)中紧集V内的函数, 有下述强逼近结果.

定理3.2　设V是C(K)中的紧集, $g(x)\in C(R^n)$且非(3.1)形式的函数; 则对任意$\varepsilon>0$, 存在正整数N, $A_i\in\mathscr{A}$, $b_i\in R^n$ $(i=1,2,\cdots,N)$, 皆与$f\in V$无关, 及$c_i(f)$, $i=1,\cdots,N$, 与$f\in V$有关, 使得

$$\left|f(x)-\sum_{i=1}^{N}c_i(f)g(A_ix-b_i)\right|<\varepsilon, \qquad (3.2)$$

对所有$f\in V$成立, 且$c_i(f)$是定义在V上的线性泛函.

证　令　$\varphi(x):=ce^{-\|x\|}$且满足$\int_{\mathbf{R}^n}\phi(x)\mathrm{d}x=1$, 记$\phi_\delta:=\delta^{-n}\phi(\delta^{-1}x)$. 定义

$$f*'\phi_\delta(x)=\int_K f(x)\phi_\delta(x-t)\mathrm{d}t, \qquad (3.3)$$

易证对任意$\varepsilon>0$，存在δ_0，当$0<\delta<\delta_0$时，对所有$f\in V$及$x\in K$

$$|f_h *' \phi_\delta(x) - f(x)| < \frac{\varepsilon}{3}, \quad (3.4)$$

把$f*'\phi_\delta(x)$写成Riemann和形式$\sum_{j=1}^{M} f(t_j)\phi_\delta(x-t_j)m(\Delta t_j)$，其中$\bigcup_{j=1}^{M}\Delta t_j$是紧集K的一个分划，$m(\Delta t_j)$是$\Delta t_j$的体积，$t_j\in\Delta t_j$. 对固定的$\delta<\delta_0$，由Arzeta—Ascoli定理知对所有$f\in V$及$x\in K$, 存在$\eta>0$，当$diam(\Delta t_j)<\eta$时，

$$|(f*\phi_\delta)(x) - \sum_{i=1}^{M} f(t_j)\phi_\delta(x-t_j)m(\Delta t_j)| < \frac{\varepsilon}{3}, \quad (3.5)$$

由定理3.2的条件和定理3.1知道，对于每个给定的j, j=1, …, M, 存在正整数N_j, $d_{ij}\in R^1$, $A_{ij}\in A$, $b_{i,j}\in R_n$, i=1, …, N_j使得对所有$x\in K$

$$|(|\phi_\delta(x-t_j) - \sum_{i=1}^{N_j} d_{ij}(A_{ij}x - X_{ij})| < \frac{\varepsilon}{3L}, \quad \frac{\varepsilon}{3}, \quad (3.6)$$

这里$L=\sup_{f\in V}\sum_{j=1}^{M}|f_h(t_j)|m(\Delta t_j)<\infty$,有意义.

联合(3.4)，(3.5)和(3.6)并适当整理排列即得存在正整数N，$A_i\in\mathscr{A}$, $b_i\in R^n$(i=1, 2, …，N)，皆与$f\in V$无关，及$c_i(f)$, i=1, …, N, 与$f\in V$有关，使得

$$|f(x) - \sum_{i=1}^{N} c_i(f)g(A_ix - b_i)| < \varepsilon, \quad (3.7)$$

从定理的证明过程可知$c_i(f)$是定义在V上的线性泛函.

类似于定理3.1，我们有下述结果：

定理3.3　设V是$L^p(K)$中的紧集，$g(x)\in L^p_{loc}(\mathbf{R}^n)\cap S'(\mathbf{R}^n)$，且在广义函数意义下至少存在一点$t^0\in R^{\#}$，使得$t^0\in supp\,\hat{\mathbf{g}}$;则对任意$\varepsilon>0$，存在正整数N，$A_i\in\mathscr{A}$, $b_i\in R^n$(i=1, 2, …，N)，皆与$f\in V$无关，常数$c_i(f)$, i=1, …，N与$f\in V$有关，使得对所有$f\in V$成立

$$\|f(x) - \sum_{i=1}^{N} c_i g(A_ix + b_i)\|_{L^p(K)} < \varepsilon, \quad (3.8)$$

而且所有$c_i(f)$是V上的连续泛函.

为证定理3.3，我们需引进如下定义的函数，对任意$f\in L^p(K)$，定义

$$f_h(x) = \frac{1}{m(B(x,h))}\int_{B(x,h)\cap K} f(t)\mathrm{d}t,$$

这里B(x, h)是R^n中以x为中心以h为半径的球，m(B(x, h))表示球B(x, h)的体积. 对固定的h>0，$f_h(x)$是

R^n上具有紧支柱的连续函数.

引理3.4[6]　V是$L^p(K)$中紧集的充要条件是：(i) V是$L^p(K)$中闭集；(ii) 存在与$f\in V$无关的常数A，使得对所有$f\in V$，成立$\|f\|_{L^p(K)}\leqslant A$.(iii) 对所有$f\in V$，当$h\to 0$时，$\|f_h-f\|_{L^p(K)}$一致收敛于零.

引理3.5　V是$L^p(K)$中的紧集，$V_h:=\{f_h:f\in V\}$，则对固定的$h>0$，有(i) V_h是C(K)中的闭集；(ii) 存在常数A_h，使得对所有$f_h\in V_h$,成立$|f_h|\leqslant A\backslash_h$.(iii) 对任给$\varepsilon>0$,存在$\delta>0$，当$x_1, x_2\in K$且$\|x_1-x_2\|<\delta$时，对所有$f_h\in V_h$时，有$|f_2(x_1)-f_2(x2)|<\varepsilon$.

证　若记T：$f\to f_h$,则T是$L^p(K)\to C(K)$上的连续线性算子.因为对$f_1,f_2\in L^p(K)$有

$$|Tf_1-Tf_2|\leqslant\frac{1}{m(B(x,h))}\int_{B(x,h)\cap K}|f_1(t)-f_2(t)|\mathrm{d}t\leqslant$$
$$\frac{1}{m(B(x,h))}\left(\int_{B(x,h)\cap K}|f_1(t)-f_2(t)|^p\mathrm{d}t\right)^{\frac{1}{p}}\left(\int_{B(x,h)\cap K}\mathrm{d}t\right)^{\frac{1}{q}}\leqslant$$
$$A_h\|f_1-f_2\|_p. \tag{3.9}$$

其中$1/p+1/q=1$,V是$L^p(K)$中紧集，得到T(V)是C(K)中紧集.Arzeta－Ascoli定理保证引理3.5的结果.

定理3.3的证明　对任意$\varepsilon>0$，由引理3.4知存在$h_0>0$，当$0<h<h_0$时，对所有$f\in V$成立：

$$\|f_h(x)-f(x)\|_{L^p(K)}<\frac{\varepsilon}{4}. \tag{3.10}$$

$\varphi(x)$及$f*'\varphi_\delta(x)$定义如3.3，则对固定的$0<h<h_0$，根据引理3.5和积分Minkowski不等式易证存在$\delta_0>0$,当$0<\delta<\delta_0$时，对所有$f_h\in V_h$有

$$\|f_h*'\phi_\delta(x)-f_h(x)\|_{L^p(K)}<\frac{\varepsilon}{4}, \tag{3.11}$$

联合(3.10)则得

$$\|f(x)-f_h*'\phi_\delta(x)\|_{L^p(K)}<\frac{\varepsilon}{2}, \tag{3.12}$$

对固定的$h<h_0,\delta<\delta_0$，把$f_h*'\varphi_\delta(x)$写成Rieman和形式$\sum_{j=1}^{M}f_h(t_j)\phi_\delta(x-t_j)m(\Delta t_j)$，其中$\bigcup_{j=1}^{M}\Delta t_j$是紧集K的一个分划，$t_j\in\Delta t_j, m(\Delta t_j)$是$\Delta t_j$的体积，则有

$$\int_K f_h(t)\phi_\delta(x-t)\mathrm{d}t-\sum_{j=1}^{M}f_h(t_j)\phi_\delta(x-t_j)m(\Delta t_j)=$$
$$\sum_{j=1}^{M}\int_{\Delta t_j}[f_h(t)-f_h(t_j)]\phi_\delta(x-t)\mathrm{d}t+$$
$$\sum_{j=1}^{M}f_n(t_j)\int_{\Delta t_j}[\phi_\delta(x-t)-\phi_\delta(x-t_j)]\mathrm{d}t=I_1+I_2, \tag{3.13}$$

根据Minkowski不等式及引理3.5可得对固定的$h<h_0, \delta<\delta_0$，存在$\eta>0$，当$diam(\Delta t_j)<\eta$时，对所有$f_h \in V_h$成立$\| I_1 \|_{L^p(K)}<\varepsilon/8$, $\| I_2 \|_{L^p(K)}<\varepsilon/8$.

联合公式(3.12)，(3.13)，(3.14)则得对所有$f\in V$

$$\|f(x)-\sum_{j=1}^{M}f_h(t_j)m(\Delta t_j)\phi_\delta(x-t_j)\|_{L^p(K)}<\frac{3\varepsilon}{4}. \qquad (3.14)$$

根据定理2.1，类似于定理3.1的余下证明证得定理.

推论3.6　若$g(x)\in L^p(R^n)\backslash\{0\}$，则对于$L^p(K)$中紧集内的函数有类似于定理3.3的结果.

致谢　感谢陈天平教授及吴宗敏教授的指导和帮助.

(本文第一作者通讯地址：上海市财经大学信息系　邮编　200433)

上海财经大学211工程资助项目.

作者单位：蒋传海　(上海财经大学信息系)
李南　(安徽大学)
徐永强　(复旦大学数学研究所)

参考文献

1　Pinkus,A.,TDI－subspace of $C(R^d)$ and some density problems from neural networks,Journal of Approximation Theory,1996,85:269～287.

2　T.Chen and H.Chen,Universal approximation capability of EBF neural networks with arbitrary activation functions,Circuits Systems,Signal Processing,1996,15(5):671～683.

3　Rudin,W.,Functional Analysis,New York,McGraw Hill,1973.

4　Mhaskar,H.N.,Micchelli,C.A.,Approximation by superposition of sigmoidal and radial basis functions,Advance Applied Mathematics,1992,13:350～373.

5　Harasymiv,S.R.,A note of a dilation in L^p,Pacific J.Math.,1967,21(3):493～501.

6　Natanson,Theory of functions of a real variable,Translated by Ediwin Hewitt,New York,1961.

收稿：1997-03-08，修回：1997-10-30.