

# 过程神经网络的若干理论问题

何新贵<sup>1</sup>, 梁久祯<sup>2</sup>

(1. 北京系统工程研究所, 北京 100101; 2. 北京航空航天大学, 北京 100083)

**[摘要]** 文章提出一种过程神经元模型, 其输入为与时间有关的函数或过程, 它是传统人工神经元模型在时间域上的扩展。基于这种过程神经元模型, 给出了一种仅含一个隐层的前馈型过程神经网络模型, 即基展开过程神经网络模型。证明了相应的连续性定理, 逼近定理, 计算能力定理等。

**[关键词]** 过程神经网络; 函数逼近能力; 计算能力; 连续性

## 1 引言

自从1943年McCulloch与Pitts提出第一个人工神经网络MP模型<sup>[1]</sup>, 人工神经网络的研究已走过了半个多世纪的坎坷道路。这期间人们曾提出了几十种较有影响力的神经网络模型, 如Perceptron(感知机)<sup>[2]</sup>, Adaline自适应线性元件<sup>[3]</sup>, Grossberg的自适应理论<sup>[4, 5]</sup>, Anderson的BSB模型<sup>[6]</sup>, Fukushima的神经认知网络理论<sup>[7]</sup>, Kohonen的自组织映射网络<sup>[8]</sup>等。特别是进入80年代, Hopfield的反馈网络模型<sup>[9]</sup>以及多层BP网络<sup>[10]</sup>的研究使人工神经网络再次进入蓬勃发展阶段。之后, 各种具有实用意义的神经网络模型不断涌现, 如Boltzmann机<sup>[11]</sup>, RBF<sup>[12]</sup>, CMAC<sup>[13]</sup>等。

在上述人工神经网络模型中, 系统的输入均为与时间无关的常量, 即网络的输入是几何点式的瞬时输入。然而, 生物学研究表明<sup>[14]</sup>: 在生物神经元中突触的输出变化与输入脉冲的相对定时有关, 即依赖于持续一定时间的输入过程。在其它实际问题中, 大多数系统的输入亦往往是一个过程, 或一个依赖于时间的函数, 而几何点式的瞬时输入只能在理论上存在。过程式输入放宽了传统神经网络模型对输入的同步瞬时限制, 使问题更为一般化。很多应用可归结为此问题, 如实时控制中, 控

制的输入往往是依赖于时间的连续函数, 控制信号的输出既依赖于空间的聚合(多输入), 又与时间的累积效应密切相关; 在变分问题中, 泛函的定义域一般可为一个与时间有关的过程区间; 依赖于时间的多因素优化问题也都可以归结为具有过程输入的情形。具有过程式输入的神经网络是人工神经网络在时间域上的扩展, 是更一般化的人工神经网络模型。因此, 研究过程神经网络模型的拓扑结构, 函数逼近性质, 学习算法等具有十分普遍性的意义, 在实际中对于求解与过程有关的众多问题有着很大的应用价值。

## 2 过程神经元

过程神经元由加权、聚合和激励三部分组成。单个过程神经元的结构图如图1所示。图1中,  $x_1(t), x_2(t), \dots, x_n(t)$  为过程神经元输入函数向量;  $w_1(t), w_2(t), \dots, w_n(t)$  为相应的权函数;  $K(t)$  为过程神经元的时间聚合基函数;  $f(\cdot)$  为激励函数, 可取线性函数、Sigmoid函数、Gauss型函数等。

过程神经元的输入与输出之间的关系为

$$Y = f((W(t) \oplus X(t)) \otimes K(\cdot) - \theta) \quad (1)$$

这里“ $\oplus$ ”表示某种空间聚合运算, “ $\otimes$ ”表示某种时间(过程)聚合运算。例如

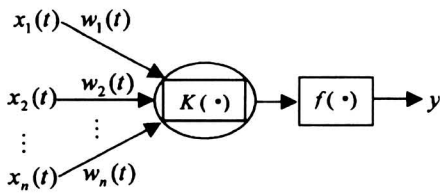


图 1 过程神经元  
Fig.1 Procedure neuron

$$W(t) \oplus X(t) = \sum_{i=1}^n w_i(t)x_i(t) \quad (2)$$

$$A(t) \otimes K(\cdot) = \int_0^T A(t)K(t)dt \quad (3)$$

其中  $K(\cdot)$  为  $[0, T]$  上的一个可积函数。或更一般地设  $K(\cdot)$  为一个一元泛函，定义

$$A(t) \otimes K(\cdot) = K(A(t)) \quad (4)$$

一般假设权函数  $W(t)$  和时间聚合基函数(泛函)  $K(\cdot)$  均连续。由式 (1~3) 描述的过程神经元，其内部运算由加权乘、积分、累加和以及激励函数组成，称之为狭义过程神经元。

其实，加权聚合运算“ $\oplus$ ”和“ $\otimes$ ”还可以取其它一般形式的运算，例如用  $\max$  和  $\min$ ，或 T-算子和 S-算子来构成“ $\oplus$ ”，激励函数  $f$  也可以是任何形式的有界函数。由式 (1, 2, 4) 描述的过程神经元被称为广义过程神经元。

### 3 过程神经网络模型

由若干个过程神经元按一定的拓扑结构组成的网络称过程神经网络。类似于传统神经网络，过程神经网络也可以分前向型和反馈型网络。本文仅讨论前向型过程神经网络。

定义一种其权函数可由一组基函数展开的过程神经网络模型。设  $U$  是泛函空间  $S = \{f(x) | x = x(t) \in R^n, t \in R, f(x) \in V \subset R^m\}$  上的一个紧致集，为讨论简单起见，取  $m = 1$ ，即多输入单输出系统。也不难将结果推广到  $m > 1$  的情况。假设过程神经网络的权利函数可被基函数  $B(t)$  展开，即将权函数的形式限制在一类较简单的函数类中来考虑问题。基函数  $B(t)$  可以是有限基，也可是可数基，可以是正交的，也可以是非正交的。这种过程神经网络如图 2 所示。图 2 中首先把权函数按基展开，中间各子层的运算如下：

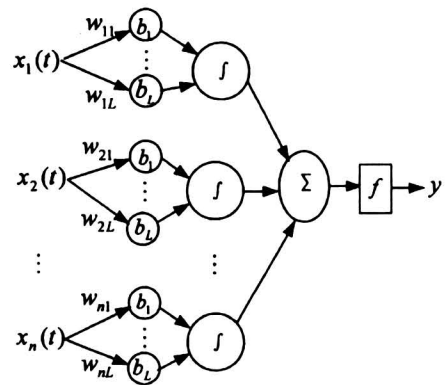


图 2 基展开过程神经网络  
Fig.2 PNN expanded on base functions

$$w_i(t) = \sum_{k=1}^L w_i^{(k)} b_k(t) \quad (5)$$

$$A(t) = \sum_{i=1}^n w_i(t)x_i(t) \quad (6)$$

$$y = f\left(\int_0^T A(t)K(t)dt - \theta\right) \quad (7)$$

或一般地有

$$y = f(K(A(t)) - \theta) \quad (8)$$

其中  $\theta$  为神经元的阈值， $f$  为神经元的激励函数。函数(泛函)  $K(\cdot)$  应根据实际需要来确定。

式 (7) 或式 (8) 表示的网络是一个单层过程神经网络。下面讨论一种具有一个隐层的网络，仍考虑多输入单输出系统，网络的拓扑结构如图 3 所示。其中，中间层(隐层)的各单元分别由图 2 中所示的神经元组成，并设隐层有  $p$  个单元。各隐单元的激励函数和阈限可以相同，也可以不同，具有相同激励函数的网络称为正规过程神经网络，具有不同激励函数的网络称为混合过程神经网络。输出层为一个非时变神经元。为简单起见，仅讨论由式 (7) 定义的网络，并设  $K(t) = 1$ ，即

$$y = g\left(\sum_{j=1}^p v_j f\left(\sum_{i=1}^n \int_0^T \left(\sum_{k=1}^L w_{ij}^{(k)} \cdot b_k(t)x_i(t)\right) dt - \theta_j\right) - \tau\right) \quad (9)$$

若假设输出层中  $g(u) = u, \tau = 0$ ，并令

$$w_{ij}(t) = \sum_{k=1}^L w_{ij}^{(k)} b_k(t) \quad (10)$$

$$c_j(x(t)) = \int_0^T \sum_{i=1}^n w_{ij}(t)x_i(t)dt - \theta_j \quad (11)$$

则

$$y = \sum_{j=1}^p v_j f(c_j(x(t))) \quad (12)$$

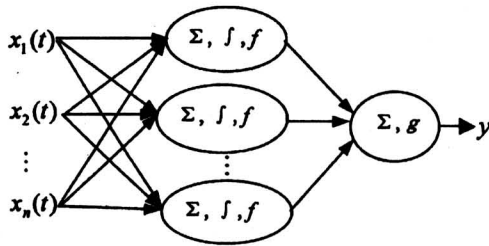


图 3 含一个隐层的基展开过程神经网络

Fig.3 PNN expanded on base functions with one hidden layer

### 4 基本定理

一个合理的神经网络模型必须具备一些最基本的特性,包括连续性、逼近能力和计算能力等。本节就来讨论这些问题。

**定理 1 (连续性定理)** 设式(12)的输入为  $x(t)$ ,  $x^*(t) \in U$ , 相应的输出分别为  $y, y^* \in V$ 。若  $f$  连续, 则对任意的  $\epsilon > 0$ , 必存在  $\delta > 0$ , 当  $\|x(t) - x^*(t)\| < \delta$  时, 有  $|y - y^*| < \epsilon$ 。

**证明:** 因  $f$  连续, 对任意的  $\epsilon_1 > 0$ , 存在  $\delta_1 > 0$ , 当  $|c_j - c_j^*| = \left| \int_0^T \left( \sum_{i=1}^n w_{ij}(t) \cdot (x_i(t) - x_i^*(t)) \right) dt \right| \leq nTW\delta_2 < \delta_1$  时, 有  $|f(c_j) - f(c_j^*)| < \epsilon_1$ , 其中  $W = \sup_{i,j} |w_{ij}(t)|$ , 且  $\|x(t) - x^*(t)\| < \delta_2 < \delta_1 / (nTW)$ 。若记  $h_j = f(c_j), h_j^* = f(c_j^*)$ , 对任意的  $\epsilon > 0$ , 存在  $\delta > 0$ , 当  $|h_j - h_j^*| < \delta$  时,  $|\sum_{j=1}^p v_j h_j - \sum_{j=1}^p v_j h_j^*| = |\sum_{j=1}^p v_j (h_j - h_j^*)| \leq \sum_{j=1}^p |v_j| |h_j - h_j^*| < pV\epsilon_1 < \epsilon$ , 其中  $V = \max_i |v_i|$ 。只需取  $\epsilon_1 = \epsilon / (pV)$ , 则定理成立。

**定理 2** 传统前馈神经网络是过程神经网络的一种特例。

**证明:** 只需在式(9)中取  $x_i(t) = x_i, b_i(t) = 1, T = 1$ , 则

$$y = g \left( \sum_{j=1}^p v_j f \left( \sum_{i=1}^n w_{ij} x_i - \theta_j \right) - \tau \right) \quad (13)$$

即成为一个非时变的传统单隐层前馈神经网络。

**定义 1** 设  $K(\cdot): R^n \rightarrow V \subset R$  为  $R^n$  到  $R$  上的任意连续函数, 记为  $K \in C(R^n)$ , 定义泛函类

$$\Sigma^n(K) = \{f: U \rightarrow V \mid f(x(t)) = \int_0^T K(x(t)) dt, x(t) \in U \subset R^n, f(x) \in V \subset R\}$$

$$x(t) \in U \subset R^n, f(x) \in V \subset R$$

**定义 2** 设  $x(t) = (x_1(t), x_2(t), \dots, x_n(t))^T$ , 其中  $x_i(t)$  为  $[0, T]$  上的连续函数 (记为  $x_i(t) \in C[0, T]$ ),  $i = 1, 2, \dots, n$ 。对任意的  $t_1, t_2 \in [0, T]$ , 若  $|x_i(t_1) - x_i(t_2)| \leq L|t_1 - t_2|, L \geq 0$ , 则称  $x_i(t)$  满足 Lipschitz 条件。若  $\|x(t_1) - x(t_2)\| \leq L_x |t_1 - t_2|, L_x \geq 0$ , 则称  $x(t)$  满足 Lipschitz 条件, 若  $|K(x(t_1)) - K(x(t_2))| \leq L_K \|x(t_1) - x(t_2)\|$ , 则称  $K(\cdot) \in C(R^n)$  满足 Lipschitz 条件。

传统神经网络的研究已经证明了下列熟知的逼近定理。

**引理 1<sup>[10]</sup>** 对任意的连续函数  $g(\cdot) \in C(R^n)$ , 存在仅含一个隐层的前馈神经网络可以以任意的精度逼近  $g$ 。

**定理 3 (逼近定理)** 对任意的由定义 1 定义连续泛函  $G(x(t)) \in \Sigma^n(K)$  和  $\epsilon > 0$ , 若满足 Lipschitz 条件, 则存在过程神经网络  $P$ , 使得  $\|G(x(t)) - P(x(t))\| < \epsilon$ 。

**证明:** 任取  $G \in \Sigma^n(K)$  和  $\epsilon > 0$ , 即

$$G(x(t)) = \int_0^T K(x(t)) dt \quad (14)$$

不失一般性, 设  $T = 1$ , 将  $K$  看成关于  $t$  的复合函数, 并将积分区间  $N$  等分,  $t_i = \frac{i}{N}$  为等分点, 则

$$G(x(t)) = \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} K(x(t)) dt \quad (15)$$

设计泛函  $\tilde{G}(x(t)) = \frac{1}{N} \sum_{i=1}^N K(x(t_i))$  作为  $G(x(t))$  的近似, 则

$$|G(x(t)) - \tilde{G}(x(t))| \leq \sum_{i=1}^N \left| \int_{\frac{i-1}{N}}^{\frac{i}{N}} K(x(t)) dt - \frac{1}{N} K(x(t_i)) \right| \quad (16)$$

由于  $K(x(t))$  对于  $t$  连续, 由积分中值定理, 存在  $\xi_i \in \left[ \frac{i-1}{N}, \frac{i}{N} \right]$ , 使

$$\int_{\frac{i-1}{N}}^{\frac{i}{N}} K(x(t)) dt = \frac{1}{N} K(x(\xi_i)) \quad (17)$$

于是

$$|G(x(t)) - \tilde{G}(x(t))| \leq \frac{1}{N} \sum_{i=1}^N L_K \|x(t_i) - x(\xi_i)\| \leq \frac{1}{N} \sum_{i=1}^N L_K L_x |t_i - \xi_i| \leq \frac{L_K L_x}{N} \quad (18)$$

其中,  $L_K, L_x$  分别为  $K(x)$  关于  $x$  和  $x(t)$  关于  $t$  的 Lipschitz 常数。于是

$$G(x(t)) = \int_0^T K(x(t))dx = \frac{1}{N} \sum_{i=0}^N K(x(t_i)) + O\left(\frac{1}{N}\right) \quad (19)$$

记  $x^{(i)} = x(t_i)$ 。由于  $K(x^{(i)}): R^n \rightarrow V$  为  $C(R^n)$  中的连续函数, 由引理 1 知可用一个传统神经网络来逼近, 再由定理 2, 该传统前馈神经网络一定能够用一个过程神经网络来替代。即

$$|K(x^{(i)}) - P_i(x^{(i)})| < \epsilon_i \quad (20)$$

其中  $\epsilon_i > 0$  为任意小数,  $i = 1, 2, \dots, N$ , 不妨取  $\epsilon_i < \epsilon / (2N)$ 。由式(19)知, 存在  $N_0$ , 当  $N > N_0$  时

$$\left| G(x(t)) - \frac{1}{N} \sum_{i=0}^N K(x(t_i)) \right| < \frac{\epsilon}{2} \quad (21)$$

记

$$P(x(t)) = \frac{1}{N} \sum_{i=0}^N P_i(x(t_i)) \quad (22)$$

则

$$\begin{aligned} & |G(x(t)) - P(x(t))| = \\ & \left| G(x(t)) - \frac{1}{N} \sum_{i=0}^N P_i(x(t_i)) \right| \leq \\ & \left| G(x(t)) - \frac{1}{N} \sum_{i=0}^N K(x(t_i)) \right| + \left| \frac{1}{N} \sum_{i=0}^N K(x(t_i)) - \right. \\ & \left. \frac{1}{N} \sum_{i=0}^N P_i(x(t_i)) \right| < \epsilon/2 + \epsilon/2 = \epsilon \quad (23) \end{aligned}$$

$P(x(t))$  即为所求。

引理 2 存在连续泛函  $K(\cdot)$ , 使得对任意的  $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$  和  $x^*(t) = (x_1^*(t), x_2^*(t), \dots, x_n^*(t))$ , 当  $x(t) \neq x^*(t)$  时, 必有  $K(x(t)) \neq K(x^*(t))$ 。

证明: 考虑  $x(t)$  的第  $i$  个分量  $x_i(t) \in C[0, T]$ , 可用如下多项式逼近:

$$x_i(t) = \sum_{j=0}^{\infty} a_{ij} t^j \quad (24)$$

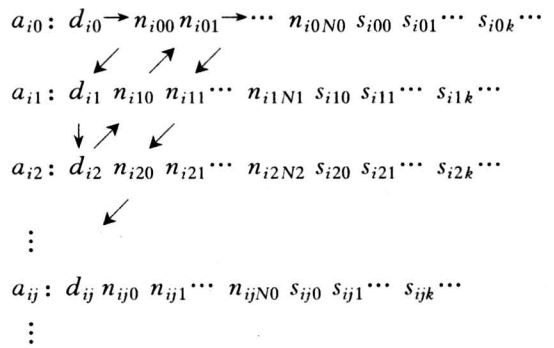
其中,  $\{a_{ij} \in R\}, i = 1, 2, \dots, n; j = 0, 1, 2, \dots$ 。不妨设

$$a_{ij} = (-1)^{d_{ij}} \cdot 10^{p_{ij}} \cdot s_{ij} \quad (25)$$

其中,  $d_{ij} \in \{0, 1\}, p_{ij} \in N$  为自然数集合,  $s_{ij} \in (0, 1)$ 。令

$$\begin{aligned} p_{ij} &= n_{i0} n_{i1} \dots n_{iN_j}, \\ n_{ik} &\in \{0, 1, \dots, 9\}, k = 0, 1, \dots, N_j, \\ s_{ij} &= s_{ij0} s_{ij1} \dots s_{ijk} \dots, \\ s_{ijk} &\in \{0, 1, \dots, 9\}, k = 0, 1, \dots, \end{aligned}$$

将  $a_{ij}, j = 1, 2, \dots$  作如下排列:



取  $z = 0.d_{10}d_{11}d_{12} \dots d_{n0}n_{100}n_{101} \dots n_{n00}d_{11}d_{21} \dots$

$d_{n1}d_{12}d_{22} \dots d_{n2}n_{110}n_{210} \dots n_{n10}n_{101}n_{201} \dots n_{n01} \dots$

即  $z$  的小数部分排列顺序为阵列中从左上角到右下按箭头所示, 并且对于每一个位置做  $i$  的从 1 到  $n$  排列。显然  $z \in [0, 1]$ 。将此映射关系记为  $K$ , 显然  $K$  为一一映射, 由  $z$  的取法知  $K(\cdot)$  连续。

定理 4 (逼近定理) 对任意的连续泛函  $G(x(t)): U \subset C(R^n) \rightarrow V \subset R$  和  $\epsilon > 0$ , 存在过程神经网络  $P$ , 使得  $\|G(x(t)) - P(x(t))\| < \epsilon$ 。

证明: 按照引理 2 的构造方法, 把一个过程神经网络的逼近问题化成了一个传统神经网络的逼近问题, 根据引理 1, 定理得证。

定理 5 (计算能力定理) 狭义过程神经网络的计算能力与图灵机等价。

证明: 首先在狭义过程神经网络中所用到的运算由式(1~3)知, 无非就是“+”, “·”, “j”以及它们的复合运算, 这些运算都可以用图灵机实现(对“j”运算只求近似值), 故过程神经网络的计算能力不会大于图灵机。另一方面, 由文献[15]知线性阈值神经元的计算能力等价于图灵机, 而线性阈值神经元又是过程神经元的特例, 因此过程神经网络的计算能力不会小于图灵机。总之, 过程神经网络的计算能力等价于图灵机。

### 5 结论

本文提出了一类新的神经网络——过程神经网络的概念。并证明了相应的连续性定理, 逼近定理, 计算能力定理等重要理论结果。过程神经网络是对生物神经网络更贴切的模拟, 实际中很多应用可归结为此问题, 包括诸如某些化工和酿造等工业过程的建模和控制、农业领域内大棚栽培和室内养殖的调控和优化等。总之, 模型对于与过程有关的众多问题的解决具有广泛的应用前景。

## 参考文献

- [1] McCulloch W S, Pitts W H. A logical calculus of the ideas immanent in neuron activity [J]. *Bulletin Mathematical Biophysics*, 1943, 5: 115~133
- [2] Rosenblatt F. *Principles of neuro-dynamics* [M]. Spartan, New York, 1962
- [3] Widrow B. Generalization and information storage in networks of adaline 'neurons' [A]. In: Yovitz M, Jacobi G, Goldstein G. *Self-organizing systems* [M], Washington DC: Spartan Books, 1962. 435~461
- [4] Grossberg S. Adaptive pattern classification and universal recoding I: parallel development and coding of neural feature detectors [J]. *Biology Cybernetics*, 1976, 23: 121~134
- [5] Grossberg S. Adaptive pattern classification and universal recoding II: feedback, expectation, olfaction and illusion [J]. *Biology Cybernetics*, 1976, 23: 187~202
- [6] Anderson J A. A simple neural network generating interactive memory [J]. *Mathematical Biosciences*, 1972, 14: 197~200
- [7] Fukushima K. Cognitron: a self-organizing multi-layered neural network [J]. *Biology Cybernetics*, 1975, 20 (3): 121~136
- [8] Kohonen T. *Self-organization and associative memory* [M]. Springer-Verlag, Berlin, 1984
- [9] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities [A]. *Proc. of the National Academy of Science* [C], U. S. A. 1983, 79: 2554~2558
- [10] Hecht-Nielsen R. Theory of the back-propagation neural network [A]. *Proc. IJCNN89* [C], 1989, 1: 593~599
- [11] Hinton G H, Sejnowskii T J, Ackley D H. Boltzmann machine: constraint satisfaction networks that learn [M]. CMU - CS - 84 - 119, Carnegie-Millon Uni., 1984
- [12] Powell M J D. Radial basis functions for multivariate interpolation: a review [R]. Department of App. Math. and Theoretical Physics, Cambridge University. Cambridge, England. 1985
- [13] Albus J S. A new approach to manipulator control, the cerebellar model articulation controller (CMAC) [J]. *Trans. of the ASME Journal of Dynamic Systems, Measurement and Control*, 1975, 97: 220~227
- [14] Zhang L I, Tao H W, Holt C E, et al. A critical window for cooperation and competition among developing retinotectal synapses [J]. *Nature*, 1998, 395: 37~44
- [15] 刘晓鸿, 戴汝为. 线性阈值单元神经网络的图灵等价性 [J]. *计算机学报*, 1995, 18 (6): 438~442

## Some Theoretical Issues on Procedure Neural Networks

He Xingui<sup>1</sup>, Liang Jiuzhen<sup>2</sup>(1. *Beijing Institute of System Engineering, Beijing 100101, China;*2. *Beijing University of Aeronautics and Astronautics, Beijing 100083, China)*

[Abstract] In this paper, a novel artificial neuron model-procedure neuron model is proposed, in which the inputs are functions or procedures associated with 'time'. Based on these neurons, a model named procedure neural network, which is also a feedforward network with only one hidden layer, is constructed. The authors call this neural network as Procedure Neural Network (PNN) expanded on certain base functions. The related continuity, function approximation ability and computational capability theorems are proved.

[Key words] procedure neural networks; function approximation ability; computational capability; continuity

\* \* \*

## 更正

本刊 2000 年第 10 期第 86 页图 3 第 4 象限内的  $V_g < c$  应为  $V_g > c$ , 特此更正, 并向作者、读者致歉。