

过程神经网络的训练及其应用

何新贵¹, 梁久楨², 许少华²

(1. 北京系统工程研究所, 北京 100101; 2. 大庆石油学院计算机系, 黑龙江 安达 151400)

[摘要] 研究过程神经网络的学习算法及其在过程模式识别中的应用。针对权值基展开的过程神经网络讨论了权值基的选取规则和对采样曲线的标准化处理问题, 给出了含一个隐层的过程神经网络的误差反传播学习算法。以聚合化学反应和渗流实验两个具体实例验证了算法的有效性, 也说明了过程神经网络的广泛应用前景。

[关键词] 过程神经网络; 学习算法; 模式识别; 化学反应; 渗流

1 引言

半个多世纪以来, 经过人们的不断努力神经网络得到了长足发展, 其应用领域也几乎渗透到各个学科。尤其是在 1989 年, 多层神经网络被证明是连续函数的一致逼近器^[1~4], 为神经网络奠定了理论基础, 也为神经网络的发展起到了很大的推动作用。近年来神经网络新的模型不断涌现, 新的学习算法也层出不穷。总的来说, 神经网络在解决 L_2 空间上的映射问题无论是在理论方面还是在实际应用中已趋于成熟。另一方面, 在许多实际问题中系统的输入为一个与时间有关的过程, 如一个化学反应过程, 股市的波动过程等, 用神经网络解决这类问题开始引起人们的关注。但是, 由于依赖于时间的采样往往数据量较大, 而传统的神经网络难于解决较大样本的学习和泛化问题, 故这类问题的解决期待着新的模型的出现。生物学研究表明^[5], 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。因此过程式输入放宽了传统神经网络模型对输入的瞬时同步限制, 使问题更为一般化。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。笔者在文献 [6] 中提出了“过程神经网络”的概念。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。

本文在文献 [6] 的基础上给出了过程神经网络的学习算法及其在过程模式识别中的应用实例。

2 过程神经网络模型

过程神经元的结构与传统 MP 模型的结构相类似, 由加权、聚合和激励运算三部分组成。与传统神经元不同之处在于过程神经元的输入和权值可以是时变的, 其聚合运算既有对空间的多输入汇聚, 亦有对时间过程的累积 [6]。由若干个过程神经元按一定的拓扑结构组成的网络称为过程神经网络。这里, 考虑了一种权函数的基展开过程神经网络模型, 为了便于讨论, 设其为一个多输入单输出系统 (不难将其推广到多输入多输出系统), 网络的拓扑结构如图 1 所示。

其中, 输入层有 n 个单元, 中间层 (隐层) 具有 m 个单元, 输出层为线性关系。过程式的输

[收稿日期] 2000-10-31; 修回日期 2001-01-11

[基金项目] “九七三”国家重点发展资助项目 (G1999032705)

[作者简介] 何新贵 (1938-), 男, 浙江浦江人, 北京系统工程研究所研究员, 博士生导师

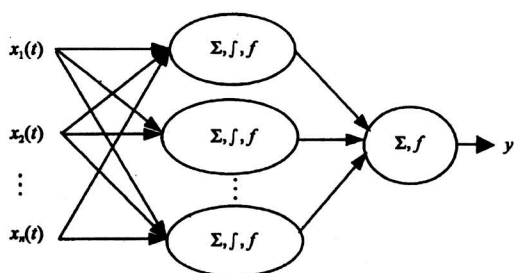


图1 含一个隐层的基展开过程神经网络

Fig.1 PNN expanded on base functions with one hidden layer

入与输出之间关系为:

$$y = \sum_{i=1}^m v_i f \left(\sum_{j=1}^n \int_0^T \left(\sum_{l=1}^L w_{ji}^{(l)} b_l(t) x_j(t) \right) dt - \theta_i \right), \quad (1)$$

式(1)中, v_i 为隐层到输出层的连接权值, $w_{ji}^{(l)}$ 为输入层与隐层的连接权值, $b_l(t)$ 为权值基函数, θ_i 为隐层的输出阈值, L 为权值基函数的个数, $[0, T]$ 为时间采样区间。由文献 [6] 知, 过程神经网络式(1)是 n 维连续函数空间 $C_n [0, T]$ 上的一致逼近器, 因此可以使用式(1)逼近 $C_n [0, T]$ 上的任意的映射。

3 学习算法

过程神经网络的学习可借鉴传统的梯度下降法, 如 BP 算法。将式(1)重写为

$$y = \sum_{i=1}^m v_i f \left(\int_0^T \left(\sum_{j=1}^n \tilde{w}_{ji}(t) x_j(t) \right) dt - \theta_i \right), \quad (2)$$

其中 $\tilde{w}_{ji}(t) = \sum_{l=1}^L w_{ji}^{(l)} b_l(t)$ 。

给定 K 个样本:

$$(x_{11}(t), x_{12}(t), \dots, x_{1n}(t); d_1)$$

$$(x_{21}(t), x_{22}(t), \dots, x_{2n}(t); d_2)$$

.....

$$(x_{K1}(t), x_{K2}(t), \dots, x_{Kn}(t); d_K)$$

其中 d_k 为第 k 个样本的输出 (期望输出)。网络误差函数可取

$$E = \sum_{k=1}^K (y_k - d_k)^2 = \sum_{k=1}^K \left(\sum_{i=1}^m v_i f \left(\int_0^T \left(\sum_{j=1}^n \tilde{w}_{ji}(t) x_{kj}(t) \right) dt - \theta_i \right) - d_k \right)^2. \quad (3)$$

由梯度下降法, 网络权值学习规则为

$$v_i = v_i + \alpha \Delta v_i \quad (4)$$

$$\tilde{w}_{ji}(t) = \tilde{w}_{ji}(t) + \beta \Delta \tilde{w}_{ji}(t), \quad (5)$$

$$\theta_i = \theta_i + \gamma \Delta \theta_i, \quad (6)$$

其中 $i=1, 2, \dots, m; j=1, 2, \dots, n; \alpha, \beta, \gamma$ 为学习速度。令

$$u_{ki} = \int_0^T \left(\sum_{j=1}^n \tilde{w}_{ji}(t) x_{kj}(t) \right) dt - \theta_i \quad (7)$$

则

$$\Delta v_i = - \frac{\partial E}{\partial v_i} = - 2 \sum_{k=1}^K \left(\sum_{i=1}^m v_i f(u_{ki} - d_k) \right) f'(u_{ki}), \quad (8)$$

$$\Delta \tilde{w}_{ji}(t) = - \frac{\partial E}{\partial w_{ji}(t)} =$$

$$- 2 \sum_{k=1}^K \left(\sum_{i=1}^m v_i f(u_{ki} - d_k) \right) f'(u_{ki}) \int_0^T x_{kj}(t) dt, \quad (9)$$

$$\Delta \theta_i = - \frac{\partial E}{\partial \theta_i} = 2 \sum_{k=1}^K \left(\sum_{i=1}^m v_i f(u_{ki} - d_k) \right) f'(u_{ki}), \quad (10)$$

若取 $f(u) = (1 + e^{-u})^{-1}$, 则 $f'(u) = f(u) \cdot (1 - f(u))$ 。

算法描述如下:

步1 给定误差精度 ϵ ; 累计学习迭代次数 $s=0$; 学习最大迭代次数 M ; 选取权值基函数 $b_l(t), l=1, 2, \dots, L$ 。

步2 初始化权值和阈值 $v_i, \tilde{w}_{ji}(t), \theta_i (j=1, 2, \dots, n; i=1, 2, \dots, m)$ 。

步3 由式(3)计算误差函数 E , 如果 $E < \epsilon$ 或 $s > M$ 转步5。

步4 按式(4)~(6)修正权值和阈值; $s+1 \rightarrow s$; 转步3。

步5 输出学习结果; 结束。

由于过程神经网络与传统的神经网络有较大的区别, 尤其是引入了权值基函数和积分运算。从数学意义上讲, 在函数空间中会有无穷多的基函数存在, 但常用的基函数有三角函数、多项式函数、Fourier 基函数、小波基函数等。过程神经网络中的基函数, 在某种意义上, 用于将与时间有关的输入函数映射成为人们熟悉的光滑函数, 即磨光算子。为避免冗余的叠加, 基函数常被选择为正交基, 在本文的实验中采用了正交基函数。在实验部分的两个例子中发现, 采用正交基函数会使得网络的训练速度明显加快。如采用如下正交基函数

$$b_l(t) = \sin(\xi_l \pi t).$$

这里 $\xi_l = l / \text{ent } T$ 。当然，也可以选择其它基函数，有时对于具体的问题还会有更好的基函数。笔者建议在选取基函数时，首先要选正交基函数，再是所选取的基函数要便于计算。但是，为什么正交基函数会使得网络的学习速度加快，还有待于更深入的研究。

时间积分（或聚合）是过程神经网络与传统神经网络的最大区别。由于在过程神经网络中输入是与时间相关的量，即与时间有关的函数，而输出只是一个静态向量，显然在输入与输出层之间应有某些神经元来完成对时间的聚合运算。时间聚合可以有多种形式，如求极值 $\max(b_l(t), x(t))$ ，做卷积运算 $b_l(t) * x(t)$ ，计算内积 $\langle b_l(t), x(t) \rangle$ 等等，其中 $b_l(t), x(t) \in [0, T]$ 。本文中选取内积运算，这也是前文中所见到的。对于输入为一连续函数的情况，实际计算时一种常用的数值处理方法是将输入 $x(t)$ 在 $[0, T]$ 上 N 等分。而对于一个实际问题，在多数情况下，不可能知道 $x(t)$ 在 $[0, T]$ 上的所有值，通过实验的方式获得的只是部分值，有时这些数据在 $[0, T]$ 上分布甚至是非等距的。对于这种情况有两种选择：**a.** 直接采用这些实际的非等距的数据作为过程神经网络的输入；**b.** 对原始数据加以处理，使得它们之间具有相同的间距。前一种选择虽然能够反映实际数据的情况，但却容易带来较大的计算误差；而后一种选择可以避免由于不同间距引起的计算误差。对原始数据的处理可分为两步，第一步是由原始数据构造拟合曲线 $x(t)$ ；第二步再将 $x(t)$ 在 $[0, T]$ 上进行 N 等分。

4 应用举例

过程神经网络可以用于任何输入为时间的函数而输出为一静态向量的问题，这类问题在实际中大量存在。这一节给出两个这样的例子，一个是有关聚合化学反应温度变化对生成物分子数的影响问题；另一个是在渗流实验中各参量的变化最终反映油藏区域采收率问题。这两个问题是石油开发和加工中遇到的重要的实际问题。

4.1 例1 丙烯酰胺均相聚合化学反应

该聚合反应是化学实验中一类重要的化学反应，在这类反应的诸影响因素中，温度是影响生成物浓度和反应时间的重要因素。为了得到温度变化与生成物浓度之间的函数关系，传统的做法是借助于微分动力学和热力学方程。但由于物理模型的复

杂性以及由实验带来的测量误差的不可测性，往往难以找到两者之间的真实关系，而过程神经网络正好适合求解这类问题。下面根据丙烯酰胺均相聚合化学反应的实验数据来处理这一问题。

表1列出了丙烯酰胺均相聚合反应实验的部分结果。表1中 t 表示从0时刻开始以分钟计的累积时间， T_i ($i = 1, 2, \dots, 9$) 表示第 i 组实验的温度 ($^{\circ}\text{C}$) 测试结果； N_m 为每组实验对应生成物的分子数 (10^6 cm^{-3})，即丙烯酰胺均相聚合反应生成物的浓度。

表1 丙烯酰胺均相聚合反应实验结果

Table 1 Experiment results of super-high molecular-weight polyacrylamide

t/min	$T_1/^{\circ}\text{C}$	$T_2/^{\circ}\text{C}$	$T_3/^{\circ}\text{C}$	$T_4/^{\circ}\text{C}$	$T_5/^{\circ}\text{C}$	$T_6/^{\circ}\text{C}$	$T_7/^{\circ}\text{C}$	$T_8/^{\circ}\text{C}$	$T_9/^{\circ}\text{C}$
0	17	18	18	18	14	15	15	14	13
10	19	16	21	18	15	16	16	15	15
20	23	16	23	18.5	17	16	18	16	16
30	27	17	26	21.5	20	16.5	21	17.5	18
40	30	18	31	24.5	22	18	25	19.5	19
50	34	19	36	28.5	25	20	30	22	21
60	39	20	43.5	34	28	22	36	24	23
70	45	21	52.5	38	33	24	40	26.5	25
80	52	22.5	59	46	38	29	54	30.5	27
90	64	24	68	56	46	32	64	35.5	30
100	73	26	73	63	52	36	73	43	34
110	81	28	77	74	60	40	76	54.5	41
120	82	32	79	78	66	44.5	77	67	44
130	84	36	83	93	72	50.5	78	79.5	49
140	88	45	87	95	76	57.5	79	83	54
150	92	54	96	97	81	66.5	84	87	65
160	101	64	98	100	83	67	86	93	77
170	113	68	102	110	94	69	89	98	80
$(N_m)^*$	18.11	18.38	18.94	18.53	14.96	17.01	17.54	19.15	16.44

* 每组实验对应生成物的分子数 (10^6 cm^{-3})

图2给出了描述丙烯酰胺均相聚合反应实验温度 ($^{\circ}\text{C}$) 随时间 (min) 变化的9组曲线，每条曲线对应一组实验生成物的分子数。对于过程神经网络而言，每条曲线对应一个与时间相关的输入序列，而生成物的分子数为过程神经网络的输出。

本例中训练过程神经网络的各参数选取如下：1个输入节点；18个权值基函数；22个隐层节点；1个输出节点。学习速率为0.05；最大学习次数为

5 000; 学习精度为 0.001。

图 3 给出了用 38 个样本训练过程神经网络的误差变化曲线, 学习过程经过在 610 次迭代后即可达到给定的误差精度 (0.001)。为了测试学习完后的过程神经网络的泛化能力, 笔者用 5 个非训练样本进行测试, 并与实际值进行比较, 测试的结果见表 2。

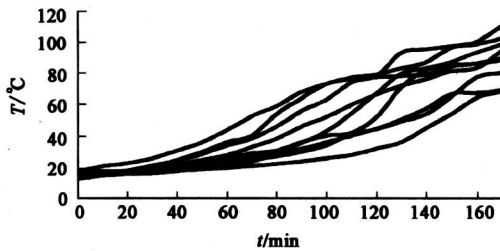


图 2 温度-时间曲线图

Fig.2 Temperature-time curve

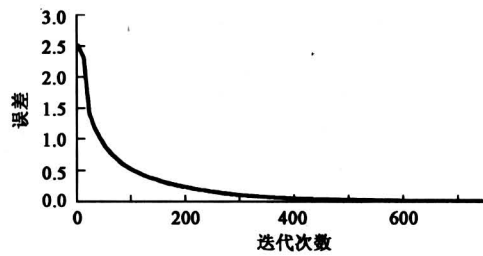


图 3 学习的误差曲线

Fig.3 Learning error curve

表 2 测试样本的识别结果

Table 2 Identification results for test samples

预测分子数/ 10^6 cm^{-3}	13.60	18.63	13.94	14.76	16.85
实际分子数/ 10^6 cm^{-3}	15.61	15.74	16.53	15.42	18.22
绝对误差/ 10^6 cm^{-3}	2.01	2.89	2.59	0.66	1.37
相对误差/%	12.9	18.4	15.7	4.3	7.5

4.2 例 2 石油地质中的渗流问题

渗流问题的研究是石油储量预测中很有意义的工作。传统预测石油储量的方法是通过一定量的实验室模拟油藏含油岩石渗流实验的数据和相关的数学力学模型来计算油层的采收率, 从而进一步估算石油储量。但这种方法需要建立和求解复杂的数学力学模型, 且渗流问题是影响因素较多的非线性问

题, 建模精确度低, 求解难度大, 适应性差, 故实际中难以满足石油储量预测的要求。

渗流是一个与过程有关的问题, 利用过程神经网络建立实验室数据与油层采收率之间的关系, 从而避免了上述复杂的数学模型问题。表 3 给出了实验室含油岩石体积为 595.75 cm^3 的渗流实验的一组记录, 其中, t 为时间, V_w 为其产水量 (mL), V_o 为其产油量 (mL), W_p 为其含水率 (单位体积内产水量与产油量和产水量之比), Δp 为压力差 (kPa)。以上 4 个量 (V_w , V_o , W_p , Δp) 是与时间有关的变化量, 作为过程神经网络的输入。与该组实验对应的油层的采收率 $R = 24.14\%$, 作为过程神经网络输出。实际上, 这样的渗流实验有若干组, 如用其中的 8 组记录就可以构成过程神经网络的 8 个学习样本。

表 3 渗流实验的一个记录

Table 3 A record for seepage experiments

t/h	V_w/mL^*	V_o/mL^*	$W_p/\%$	$\Delta p/\text{kPa}$
0	0	8.4	0	5.59
4	8.4	0.8	91.3	3.14
8	8	0.4	95.24	2.99
12	8	0.3	96.39	3.04
16	8.3	0.25	97.08	2.94
20	8.8	0.2	97.78	2.89
24	9	0.15	98.36	2.84
28	9.4	0.1	98.95	2.96
32	9.7	0.1	98.98	2.84
36	9.5	0.1	98.96	2.89
40	9.55	0.05	99.48	2.89
44	9.75	0.05	99.49	2.89
48	7.15	0.05	99.31	2.84

* 含油岩石体积为 595.75 cm^3 的产水量和产油量

本例中训练过程神经网络的参数选取如下: 4 个输入节点; 43 个权值基函数; 20 个隐层节点, 1 个输出节点。学习速率为 0.05; 最大学习次数为 30 000; 学习精度为 0.05。图 4 给出了用上述样本和学习参数训练过程神经网络的误差变化曲线。这种方法用于发现渗流过程实验数据与采收率之间的映射关系比较简单, 易操作, 所得结果能够满足实际需求。

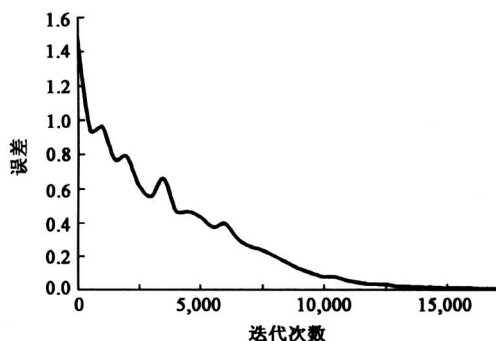


图4 学习误差曲线

Fig.4 Learning error curve

5 结语

过程神经网络是传统神经网络在时间域上的扩展，从神经生物学的角度出发更符合生物神经元的运行机制，可在求解大量与时间有关的过程式输入问题方面发挥较大作用。由于这种模型属首次提出，还有许多问题，尤其是网络的学习与泛化，有待研究和完善。由于过程神经网络的输入是一个过程，其数据量一般较大，在网络训练中往往要花费较大的时间和空间开销。而对于较少的训练样本，

网络的泛化能力一般较低。因此，如何选择权值基函数和提高网络的学习速度是一个值得研究的问题，如何精选学习样本以便提高学习速度和泛化能力，是过程神经网络重点研究的另一个问题。

参考文献

- [1] Hecht-Nielsen R. Theory of the backpropagation neural network [A]. Proceedings of International Joint Conference on Neural Networks [C]. 1989, 1: 593~599
- [2] Cybenko G. Approximation by superposition of a sigmoid function [J]. Mathematics of Control, Signals and Systems, 1989, 2: 303~314
- [3] Hornik K, Stinchcombe M, White H. Multilayer feed-forward networks are universal approximators [J]. Neural Networks, 1989, 2: 359~366
- [4] Funahashi K. On the approximate realization of continuous mappings by neural networks [J]. Neural Networks, 1989, 2: 183~192
- [5] Zhang Li I, Tao Huizhong W, Holt C E, et al. A critical window for cooperation and competition among developing retinotectal synapses [J]. Nature, 1998, 395: 37~44
- [6] 何新贵, 梁久祯. 过程神经网络若干理论问题[J]. 中国工程科学, 2000, 2(12): 40~44

Learning and Applications of Procedure Neural Networks

He Xingui¹, Liang Jiuzhen², Xu Shaohua²

(1. Beijing Institute of System Engineering, Beijing 100101, China; 2. Dept. of Computer Science, Daqing Petroleum Institute, Anda, Heilongjiang 151400, China)

[Abstract] This paper deals with learning algorithms for procedure neural networks (PNN) and its applications in aggregation chemical reaction and seepage test in oil geology. Weight bases selection rules and pattern curve standard problems are also discussed. These examples show that PNN have extensive applications.

[Key words] procedure neural networks; learning algorithm; pattern recognition; chemical reaction; seepage

* * * * *

《中国工程科学》(月刊)是中国工程院院刊

欢迎读者直接向本刊编辑部订阅