前馈神经网络 精确映射 算法 网络结构 结构设计 (17)

计算机科学 1999Vol. 26№. 4

63-65

能实现精确映射的前馈神经网络快速算法与结构设计

A Fast Training Algorithm for Feedforward Neural Nets with High Precision and MLP Configuration Design

张代远 虞厥邦 邱玉辉

(电子科技大学 成都 610054)(西南师范大学计科系 重庆 400715)

TP18

Abstract Some sufficient conditions for a three layered perceptron to implement precise I/O mapping are proven in the present paper. Two fast training algoritbms based on these conditions together with some configuration design guidelines are given. The validity of the main results is verified via numerical examples, the simulation results show also that the training speed as well as the mapping accuracy are much higher than the conventional BP algorithm

Keywords Neural network, Multi-layered perceptron, Moore-Penrose Pseudoinverse, BP algorithm

一、引言

人们常用 BP 算法[1]训练多层感知机，但 BP 算法的缺点[2,3]使它在工程应用上受到了限制，尽管从理论上人们证明了对 BP 算法而言，如果不限制神经元的个数，则三层前馈神经网络可以以任意精度实现给定样本的映射[4]，但对于给定的实际问题，BP 算法并没有具体给出确定神经元个数的计算方法，使得应用时须凭经验选择[5]。

本文正是针对 BP 算法的这一缺点，提出了一种基于 Moore-Penrose 广义逆的代数方法。该方法给出了在实现精确映射要求下，确定神经元个数的两种充分条件，并给出了具体的计算公式。这对于那些要求高精度逼近的场合无疑具有指导意义。本文建议的这种新训练算法，其另一个主要优点是：在实现了高精度 I/O 映射的同时还获得远较 BP 算法为快的计算速度，原因是本算法无需经过迭代计算。

二、新算法原理与网络结构设计

三层前馈网络的输入输出关系可由下式给出：

$$X_{i+1}=f_{i+1}(\overline{W}_i\overline{X}_i) \qquad i=1,2 \tag{1}$$

这里 $X_i\in R^{s_i\times k}$ 是第 i 层的输出矩阵（s_i 为第 i 层的神经元个数，k 为需要训练的样本个数），$\overline{X}_i=\begin{bmatrix}X_i\\E\end{bmatrix}$，称 $\overline{X}_i$ 为 X_i 的扩充矩阵，其中 $E\in R^k$ 是所有元素都为 1 的（行）向量。$W_i\in R^{s_{i+1}\times s_i}$ 是第 i 层与第 $i+1$ 层间的连结权矩阵，$\overline{W}_i=[W_i b_{i+1}]$，$b_{i+1}$ 是第 $i+1$ 层各神经元的阈值（列）向量，$b_{i+1}\in R^{s_{i+1}}$。f_{i+1} 为第 $i+1$ 层神经元的变换函数，这样（1）式便是包含了阈值在内的一个统一的输入输出关系表达式。前馈网络的训练目的就是选择适当的 $\overline{W}_i(i=1,2)$ 以满足给定样本对的映射。

令 $P=[u^{(1)},u^{(2)},\cdots,u^{(k)}]$，$T=[t^{(1)},t^{(2)},\cdots,t^{(k)}]$ 为给定的样本对矩阵，其中 $u^{(p)}$ 为输入样本（列）向量，并设 $u^{(i)}\neq u^{(j)}(i\neq j)$；$t^{(p)}$ 为监督（列）向量（或称教师向量），$u^{(p)}\in R^{s_1}$，$t^{(i)}\in R^{s_3}$，$P=1,2,\cdots,k$；$Y=[y^{(1)},y^{(2)},\cdots,y^{(k)}]$，为网络输入端输入 P 时，在网络输出端得到的实际输出矩阵，$y^{(p)}$ 为输入端输入 $u^{(p)}$ 时，输出端的实际输出（列）向量，$y^{(p)}\in R^{s_3}$。

又令
$$J=\frac{1}{2}\|Y-T\|_F^2=\frac{1}{2}\sum_{p=1}^{k}\|y^{(p)}-t^{(p)}\|_2^2 \tag{2}$$

并称 J 为网络的价值函数。为以后的叙述方便我们给出如下一个定义：

定义 若对给定的样本对 P，T，如果存在权 $\overline{W}_1$ 和 $\overline{W}_2$ 使一个三层前馈网络的价值函数 $J=0$，则我们称该网络对给定的样本对存在一个精确映射 M：$u^{(p)}\to t^{(p)}$。

为了证明这种精确映射 M 的存在性，给出一些情况下 M 存在的充分条件，用以进行网络结构设

计，下面将讨论两种情形。

情形 1 隐层神经元由样本个数确定。

由(1)有 $X_2=f_2(\overline{W}_1\overline{X}_1)$ (3)

$$X_3=T=f_3(\overline{W}_2\overline{X}_2) \tag{4}$$

于是：$\overline{W}_1\overline{X}_1=g_2(X_2)$ (5)

$$\overline{W}_2\overline{X}_2=g_3(T) \tag{6}$$

式中 $g_2=f_2^{-1}$，$g_3=f_3^{-1}$ 分别是隐层和输出层神经元变换函数的逆变换。按广义逆矩阵原理由(6)得：

$$\overline{W}_2=g_3(T)\overline{X}_2^+ +\eta(I_{s_2+1}-\overline{X}_2\overline{X}_2^+) \tag{7}$$

这里$\overline{X}_2^+$ 是$\overline{X}_2$的 Moore-Penrose 广义逆，I_{s_2+1}是 s_2+1 阶单位方阵，$\eta\in R^{s_3\times(s_2+1)}$是一个任意矩阵。(7)式求得的$\overline{W}_2$应满足下式：

$$J=\min\|\overline{W}_2\overline{X}_2-g_3(T)\|_F^2$$
$$=\min\|g_3(T)\overline{X}_2^+\overline{X}_2+\eta(I_{s_2+1}-\overline{X}_2\overline{X}_2^+)\overline{X}_2-g_3(T)\|_F^2$$
$$=\min\|g_3(T)(\overline{X}_2^+\overline{X}_2-I_k)\|_F^2 \tag{8}$$

其中 $\|\cdot\|_F$ 是 Frobenius 范数。由于 $\|\cdot\|_F\geq 0$，故(8)式 J 的最小值是 0，由广义逆的性质知如$\overline{X}_2$列满秩，即：

$$\mathrm{rank}(\overline{X}_2)=k \tag{9}$$

则$\overline{X}_2^+\overline{X}_2-I_k=0$，此时有 $J=0$，于是我们有：

定理 1 如果三层前馈网络隐层输出扩充矩阵列满秩，则该网络一定能实现精确映射。

由定理 1 所给出的充分条件，可以对网络的结构进行设计。结构设计主要是指确定隐层神经元个数(输入层和输出层的结点或神经元个数则可针对具体的应用问题来确定)。

由线性代数知(9)成立的一个必要条件是$\overline{X}_2$的行数不小于列数，即：

$$s_2\geq k-1 \tag{10}$$

这样我们就有如下准则 1：

准则 1 对任意给定的样本对 P、T，要保证三层前馈网络实现精确映射，其隐层神经元的个数不能小于样本数减 1。

另外，应当指出的是由(3)式知，为使 rank $\overline{X}_2$=k，除了应满足(10)式外，还应选择好$\overline{W}_1$及 f_2。文[6]给出了一种十分有效的方法。

针对情形 1 的具体算法就是选择$\overline{W}_1$满足(9)式，然后由(6)式求$\overline{W}_2$。

情形 2 隐层神经元个数由输出层个数决定。

由(6)式得：

$$[W_2, b_3]\begin{bmatrix}X_2\\E\end{bmatrix}=g_3(T)$$

即 $X_2=W_2^+(g_3(T)-b_3E)+(I_{s_2}-W_2^+W_2)\zeta$ (11)

式中：b_3 为输出层阈值向量，I_{s_2} 为 s_2 阶方阵，$\zeta\in R^{s_2\times k}$为任意矩阵。(11)式的 X_2 将使下式成立：

$$J=\min\|\overline{W}_2\overline{X}_2-g_3(T)\|_F^2=$$
$$=\min\|(W_2W_2^+-I_{s_3})(g_3(T)-b_3E)\|_F^2 \tag{12}$$

于是我们可以选择 W_2 行满秩，使(12)的 $J=0$，为使 W_2 行满秩，W_2 的行数 s_3 不应大于它的列数，即：

$$s_2\geq s_3 \tag{13}$$

显然，为实现精确映射，由(11)式求出的 X_2 还应满足(5)式，通过与引出(8)式类似的推导可得：

$$J_1=\min\|g_2(X_2)(\overline{X}_1^+\overline{X}_1-I_k)\|_F^2=0 \tag{14}$$

式中$\overline{X}_1^+$ 为$\overline{X}_1$的广义逆。显然，如果$\overline{X}_1$列满秩，则(14)恒成立，综合(13)与(14)我们有：

定理 2 如果三层前馈神经网络隐层与输出层间的权矩阵行满秩且输入样本扩充矩阵列满秩，则该网络一定能实现精确映射。

另外，$\overline{X}_1$列满秩的一个必要条件是 $s_1+1\geq k$，即：

$$s_1\geq k-1 \tag{15}$$

(13)和(15)式给出了结构设计的另一准则：

准则 2 对任意给定的样本对 P、T，要保证三层前馈网络实现精确映射，隐层神经元的个数不应少于输出层神经元个数且同时输入层神经元个数不少于样本个数减 1。

针对情形 2 的具体算法是选择$\overline{W}_2$使 rank $\overline{W}_2$=s_3，然后由(11)及(5)式求$\overline{W}_1$。

三、数值模拟实验

情形 1 的适应面较宽，如不限制隐层神经元的个数，则只要由(10)式选择 s_2，就一定能实现精确映射。结合情形 1 的算例结果见表 1。

情形 2 适用于输入样本维数很高，输出样本维数较低，而待训练的样本个数不是很多(满足(15)式)的情况，这时可以把隐层神经元个数选得较少((13)式)。结合情形 2 的算例结果见表 2。

表 1：输入节点数为 3，输出层神经元个数为 1，隐层神经元个数为样本个数减 1。BP 算法精度按 $J_{BP}=\frac{1}{2}\sum_{p=1}^{k}\|y^{(p)}-t^{(p)}\|_2^2<\epsilon$ 停止计算，其中取 $\epsilon=0.01k$。表 1 结果对应于情形 1。

表 2：输入结点 300，隐层、输出层各为 2 个神经元，样本个数由 50～300(间隔 50)。BP 算法的结果均是迭代 10000 次时的结果。表 2 结果对应于情形 2。

注：表 1 和表 2 中的样本由以下混沌 kent 序列生成：

$$x_{n+1}=\begin{cases} x_n/a & 0\leqslant x_n\leqslant a \\ (1-x_n)/(1-a) & a\leqslant x_n\leqslant 1 \end{cases}$$

其中：$a=0.2$，取初值 $x_1=0.3$，表 1 和表 2 的样本对分别为：

$(x_i,x_{i+1},x_{i+2})\to x_{i+3}$，$(x_i,\cdots,x_{i+299})\to(x_{i+300},x_{i+301})$；$i=1,2,\cdots$

表 1　　　　　　　　表 2

样本个数 k		100	150	200	250	300	350	50	100	150	200	250	300
新算法	精度 J	1.9×10^{-23}	2.8×10^{-23}	1.6×10^{-22}	5.3×10^{-21}	6.5×10^{-20}	8.3×10^{-20}	1.48×10^{-28}	9.75×10^{-28}	4.21×10^{-26}	9.26×10^{-24}	1.09×10^{-24}	1.18×10^{-24}
	耗时（秒）	4	8	16	23	35	46	3	8	15	30	48	68
BP 算法	精度 J_{BP}	0.99	1.49	1.98	2.50	3.00	2.48	0.0012	0.007	0.045	3.75	5.54	8.86
	耗时（分）	0.62	2.83	6.68	13.60	26.10	46.13	6.25	8.70	11.11	16.67	20	23.53

参考文献

1 Rumelhart D E, et al. Learning Representation by Back-Propagation Errors. Nature, 1986, 533～536

2 Haykin S. Neural Networks. Macmillan College Publishing Company, 1994. 190～192

3 Baldi P. Gradient Descent Learning Algorithm Overview: Greneral Dynamical Systems Perspective. IEEE Trans Nenral Networks, 1995, 6(1). 182～195

4 张立明. 人工神经网络的模型及应用. 复旦大学出版社, 1992. 44～46

5 焦李成. 神经网络系统理论. 西安电子科技大学出版社, 1990. 36

6 Zhang Daiyuan, Yu Juebang. An Algebraic Algorithm for 3-Layered Perceptron Training with δ Activation Function in Hidden Layer. In: Huang Xinhan, et al. eds. The Intl. Conf. on Artificial Intelligence for Engineering. 1998. 310～312

（上接第 58 页）

（a）　　　　（b）

图 1　位率 0.106bpp 时的 Lena 重建图像

参考文献

1 Barnsley M F, Sloan A D. A better way to compress images. BYTE Magazine, 1988, 13(1): 215～223

2 Jacquin A E. Image coding based on a fractal theory of iterated contractive image transformations. IEEE Trans on Image Processing, 1992, IP-1(1): 18～30

3 Fisher Y. Fractal Image Compression. New York: Springer-Verlage, 1995

4 Barnsley M F, Hurd L P. Fractal Image Compression. Wellesley: AK Peters, 1992

5 Yang Yangyi, et al. Projection-based spatially adaptive reconstruction of block-transform compressed Images. IEEE Trans on Image Processing, 1995, 4(7): 896～908

6 Chou J, et al. A simple algorithm for removing blocking artifacts in block-transform coded images. IEEE Signal Processing Letters, 1998, 5(2): 33～35

7 Jain A. Fundamentals of Digital Processing. New York: Van Nostrand Reinhold, 1993