

二阶神经网络映射能力的研究

李金艳 余英林

(华南理工大学无线电与自动控制研究所·广州, 510641)

摘要: 本文讨论了二阶神经网络的映射能力. 主要包括: (1) 从理论上严格地证明了二阶神经网络能以任意精度逼近任意连续函数. (2) 给出了二阶神经网络的 BP 算法. (3) 模拟实验结果. 模拟实验表明: 在中间隐层单元数目相同的条件下, 二阶神经网络的误差函数比一阶神经网络下降得快; 在误差精度相同的条件下, 二阶神经网络的隐层单元数目远比一阶神经网络少.

关键词: 二阶神经网络; 逼近; 二阶神经网络的 BP 算法

1 导 言

近几年来, 标准多层前馈神经网络(一阶)的映射能力受到极大的关注和深入的研究并在这方面取得了许多重要的成果. 例如, Hornik 和 White^[1]用平滑逼近的概念和 Hornik^[2]用 $L^p(u)$ 逼近的概念分别得到的结果; Ito^[3]用在 R^d 上一致逼近概念所取得的成果; Leshno, Lin, Pinkus 和 Schocken^[4]得到当网结结点激励函数更一般时的结果等等. 这些结果都是以不同的逼近方式或者不同的激励函数为立足点进行研究以后取得的, 唯一保持不变的是这些网络的连接权都是一阶的. 这是一个局限. 众所周知, 神经网络是基于人的大脑这个纯生物物体进行抽象而建立数学模型加以研究和分析的. 脑细胞之间的作用广泛又复杂, 一个脑细胞, 不仅有其他细胞对它的单一作用, 而且还有其他细胞对它的联合作用. 若把这个生物特性反映到神经网络上去, 那么很有必要为神经网络增加二阶连接权. 只有这样, 神经网络才能具有更强的映射能力, 更高的存储水平以及更广的容错范围. 映射能力强、存储水平高和容错范围广这些特点正是神经网络过渡到实用阶段所必备的条件, 综上所述, 对二阶神经网络的研究和分析将具有重要的理论意义和实用价值.

下面分三个部分进行讨论.

2 二阶神经网络的逼近理论

2.1 二阶三层前馈网络结构

二阶前馈网络的结构一般由输入层、输出层和一个或多个隐层所构成. 通常地, 我们只对具有单个隐层的高阶神经网络进行研究. 二阶三层前馈网络的大致结构如图 1 所示.

在这个图中, w_i 表示从隐层的第 i 个单元到输出单元的连接权, $u_{ij}^0 (i \neq j)$ 表示从隐层的第 i, j 个单元到输出单元的二阶连接权; w_{ij} 表示从输入层的第 j 个单元到隐层的第 i 个单元的连接权; $v_{ij}^k (i \neq j)$ 表示输入层的第 i, j 个单元到隐层的第 k 个单元的二阶连接权. $\sigma(x)$ 表示隐层单元的激励函数. 输入、输出层的激励函数都是线性的. θ_i 表示隐层单元的阈值. 从而有:

$$\begin{aligned}
 f(\bar{X}) = & \sum_{k=1}^m w_k \sigma \left(\sum_{j=1}^n x_j w_{kj} + \sum_{i \neq j} \sum x_i x_j v_{ij}^k + \theta_k \right) \\
 & + \sum_{r \neq s} \sum [v_{rs}^o \sigma \left(\sum_{j=1}^n x_j w_{rj} + \sum_{i \neq j} \sum x_i x_j v_{ij}^r + \theta_r \right) \sigma \left(\sum_{j=1}^n x_j w_{sj} \right. \\
 & \left. + \sum_{i \neq j} \sum x_i x_j v_{ij}^s + \theta_s \right)].
 \end{aligned} \quad (2.1)$$

当隐层到输出单元的连接权只有一阶权值时, (2.1) 式变为

$$f(\bar{X}) = \sum_{k=1}^m w_k \sigma \left(\sum_{j=1}^n x_j w_{kj} + \sum_{i \neq j} \sum x_i x_j v_{ij}^k + \theta_k \right). \quad (2.2)$$

在这里, m 表示隐层单元数目. 下面, 主要研究 (2.2) 式所反映的神经网络的映射能力. 有关隐层与输出单元之间存在高阶连接权时这方面的结果我们将在另一篇论文中给出. 所以本文讨论的主要问题是: 对一个任意的连续函数 $f(X)$, 当 m 充分大时, $\sum_{k=1}^m w_k \sigma \left(\sum_{j=1}^n x_j w_{kj} + \sum_{i \neq j} \sum x_i x_j v_{ij}^k + \theta_k \right)$ 是否能够一致逼近 $f(\bar{X})$? 下面给出肯定的证明.

2.2 逼近原理

定义 $\Sigma_n = \text{span} \left\{ \sigma \left(\sum_{j=1}^n x_j w_{kj} + \sum_{i \neq j} \sum x_i x_j v_{ij}^k + \theta \right) : w_{kj}, v_{ij} \in \mathbb{R}, i, j = 1, \dots, n/\theta \in \mathbb{R} \right\}$.

在这里, $\sigma(x) \in C^\infty$. (C^∞ 表示具有任意阶导数的所有连续函数所构成的集合).

引理 一个定义在紧致子集上的任意 S 元连续函数 $f(x_1, x_2, \dots, x_s)$ 可由 S 元多项式以任意精度逼近.

证明见参考文献[6].

定理 如果 $\sigma(x) \in C^\infty$ 且是一个非多项式, 那么 Σ_2 在 $C(A)$ 中稠密. ($A = [a, b] \times [c, d]$).

证 略.

推论 若 $\sigma(x) \in C^\infty$ 且是一个非多项式函数, 那么 Σ_n 在 $C(A)$ 中稠密. A 是 n 维欧几里得空间的一个紧致子集.

这个推论的证明是比较简单的. 证明过程与上述定理的证明过程类似. 这里不再详述.

这个推论的含义是: 对于任意小的正数 ϵ , 若 $f(\bar{X}) \in C(A)$, 那么必存在函数:

$$\sigma \left(\sum_{j=1}^n x_j w_{kj} + \sum_{i \neq j} \sum v_{ij}^k x_i x_j + \theta_k \right)$$

的一个线性组合使得

$$\sum_{k=1}^m [w_k \sigma \left(\sum_{j=1}^n x_j w_{kj} + \sum_{i \neq j} \sum x_i x_j v_{ij}^k + \theta_k \right) - f(\bar{X})] < \epsilon.$$

这里, m 代表隐层单元数目.

上面详细地讨论了当输入层与隐层之间增加二阶连接权时, 三层前馈网络可以映射任

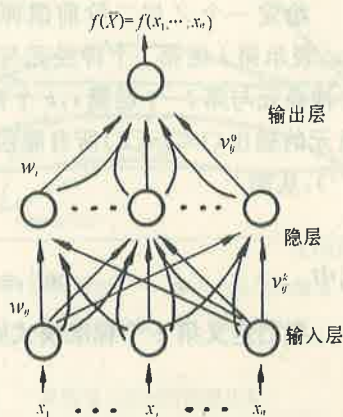


图 1 二阶三层前馈网络结构

意的连续函数,从定理的证明之中可明显看出当输入层与隐层之间增加高阶直至 n 阶连接权时,三层前馈网络同样具有这种性质(n 代表被映射函数的自变量个数).

3 二阶神经网络的 BP 算法

在第二部分的讨论之中,我们已经知道二阶神经网络能够以任意精度逼近任意连续函数.但目前还存在着一个问题就是:怎样去找到这些神经网的连接权?在一阶多层前馈网络中,我们是利用 BP(back propagation)算法去不断地修正神经网络的连接权值,使得网络输出与期望输出之间的误差不断减小.这就是所谓的梯度下降算法.对于二阶多层前馈网络,也有类似一阶的 BP 算法,下面详细给出.

给定一个 Z 层二阶前馈神经网络,输入层定义为第一层,……,输出层定义为第 Z 层. w'_{ij} 表示第 l 层第 i 个神经元与第 $l-1$ 层第 j 个神经元之间的连接权. w'_{ijk} 表示第 l 层第 i 个神经元与第 $l-1$ 层第 i, k 个神经元之间的二阶连接权. ($2 \leq l \leq Z$) x'_i 表示第 l 层第 i 个神经元的输出 ($1 \leq l \leq Z$) 所有隐层与输出层的激励函数为 sigmoid 函数 $\sigma(x)$, $\sigma(x) = 1/(1 + e^{-x})$, 从而

$$x'_i = \sigma(\text{net}'_i) \quad (3.1)$$

其中

$$\text{net}'_i = \sum_j w'_{ij} x'^{l-1}_j + \sum_{j \neq k} w'_{ijk} x'^{l-1}_j x'^{l-1}_k.$$

我们定义第 p 个标准模式的误差的能量函数为

$$E_p = \frac{1}{2} \sum_i (d_{pi} - x'^{l}_{pi})^2, \quad (3.2)$$

那么,总的误差能量函数为

$$E = \sum_p E_p = \frac{1}{2} \sum_p \sum_i (d_{pi} - x'^{l}_{pi})^2. \quad (3.3)$$

应用梯度下降法,我们得到

$$\Delta_p w'_{ij} = -\eta \frac{\partial E_p}{\partial w'_{ij}} \quad \text{和} \quad \Delta_p w'_{ijk} = -\eta \frac{\partial E_p}{\partial w'_{ijk}}.$$

1) 对于输出层,即 $l=Z$ 时,我们可得

$$\Delta_p w'_{ij} = \eta (d_{pi} - x'^{l}_{pi}) \sigma'(\text{net}'_{pi}) x'^{l-1}_{pj} \quad (3.4)$$

和

$$\Delta_p w'_{ijk} = \eta (d_{pi} - x'^{l}_{pi}) \sigma'(\text{net}'_{pi}) x'^{l-1}_{pj} x'^{l-1}_{pk} \quad (3.5)$$

2) 对于隐层,即 $2 \leq l \leq Z-1$, 我们得到

$$\Delta_p w'_{ij} = \eta \sigma'(\text{net}'_{pi}) \sum_k \delta'^{l+1}_{pk} [w'^{l+1}_{ki} + \sum_{j \neq i} w'^{l+1}_{kij} x'^{l-1}_{pj}] x'^{l-1}_{pj} \quad (3.6)$$

和

$$\Delta_p w'_{ijk} = \eta \sigma'(\text{net}'_{pi}) \sum_k \delta'^{l+1}_{pk} [w'^{l+1}_{ki} + \sum_{j \neq i} w'^{l+1}_{kij} x'^{l-1}_{pj}] x'^{l-1}_{pj} x'^{l-1}_{pk}. \quad (3.7)$$

这里 $\delta'_{pi} = -\frac{\partial E_p}{\partial \text{net}'_{pi}}$. 总之

$$\Delta w'_{ij} = \sum_p \Delta_p w'_{ij}, \quad (3.8)$$

$$\Delta w'_{ijk} = \sum_p \Delta_p w'_{ijk}. \quad (3.9)$$

在实际应用过程中,有不少改进的一阶网络 BP 算法,当然,对于二阶网络也存在着改进的二

阶网络 BP 算法. 这将是以后研究内容.

4 模拟结果

利用二阶神经网络的 BP 算法, 我们在中间隐层单元数目和网络误差的下降速度两方面对二阶神经网络和一阶神经网络进行了比较. 模拟实验表明: 高阶神经网络具有很多优点.

4.1 一阶与二阶网络对 XOR 函数实现的比较

比较一阶与二阶网络的性能的第一个例子分别由它们去实现 XOR 函数. 训练集如下: 输入数据集 $\{(0,0), (0,1), (1,0), (1,1)\}$, 相应地, 期望输出集的 $\{0, 1, 1, 0\}$. 当一阶网的一阶连接权值、隐层单元数目和训练样本都与二阶网相同时, 用 BP 算法分别训练它们. 我们可以看出二阶网在收敛速度方面优于一阶网. (详见表 1, 实验数据由 Dx386-33 机器给出).

4.2 随机训练样本实验

随机地产生一个四个元素的训练集. 分别由二阶网和一阶网去训练它们. 在训练次数相同的情况之下, 二阶网的训练误差是最小的 (详见表 2). 同时, 从表 2 中可以看出随着一阶网隐层单元数的增加, 它们的训练误差不断地下降.

表 1 训练时间与误差的比较

网络结构(输入-隐层-输出)	训练时间	训练次数	训练误差
2-3-1(一阶)	15 秒	3128	0.006056
2-3-1(二阶)	15 秒	2885	0.001136
2-3-1(一阶)	30 秒	6332	0.002751

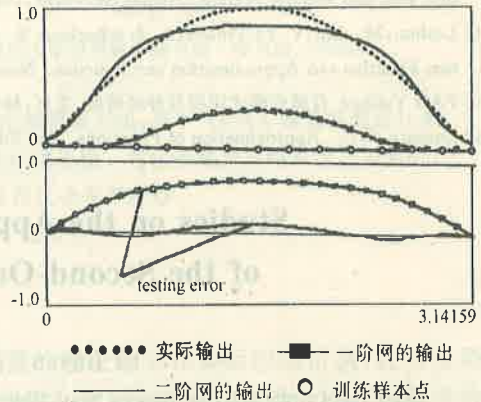


图 2 一阶网与二阶网的性能比较

表 2 中间隐层单元数量比较

网络结构	训练次数	训练误差
2-10-1(一阶)	5000	0.007619
2-15-1(一阶)	5000	0.007470
2-20-1(一阶)	5000	0.007366
2-10-1(二阶)	5000	0.005604

4.3 函数 $\sin(x)$ 的逼近实验

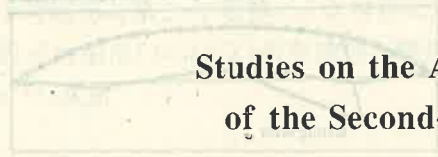
训练集为 $\{\sin(\pi/14 * i); i = 0, 1, 2, \dots, 14\}$. 分别由一阶网(1-5-1)和二阶网(1-5-1)去训练它们. 图 2 显示 2 个网络应用 BP 算法训练 2000 次以后的结果. 可以看出二阶网的训练误差和测试误差精度都优于一阶网.

5 结 论

本文从一个方面讨论了二阶神经网络的映射能力问题. 证明了二阶神经网络能够以任意精度逼近任意连续函数. 实验也表明高阶网具有很多优点, 以后, 我们将进一步研究高阶神经网络的逼近理论、存储能力及其应用.

参 考 文 献

- [1] Hornik, K., Stinchcombe, M. & White, H. . Multilayer Feedforward Networks Are Universal Approximators. *Neural Network*, 1989, 2; 359—366
- [2] Hornik, K. . Approximation Capabilities of Multilayer Feedforward Network. *Neural Network*, 1991, 4; 251—257
- [3] Ito, Y. . Approximation of Continuous Functions on R^d by Lineal Combination of Shifted Rotations of a Sigmoid Function with and without Scaling. *Neural Networks*, 1992, 5; 105—116
- [4] Leshno, M. , Lin, V. Ya, Pinkus, A. & schocken, S. . Multilayer Feedforward Networks with Non-Polynomial Activation Function can Approximation any Function. *Neural Networks*, 1992, 6; 861—867
- [5] PAO Yohhan. 自适应模式识别与神经网络. 北京: 科学出版社, 1992, 50—70
- [6] Lorentz, G. G. . Approximation of Functions. 2nd Edition, New York: Chelsea Publishing Company, 1986, 10—11



Studies on the Approximation Capability of the Second-Order Neural Network

LI Jinyan and YU Yinglin

(Institute of Radio and Automation, South China University of Technology • Guangzhou, 510641, PRC)

Abstract: The approximation capability of the second-order neural network is investigated in this paper and the following results have been obtained: 1) It has been proved that the second-order neural network can approximate any continuous function with any degree of accuracy; 2) the BP algorithm for second-order neural network and the simulated results are given in this paper. The simulated experiments show that when the number of hidden neurons in the second-order neural network is equal to the first-order one's, the error of the second-order neural network decreases faster than the first-order one's; when the accuracies of both of the second-order and first-order neural networks are equal, the number of hidden neurons in the second-order is far smaller than the first-order one's.

Key words: second-order neural network; approximation; BP algorithm for the second-order neural network

本文作者简介

李金艳 1969年生, 华南理工大学无线电与自动控制研究所博士研究生, 主要研究领域是高阶神经网络。

余英林 见本刊1996年第2期第190页。