

基于跨越连接的多层前馈神经网络结构分析

唐云岚^{1,2}, 高妍方¹, 谭旭¹, 陈英武¹

TANG Yun-lan^{1,2}, GAO Yan-fang¹, TAN Xu¹, CHEN Ying-wu¹

1.国防科学技术大学 信息系统与管理学院, 长沙 410073

2.武警工程学院 通信工程系, 西安 710086

1.Institute of Information System and Management, National University of Defense Technology, Changsha 410073, China

2.Department of Communication Engineering, Engineering College of the Chinese People's Armed Police Force, Xi'an 710086, China

E-mail: tcloudyl@163.com

TANG Yun-lan, GAO Yan-fang, TAN Xu, et al. Research on structure of multi-layer feed-forward neural network with cross connections. *Computer Engineering and Applications*, 2009, 45(32): 45-47.

Abstract: Neural networks with any kind of connections can always be sorted as cross-connected ones. According to traditional multi-layer feed-forward neural network, this paper elaborates the concept of completely-fully connected neural network and then puts forward a cross-connected multi-layer feed-forward neural network algorithm. By analyzing the error function of multi-layer feed-forward neural network, it can be theoretical proved that the cross-connected neural network can reach ideal results with more concise framework compared with the non-cross connected neural network. Lastly, the proposed algorithm is tested on the XOR problem, which is well solved by using only one hidden neuron.

Key words: cross connections; multi-layer feed-forward neural network; structure of hidden layer; XOR problem

摘要:任何连接方式的神经网络总可以归结为跨越连接网络。在传统多层前馈神经网络算法的基础上,提出了完全全连接神经网络的概念,给出了基于跨越连接的多层前馈神经网络算法。通过分析多层前馈神经网络的误差函数,从理论上证明了:相对于无跨越连接网络,基于跨越连接的多层前馈神经网络能以更加简洁的结构逼近理想状态。最后,用一个隐层神经元解决了 XOR 问题。

关键词:跨越连接;多层前馈神经网络;隐层结构;XOR 问题

DOI: 10.3778/j.issn.1002-8331.2009.32.014 文章编号: 1002-8331(2009)32-0045-03 文献标识码: A 中图分类号: TP13

具有误差反向传播学习功能的多层前馈神经网络具有结构严谨、工作状态稳定、可操作性强等特点,同时由于隐层节点的引入,使得一个三层非线性网络可以以任意精度逼近任何连续函数,因而在模式识别、非线性映射、预测等许多领域得到了广泛的应用。但是,神经网络结构优化问题一直是国内外研究的热点。网络的输入节点数和输出节点数可由实际问题决定,网络结构优化的重点与难点在于隐层结构的设计,具体是指隐层的数目和各隐层的神经元数目。确定隐层的结构在很大程度上决定着网络的记忆容量、泛化能力、训练速度和输出响应的质量。隐层节点数过少,学习可能不收敛,网络的识别能力和泛化能力降低;隐层节点数过多,往往偏向于冗余,训练长时间不收敛,网络的容错性能下降。如何以较少的隐层节点数实现较优的网络性能,成为了网络结构设计中的一大难题。

传统的多层前馈神经网络是采用全连接的方式进行计算的。这种全连接中,上层节点与下层的每一个节点均有连接通路,但不存在跨层连接,因此,所谓的全连接是相邻层节点之间的全连接,是一种非完全的全连接。在实际问题中,系统的内部机理十

分复杂,各种因素(节点)错综复杂,相互制约,采用传统的非完全全连接神经网络建模,难以准确描述事物相互关联的本质,因此,往往建立结构异常复杂的网络也难以获得理想的结果。

在借鉴传统的多层前馈神经网络算法基础上,提出了完全全连接神经网络的概念,给出了基于跨越连接的多层前馈神经网络算法。通过对多层前馈神经网络误差函数的分析,从理论上证明了:相对于无跨越连接网络,基于跨越连接的多层前馈神经网络能以更加简洁的结构逼近理想状态。最后,用一个隐层神经元解决了 XOR 问题。

1 神经网络的连接方式

神经网络的连接方式包括前馈、跨越、侧向、反馈等四种^[1],如图 1。

其中,前馈连接是最基本的连接方式,它普遍存在于其余三种连接方式中。而有侧向连接的网络总可以转化为有跨越连接和反馈连接的网络(图 2);反馈连接网络可以理解为连接权值为负的前馈连接和跨越连接网络。因此,任何连接方式的神

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.70272002);高等院校博士学科点专项科研基金(the China Specialized Research Fund for the Doctoral Program of Higher Education under Grant No.20059998019)。

作者简介:唐云岚(1979-),男,博士研究生,主要研究方向为神经网络、多目标进化算法、计算智能等。

收稿日期:2008-06-27 修回日期:2008-10-09

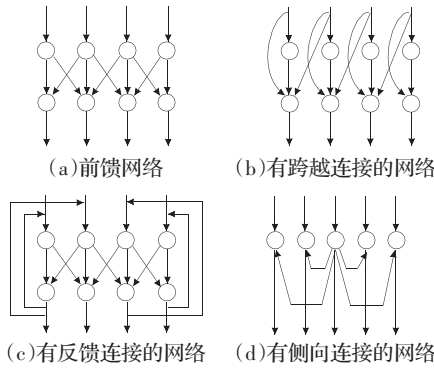


图1 神经网络结构的几种基本形式

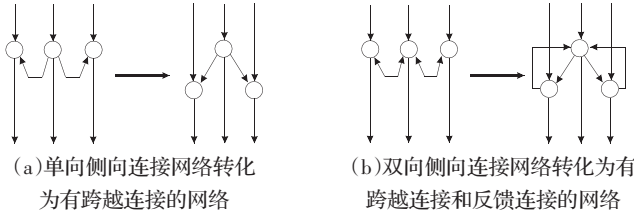


图2 侧向连接网络的转化

神经网络总可以归结为跨越连接网络。

显然,跨越连接神经网络不仅是相邻层节点之间的全连接,也是非相邻层节点之间的全连接,因此,是一种完全的全连接神经网络。

2 基于跨越连接的多层前馈神经网络算法

设基于跨越连接的多层前馈神经网络有 S 个隐层。输入层为 I , 有 I 个输入信号, 其中任一输入信号用 i 表示; 第 s 个隐层为 M_s , 有 M_s 个神经元, 其中任一神经元用 m_s 表示; 输出层为 O , 有 O 个输出神经元, 其中任一神经元用 o 表示。层 $P(P \in \{I, M_1, M_2, \dots, M_S\})$ 与任一后续层 $Q(Q \in \{M_1, M_2, \dots, M_S, O\})$ 的连接权值为 w_{pq} 。神经元的输出用 u 表示, 激励输出用 v 表示, u 和 v 的上标表示层, 下标表示层中某个神经元。输入的学习样本集为 $X=[X_1, X_2, \dots, X_k, \dots, X_N]$, 对应任一学习样本为 $X_k=[x_{k1}, x_{k2}, \dots, x_{ki}, \dots, x_{kI}]$ ($k=1, 2, \dots, N$), 与 X_k 相对应的网络实际输出为 $Y_k=[y_{k1}, y_{k2}, \dots, y_{ko}]$, 期望输出为 $d_k=[d_{k1}, d_{k2}, \dots, d_{ko}]$, n 为迭代次数, 则权值和实际输出是 n 的函数。基于跨越连接的多层前馈神经网络学习过程可以描述如下:

2.1 工作信号正向传播

输入信号从输入层经隐层单元, 传向输出层, 在输出端产生输出信号, 这是工作信号的正向传播。在信号向前传递的过程中网络的权值是固定不变的, 与传统多层前馈神经网络算法不同, 每一层神经元的状态不仅只影响下一层神经元的状态, 还影响与其有跨越连接的神经元的状态。如果在输出层不能得到期望的输出, 则转入误差信号反向传播。正向传播的工作信号为:

第 1 个隐层 M_1 的输入和输出为:

$$u_{m_1} = \sum_{i=1}^I w_{im_1} x_{ki}, v_{m_1} = f\left(\sum_{i=1}^I w_{im_1} x_{ki}\right), m_1=1, 2, \dots, M_1 \quad (1)$$

第 s 个隐层 M_s 的输入和输出为 ($s=2, 3, \dots, S$):

$$u_{m_s} = \sum_{i=1}^I w_{im_s} x_{ki} + \sum_{i=1}^{s-1} \sum_{m_i=1}^{M_i} w_{m_i m_s} v_{m_i}, v_{m_s} = f(u_{m_s}), m_s=1, 2, \dots, M_s \quad (2)$$

输出层 O 的输入和网络输出为:

$$u_o = \sum_{i=1}^I w_{io} x_{ki} + \sum_{i=1}^S \sum_{m_i=1}^{M_i} w_{m_i o} v_{m_i}, y_{ko} = v_o = f(u_o), o=1, 2, \dots, O \quad (3)$$

输出层第 o 个神经元的误差信号为 $e_{ko}(n) = d_{ko}(n) - y_{ko}(n)$ 。

定义神经元 p 的误差能量为 $\frac{1}{2} e_{ko}^2(n)$, 则输出层所有神经元的

误差能量总和为 $E = \frac{1}{2} \sum_{o=1}^O e_{ko}^2(n)$ 。

2.2 误差信号反向传播

网络的实际输出与期望输出之间的差值 $e_{ko}(n)$ 即为误差信号, 误差信号由输出端开始逐渐向回传播, 这是误差信号的反向传播。在误差信号反向传播过程中, 网络的权值由误差反馈进行调节。通过权值的不断修正使网络的实际输出更接近期望输出。反向传播中每一层神经元的局部梯度为:

输出层 O 的局部梯度为:

$$\delta_o^O(n) = y_{ko}(n)(1 - y_{ko}(n))(d_{ko}(n) - y_{ko}(n)), o=1, 2, \dots, O \quad (4)$$

第 S 个(即最后一个)隐层 M_S 的局部梯度为:

$$\delta_{m_S}^M(n) = v_{m_S}^M(n)(1 - v_{m_S}^M(n)) \sum_{o=1}^O \delta_o^O(n) w_{m_S o}(n), m_S=1, 2, \dots, M_S \quad (5)$$

第 s 个隐层 M_s 的局部梯度为 ($s=1, 2, \dots, S-1$):

$$\delta_{m_s}^M(n) = v_{m_s}^M(n)(1 - v_{m_s}^M(n)) \left[\sum_{i=s+1}^S \sum_{m_i=1}^{M_i} \delta_{m_i}^M(n) w_{m_s m_i}(n) + \sum_{o=1}^O \delta_o^O(n) w_{m_s o}(n) \right], m_s=1, 2, \dots, M_s \quad (6)$$

通过局部梯度调整网络权值为:

第 S 个(即最后一个)隐层与输出层的连接权值 $w_{m_S o}$ 调整为:

$$\Delta w_{m_S o}(n) = \eta \delta_o^O(n) v_{m_S}^M(n), w_{m_S o}(n+1) = w_{m_S o}(n) + \Delta w_{m_S o}(n), m_S=1, 2, \dots, M_S, o=1, 2, \dots, O \quad (7)$$

第 s 个隐层与第 $s+1$ 个隐层的连接权值 $w_{m_s m_{s+1}}$ 调整为 ($s=1, 2, \dots, S-1$):

$$\Delta w_{m_s m_{s+1}}(n) = \eta \delta_{m_{s+1}}^M(n) v_{m_s}^M(n), w_{m_s m_{s+1}}(n+1) = w_{m_s m_{s+1}}(n) + \Delta w_{m_s m_{s+1}}(n), m_s=1, 2, \dots, M_s, m_{s+1}=1, 2, \dots, M_{s+1} \quad (8)$$

输入层与第一隐层的连接权值 w_{im_1} 调整为:

$$\Delta w_{im_1}(n) = \eta \delta_{m_1}^M(n) x_{ki}(n), w_{im_1}(n+1) = w_{im_1}(n) + \Delta w_{im_1}(n), i=1, 2, \dots, I, m_1=1, 2, \dots, M_1 \quad (9)$$

3 结构分析

刘耦耕^[2]提出由多层前馈神经网络误差函数构成的非线性方程组的独立方程个数和网络待求未知变量的个数应该相等, 由此导出了无跨越连接的多层前馈人工神经网络隐层层数判别式为:

$$N_2 \leq \text{int} \left[\frac{J(K-1) - (I-1)}{2} \right] \quad (10)$$

在有跨越连接的多层前馈人工神经网络中, 假定输入层神经元个数为 I , 输出层神经元个数为 J , 隐层层数为 N_1 ; 第 n 个隐层神经元个数为 $M_n, n=1, 2, \dots, N_1$; 标准样本个数为 K , 第 k 个样本为 x_k , 期望输出为 d_k , 实际输出为 $y_k, k=1, 2, \dots, K$ 。于是, 输入向量写为 $X_k = \{x_{ki}\}, i=1, 2, \dots, I$; 期望输出向量写为 $D_k = \{d_{ko}\}$, 实际输出向量为 $Y_k = \{y_{ko}\}, j=1, 2, \dots, J$ 。

设待求的神经元权值和阈值为未知量, 则网络第 1 隐层对

应未知量个数为 $(I+1)M_1$, 网络第 2 隐层对应未知量个数为 $(I+1+M_1)M_2$, 网络第 n 隐层对应未知量个数为 $(I+1+M_1+M_2+\dots+M_{n-1})M_n$, 网络输出层对应未知量个数为 $(I+1+\sum_{i=1}^{N_i} M_i)J$. 因此, 总的未知量个数 L 可表示为:

$$L=(I+1)M_1+\sum_{p=1}^{N_i-1}\left[(I+1+\sum_{i=1}^p M_i)M_{p+1}\right]+(I+1+\sum_{i=1}^{N_i} M_i)J \quad (11)$$

L 个未知量 z_1, z_2, \dots, z_L 构成待求向量, 则实际输出是未知向量 $\mathbf{Z}=[z_1, z_2, \dots, z_L]^T$ 的函数, $\mathbf{Y}_k=\{y_k^j(\mathbf{Z})\}$. 因此, 总的平均误差函数为:

$$E(\mathbf{Z})=\frac{1}{2K}\sum_{k=1}^K\sum_{j=1}^J(d_k^j-y_k^j(\mathbf{Z}))^2 \quad (12)$$

在多层前馈神经网络算法中, 误差函数对权值的梯度方向修正权值, 最终找到一组待求向量 \mathbf{Z} 值, 通过网络学习训练, 使实际输出逼近期望输出, 误差函数 $E(\mathbf{Z})$ 逼近于零. 换句话说, 按照理想状态有:

$$E(\mathbf{Z})=\frac{1}{2K}\sum_{k=1}^K\sum_{j=1}^J(d_k^j-y_k^j(\mathbf{Z}))^2=0 \quad (13)$$

展开可得 $K \cdot J$ 个非线性方程构成的非线性超越代数方程组, 其方程总个数 S 记为:

$$S=K \cdot J$$

多层前馈神经网络误差函数构成的非线性方程组的独立方程个数和网络待求未知变量的个数应该相等, 导出有跨越连接的多层前馈人工神经网络结构方程式:

$$(I+1)M_1+\sum_{p=1}^{N_i-1}\left[(I+1+\sum_{i=1}^p M_i)M_{p+1}\right]+(I+1+\sum_{i=1}^{N_i} M_i)J=K \cdot J \quad (14)$$

进而导出隐层数判别式:

$$N_i \leq \text{int}\left[\frac{\sqrt{4I^2+4J^2+4I-4J+8KJ+1}-(2I+2J+1)}{2}\right] \quad (15)$$

令 $N_1 \leq N_2$, 则可导出一个输入输出层神经元个数与样本个数的关系式:

$$K \geq \frac{\sqrt{4I^2+4J^2-4I-4J+1}+(2-I)}{J} \quad (16)$$

一般情况下, 上式均成立. 因此, 可以认为, 相对于无跨越连接网络, 跨越连接网络能以更加简洁的结构逼近神经网络的理想状态.

4 仿真实验

XOR (异或) 问题在神经网络研究中经常用来作为检验网

络性能的例子. 众所周知, 只有一个隐元的传统 S-型网络不能解决 XOR 问题^[3]. 但是, 如果采用基于跨越连接的多层前馈神经网络, 那么只用一个隐元就能解决这个问题. 图 3 是这种网络的结构, 式(17)是网络达到收敛时的一组权值数据:

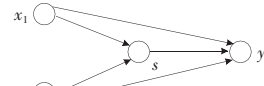


图3 求解 XOR 问题的跨越连接网络

$$W=\{w_{x_1-s}, w_{x_1-y}, w_{x_2-s}, w_{x_2-y}, w_{s-y}\} =$$

$$\{13.124\ 8, -28.771\ 6, 13.124\ 6, -28.464\ 5, 55.024\ 9\} \quad (17)$$

传统的多层前馈神经网络至少需要两个隐层神经元才能解决 XOR 问题. 设训练精度为 $1e-10$, 隐层和输出层激活函数为常规的 S-型激活函数, 则两种网络的性能比较如表 1 所示.

表 1 两种网络求解 XOR 问题时的性能比较

网络类型	网络结构	实验收敛次数	平均/最大/最小迭代次数
基于跨越连接的多层前馈神经网络	2-1-1	20/17	19.06/25/15
传统的多层前馈神经网络	2-2-1	20/9	20.22/13/30

由表 1 可知, 尽管基于跨越连接的多层前馈神经网络只采用一个隐元, 但收敛性能明显优于传统网络, 且具有更好的稳定性.

5 结束语

基于跨越连接的多层前馈神经网络, 通过对连接方式的改进, 摒弃了基于非完全全连接的传统神经网络结构模式, 有助于降低网络的结构复杂性, 为神经网络结构优化提供了一种新的思路. 仿真实验表明, 该网络的收敛性和稳定性较传统网络有较大提高. 当然, 该文仅以经典的 XOR 问题说明了上述性能的改善, 在进一步的研究中, 将试图结合具体的工程实践问题及其他经典测试函数, 拓展基于跨越连接的多层前馈神经网络在各个领域的应用.

参考文献:

- [1] 加卢什金. 神经网络理论[M]. 北京: 清华大学出版社, 2002.
- [2] 刘耦耕, 李圣清, 肖强晖. 多层前馈人工神经网络结构研究[J]. 湖南师范大学自然科学学报, 2004, 27(1): 26-30.
- [3] 吴佑寿, 赵明生. 激活函数可调的神经元模型及其有监督学习与应用[J]. 中国科学: E 辑, 2001, 31(3): 263-272.

(上接 26 页)

- [3] Ratnaweera A, Halgamuge S K, Watson H C. Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients[J]. IEEE Trans on Evolutionary Computation, 2004, 8(3): 240-255.
- [4] Wang L F, Singh C. Environmental/economic power dispatch using a fuzzified multiobjective particle swarm optimization algorithm[J]. Electr Power Syst Res, 2007, 77(12): 1654-1664.
- [5] Wang L F, Singh C. Multicriteria design of hybrid power generation systems based on a modified particle swarm optimization algorithm[J].

IEEE Transactions on Energy Conversion, 2009, 24(1): 163-172.

- [6] He P, Jagannathan S. Reinforcement learning neural network based controller for non-linear discrete time systems with input constraints[J]. IEEE Trans on Syst, Man, and Cybern, Part B: Cybern, 2007, 37(2): 425-436.
- [7] Sharma K D, Chatterjee A, Rakshit A. A hybrid approach for design of stable adaptive fuzzy controllers employing Lyapunov theory and particle swarm optimization[J]. IEEE Transactions on Fuzzy Systems, 2009, 17(2): 329-341.