

# 高阶神经网络连接权的稀疏化及其删减算法<sup>1</sup>

李守丽 李金艳\* 李望超

(河北工业大学计算机系 天津 300130)

\*(The University of Melbourne, Australia)

**摘要** 本文首先研究完全连接型高阶神经网络的逼近能力,并证明了定义在 $\{0,1\}^N$ 上的任意布尔函数都可以由完全连接的高阶神经网络来实现。接着提出了旨在简化网络结构的去除冗余连接权删减算法,并用于高阶神经分类器的稀疏化实现。模拟实验结果证明了这种算法的有效性。

**关键词** 高阶神经网络, 冗余连接权, 稀疏化连接, 删减算法

**中图分类号** TN-052

## 1 引言

对传统的一阶神经网络的研究,一直受到广泛的关注和重视。相对于一阶神经网络来说,高阶神经网络具有较大的存储容量<sup>[1,2]</sup>,高阶连接权的引入可加快网络的收敛速度<sup>[3]</sup>,还有一些研究成果表明,高阶神经网络具有语义推断和目标检测的功能<sup>[4,5]</sup>,此外,S.Amari教授所提出的高阶神经网络逼近理论具有开创性质,开辟了神经网络研究的新领域<sup>[6]</sup>,等等。尽管有关高阶神经网络的这类报道很多,但普遍使用的都是完全连接型或稠密型高阶神经网络,在神经元数量较多的情况下,常常会由于连接权数目过于巨大而给网络实现造成困难。为克服这方面的困难,本文就连接权稀疏化的高阶神经网络进行研究,并提出了去除冗余连接权的删减算法。模拟结果表明,所提出的算法对冗余连接权的删除是相当有效的,借此可设计出网络规模尽可能小的网络结构而不降低网络的整体性能。

## 2 完全连接型和稀疏连接型的高阶神经元

### 2.1 全连接的高阶神经元

全连接高阶神经元的构成框架如图1所示,图中左侧的矩形框内表示神经元的三个基本输入项,右侧矩形框内表示高阶扩展项。图1所表达的主要特性是高阶扩展项的引入。

### 2.2 稀疏化的高阶神经元

稀疏化的高阶神经元的构成框架如图2所示,图中左侧表示基本输入项,右侧表示高阶扩展项,可以看出,在左右两个矩形框内都有被删去的连接权。

## 3 用高阶网络实现任意布尔函数<sup>[7,8]</sup>

基于以上的描述,作者提出了一个引理和一个定理。根据该引理和定理,可以导出以下结果:

- (1) 对布尔函数的逼近能力
- (2) 决定所有高阶权值的递推算法
- (3) 各阶连接权的分布状况

<sup>1</sup> 1997-08-25 收到, 1998-07-26 定稿

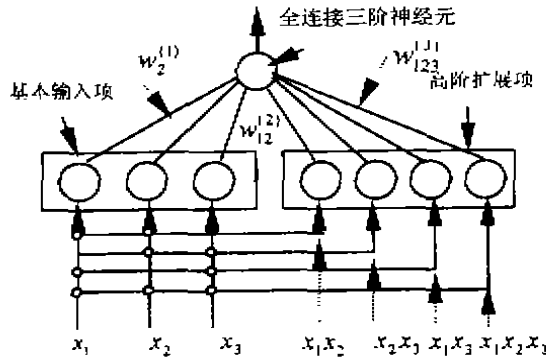


图 1 全连接三阶神经元

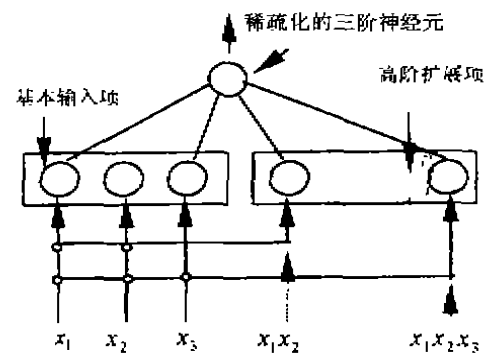


图 2 稀疏化的三阶神经元

限于篇幅, 这里略去上述有关结果的细节, 对这些内容有兴趣的读者可查阅参考文献 [7,8].

#### 4 实现稀疏化的删减算法

首先做这样的约定: 如果一个连接权为零, 那么可以把它理解为已被消除。由于完全连接型高阶神经分类器之中总有一些不太有用的连接权, 而大量数目的连接权又构成了 VLSI 硬件实现的极大困难。因而有必要去稀疏化完全连接型高阶神经分类器。消除原则主要是基于绝对值较小的连接权就应该删除这个思想。根据这个思想, 删除过程描述如下:

第一步 根据“递推算法”计算出完全连接型高阶神经网络的所有初始权值。

第二步 选出具有最大绝对值的连接权, 然后把所有的连接权都除以这个连接权。明显地, 正确分类的条件仍然具备。

第三步 修改连接权:  $w_i^{k+1} = w_i^k + \Delta w_i^k$ 。

第四步 选出具有最大绝对值的连接权, 然后把所有的连接权都除以这个连接权。

第五步 消除不重要的连接权。即  $w \rightarrow 0$ , 如果  $|w| < T$ ,  $T > 0.0$ 。其中  $T$  表示阈值 (Threshold)。

第六步 回到第三步直到条件 1(见下文) 被满足。如果满足了条件 3, 结束这个算法。

第七步 修改所有连接权:  $w_i^{k+1} = w_i^k + \Delta w_i^k$ 。

第八步 选出具有最大绝对值的连接权, 然后把所有的连接权都除以这个连接权。

第九步 回到第七步直到条件 2 被满足。

第十步 回到第三步。

条件 1 训练集之中首次有一个分类发生差错。

条件 2 如训练集中没有任何分类发生差错。

条件 3 被消除的连接权的数目超过限值。

#### 5 模拟结果

将上述算法用于稀疏化一个带有 9 个输入单元的完全连接型高阶神经分类器, 所得结果如下所述。

最初, 这个完全连接型高阶神经分类器共有 512 条连接权. 经过上述稀疏化算法的删减, 一般只剩下 30 至 60 条连接权. 具体剩下多少条连接权要视所选择的参数而定. 详细的实验结果可参看表 1. 由表中可见, 经过 9 个删减循环之后, 只剩下 34 条连接权. 这表明这个稀疏化算法是很有效的.

表 1 删减算法执行过程中非零权减少的情况

循环	1	2	3	4	5	6	7	8	9	10
非零权	251	211	151	130	80	80	39	38	34	4
误差	0.000539	0.000516	0.000516	0.000516	6.19998	0.000516	0.000531	0.000525	0.000516	2.50012

## 6 讨论和注解

(1) 神经网络稀疏化这个思想起源于稀疏矩阵的概念. 如果一个  $n \times n$  矩阵的非零元素远远小于  $n^2$ , 那么它可以称为一个稀疏矩阵. 然而本章中所用的稀疏化概念在于描述简化神经网络结构的过程.

(2) 所提出的删减算法之中, 第二步和第八步是必要的. 没有这两步可能会导致算法的收敛性恶化, 这可以从以下得到解释. 首先, 即便是所有的连接权除以或乘以一个正实数, 高阶神经分类器的输出不会有任何变化. 因为输出单元的传递函数是  $\sigma(x)$ . 这里,  $\sigma(x) = 1$ , 如果  $x > 0$ ; 否则  $\sigma(x) = 0$ . 第二, 因为所有完全连接型高阶神经单元的连接权都是由一个递推算法唯一地决定, 具有最大绝对值的连接权可能会达到 100 或 1000, 而最小的才是 1 或 0. 第三, 如果不包括第二步和第八步, 明显地  $f(x)$  会变得很大. 因而即便是  $\Delta w_{\{n(i)\}_k}^{(k)} = 0$ , 训练集中仍然会存在错误的分类. 例如: 如果训练集中有一些错误分类, 那么  $\text{error}(x) \geq 0.5$  或者  $\text{error}(x) \rightarrow 1$ . 在  $\text{error}(x) \rightarrow 1$  的情况之下, 有  $1 - \text{error}(x) \rightarrow 0$ . 因为  $1 - \text{error}(x)$  变得非常之小, 以致于在数字模拟实验中并不能保证  $1 - \text{error}(x)$  不等于零. 而事实上,  $1 - \text{error}(x)$  是非零的.

(3) 控制稀疏算法的参数如  $c$  和  $T$  都是重要的. 为了避免因太多的删减而引起分类错误, 选择  $T$  时要适当, 不能太小. 另外, 参数  $c$  也不能太小.

(4) 本章删减连接权的标准是依赖于连接权的绝对值. 我们认为删减连接权的标准应更加普遍化, 尽管所提出的稀疏算法很有效. 如: 当训练集中有分类错误时, 是不是可以把权值最大的连接权删去? 在一般的神经网络系统之中, 如何定义“最不重要连接权”等等. 诸如此类问题都有待于进一步研究.

## 7 结 论

(1) 本文证明了完全连接型的高阶神经元能够逼近定义在  $\{0, 1\}^N$  上的任意布尔函数.

(2) 完全连接型的高阶神经元的连接权可全部和唯一地由本文首次提出的递推算法所决定. 这个算法的最重要的特点是: 只要一个布尔函数的真值表给定, 神经单元的连接权就可以直接算出.

(3) 所提出的这个稀疏化算法, 主要用于删减完全连接型高阶神经分类器之中的不太重要的连接权. 模拟实验结果表明: 这个稀疏化算法对于消除多余的连接权是有很有效的. 从而可得到尽可能小的网络结构.

## 参 考 文 献

- [1] Paretto P, Niez J J. Long term memory storage capacity of multiconnected neural networks, *Biol. Cybern.*, 1986, 54(3): 53-63.
- [2] Baldi P. Neural networks, orientations of the supercube and algebraic threshold functions, *IEEE Trans. on Inform. Theory*, 1988, IT-34 (3): 523-530.
- [3] Giles C L, Maxwell T. Learning, invariance, and generalization in high-order neural networks, *Applied Optics*, 1987, 26(23): 4972-4978.
- [4] Giles C L, Chen D, Miller C B, *et al.* Second-order recurrent neural networks for grammatical inference, *Proc. Int. Joint Conf. Neural Networks, IJCNN91, Seattle: vol.2, 1991, 273-281.*
- [5] Liou R, Azimi-Sadjadi M R, Dent R. Detection of dim targets in high cluttered background using high order correlation neural network, *Proc. Int. Conf. Neural Networks, IJCNN91, Seattle: vol.1, 1991, 701-706.*
- [6] Shun-ichi Amari. Dualistic geometry of the manifold of higher-order neurons, *Neural Networks*, 1991, 4(5): 443-451
- [7] 李金艳. 稀疏高阶神经网络的逼近能力及其算法的研究: [博士论文]. 华南理工大学, 1996.
- [8] 李金艳, 余英林. 二层高阶神经网络对任意布尔函数的实现, *华南理工大学学报*, 1995, 23(10): 111-116.
- [9] Fahner G, Eckmiller R. Structural adaption of parsimonious higher-order neural classifiers, *Neural Networks*, 1994, 7(2): 279-289.

## SPARSE CONNECTION WEIGHTS OF HIGHER-ORDER NEURAL NETWORK AND ITS PRUNING ALGORITHM

Li Shouli    Li Jinyan\*    Li Wangchao

*(Hebei University of Technology, Tianjin 300130)**\*(The University of Melbourne, Australia)*

**Abstract** In this paper, the fully-connected higher-order neuron and sparsed higher-order neuron are introduced, the mapping capabilities of the fully-connected higher-order neural networks are investigated, and that arbitrary Boolean function defined from  $\{0, 1\}^N$  can be realized by fully-connected higher-order neural networks is proved. Based on this, in order to simplify the networks' architecture, a pruning algorithm for eliminating the redundant connection weights is also proposed, which can be applied to the implementation of sparsed higher-order neural classifier. The simulated results show the effectiveness of the algorithm.

**Key words** Higher-order neural networks, Redundant connection weights, Sparsed connection, Pruning algorithm

李守丽: 女, 1973年生, 硕士研究生, 从事神经网络研究.

李金艳: 男, 1969年生, 博士, 现在澳洲墨尔本大学从事神经网络研究.

李望超: 男, 1933年生, 教授, 主要研究方向: 非线性系统理论, 神经网络等.