

# 神经网络设计的特征空间序贯划分算法<sup>\*</sup>)

孙功星 戴贵亮

(中国科学院高能物理研究所计算中心 北京100039)

## A Geometrical Strategy of Constructive Initial Neural Networks

SUN Gong-Xing DAI Gui-Liang

(Computing Center of Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100039, China)

**Abstract** A geometrical strategy for constructive neural networks is proposed in the paper. Firstly it can acquire quickly initial input weight parameters and topology by sequentially partitioning feature space with the presented geometrical method. Secondly with SVD, its initial output weights are obtained very quickly. Finally these weights are returned with BP algorithms. Its distinctive features are that it can construct quickly an initial neural networks using geometrical method other than backpropagation algorithms so that overtraining and undertraining are avoided automatically, and experimentally it performs better on two-spiral classification than Cascade-Correlation Algorithm.

**Keywords** Geometrical strategy, Overtraining, Singular value decomposition

## 1 引言

神经网络已广泛地用于处理实际问题,如语音处理、图像处理和计算机视觉、模式分类和识别等。面对越来越复杂的应用,传统的神经网络学习算法变得不能适应。与许多其他的有效算法相比,神经网络的学习速度慢和固定拓扑结构的适应性两个缺陷显得异常突出。

训练前置神经网络通常采取的办法是用误差反转(Back Propagation, BP)学习算法<sup>[1]</sup>,即利用梯度下降法对二次型误差标准求极小值,在合适的初始条件和网络结构下,它是一个很好的算法。但它随机地选取初始权重时,其训练过程变得极为缓慢。这是因为误差表面存在着许多近似平坦的区域和长而窄的区域,在接近平坦的区域,由误差梯度的信息极小而导致权重改变极小,这种权重微小的改变很容易被隐含神经元的非线性所吸收,因此训练的收敛速度愈加变慢。

利用BP算法训练网络之前,必须要事先指定网络的拓扑结构。这种人为指定的方法具有很大的随机性,因而很不适用。获得适合的神经网络结构的一种方法是修剪(pruning)技术<sup>[2,3]</sup>,它是从一个很大的神经网络开始,目标是修剪掉多余的神经元和权重。修剪多余的神经元和权重的一个重要的技术是所谓的权衰减(weight-decay)技术,在训练过程中,它能衰减掉那些不受BP训练重视的神经元和权重,而使神经网络结构得以优化。获得适合的神经网络结构的另一种办法是增长(growing)技术。与修剪技术相反,它是从一个小的网络开始,逐步增加隐含神经元或子网以适应给定的任务<sup>[4,5]</sup>,最终能够获得具有紧致结构的神经网络。但是,这种神经网络的结构算法往往只注重对训练集的学习,并不能给出一个合理的标准以设计合理大小的神经网络,所以过拟合现象十分严重。研究者们提出的多网络模型虽然可以缓解过拟合问题<sup>[6,7]</sup>,但是缓解的程度非常依赖于单个神经网络的设计。

本文提出一种结构神经网络和初始权重的新方法,它利用序贯方法对训练样本的特征空间进行划分,每次划去的区域里只包含一类样本,直到每个区域中只有一类样本为止。在

对训练模式的特征空间的划分过程中,同时完成确定神经网络隐节点数和初始化隐含层权重两个任务。再用奇异值分解的方法在训练模式的内部表示和理想输出之间作线性回归逼近,确定初始输出权重。最后,再用标准的BP学习算法进一步训练网络。

## 2 神经网络的特征划分结构策略的依据

一个 $n$ 输入的神经元是这样的一种元件,它能执行两种操作,首先它能对其输入 $x = (x_1, x_2, \dots, x_n)^T$ 和连接权 $w = (w_1, w_2, \dots, w_n)^T$ 作点乘操作,即 $sum = \sum_i x_i w_i = x^T w$ 其次,它能产生一个输出 $y = \sigma(sum + \theta)$ ,这里 $\sigma$ 一般为sigmoid函数, $\theta$ 为阈值。因此, $n$ 输入的神经元与一有向的 $n$ 维欧几里得空间 $R_n$ 中的超平面 $H_{w,\theta}$ 相联系

$$H_{w,\theta} = \{x: x^T w + \theta = 0\}$$

$H_{w,\theta}$ 的正侧沿 $w$ 方向,它离原点的距离为 $d = |\theta| / \|w\|$ , $H_{w,\theta}$ 把 $R^n$ 分成两个半空间, $H_{w,\theta}^+ = \{x, x^T w > 0\}$ 和 $H_{w,\theta}^- = \{x, x^T w \leq 0\}$ 超平面则是它们的边界。对于具有 $m$ 个神经元的隐含层(每个神经元有 $n$ 个输入),假定它们为 $N_1, N_2, \dots, N_m$ ,设 $N_k$ 的权值为第 $k$ 个神经元的连接权,让 $W$ 是矩阵,它的列均为 $n$ 维矢量,那么,模式 $x$ 的内部表示为

$$x_{ip} = \sigma(x^T W) = (\sigma(x^T w_1), \sigma(x^T w_2), \dots, \sigma(x^T w_m))^T$$

这 $m$ 个神经元相应于 $m$ 个超平面,它可以被看作是对特征空间 $m$ 次划分。

对两类分类问题而言,假定输出层由单个 $n$ 输入、一个输出的神经元组成。在内部空间中,它相应于一个超平面,倘若在内部表示空间中两类模式位于该超平面的两侧,则称这个分类任务是两层网络可分的。

从上述分析可以看出,隐含层的作用实质上是实现训练模式从输入特征空间到内部表示空间的映射,在隐含神经元函数固定的情况下,这种变换唯一地由特征的划分所确定。输出层的作用是实现模式由内部表示空间向输出空间的转换,同样地,这种转换唯一地由如何划分内部表示空间而确定。因

<sup>\*</sup> 本项目为中国科学院高能所创新资金支持项目(项目编号为U-512)。孙功星 博士,副研究员,主要研究兴趣包括集群计算技术、Grid计算、MSS、人工神经网络、信息处理和理论物理等。戴贵亮 研究员,博士生导师,主要研究兴趣包括电子学读出系统、人工神经网络、计算机网络等。

此,设计神经网络可以通过划分模式的特征表示空间和内部表示空间而完成。

### 3 序贯的特征表示空间的划分策略

设训练样本由  $n$  个样本组成,每个输入样本均为  $n$  维欧几里得空间中的一点。为了方便起见,令序贯划分特征表示空间的过程如下:首先,在模式的特征表示空间中任意选取一点  $p_0$  作为参考点,在完成第  $k$  次对特征表示空间的划分之后,求得剩余训练样本集中与参考点距离最远的样本,记为  $p$ ,并把向量  $p-p_0$  作为超平面的法线方向。然后,求得与  $p$  异类样本中在  $p-p_0$  方向上投影值最大的样本,记为  $p_k$ ,并让该超平面经过该点,由此可以完全确定超平面,即为与第  $i$  个隐节点相联系的那个超平面,它把模式的特征表示空间划分为两个半空间  $H_{w,\theta}^+$ (在超平面的正侧)和  $H_{w,\theta}^-$ 。我们总是把位于半空间的样本除去。重复这样的过程,直到剩余样本集中只有一类样本为止。在训练样本有限且可分的时候,由于每一次至少分离一个样本,所以整个训练过程是收敛的。

假定在序贯空间的划分过程共生成  $m$  个超平面,它们相应于  $m$  个隐含节点。对于第  $k$  个隐节点,输入节点到它的连接权为

$$w_k = p_k - p_0 \quad (1)$$

$$\theta = -p_r^T(p_k - p_0) \quad (2)$$

其中,  $p_0$  是参考点,  $p_k$  是本次选择的距参考点最远的点,  $p_r$  为与  $p_k$  异类在  $p_k - p_0$  上投影值最大的点。

### 4 奇异值分解方法和输出权重的确定

假定一个系统有  $m$  个输入,每个输入有  $n$  个属性,则该系统的每一输出都服从相应线性回型,表示如下

$$y_1 = w_{01} + w_{11}x_1 + \dots + w_{n1}x_n$$

$$y_2 = w_{02} + w_{12}x_1 + \dots + w_{n2}x_n$$

.....

$$y_m = w_{0m} + w_{1m}x_1 + \dots + w_{nm}x_n$$

此处,  $x_i, y_i$  均为列矢量,将上面的线性系统表示为矩阵形式为

$$Y = XW \quad (3)$$

其中,  $X = (x_1, x_2, \dots, x_{N_i})^T \in R^{n \times m}$ ,  $Y = (y_1, y_2, \dots, y_{N_i})^T \in R^{m \times m}$ ,只有当  $Y$  位于  $W$  所张开的子空间时,方程才有精确的解。在  $W$  为非奇异的特殊情况下,可以在方程两边左乘求得的解

$$W = X^{-1}Y \quad (4)$$

但是,在大部分实际的回归问题中,样本数远远地大于回归模型中的参数个数。 $Y$  位于  $W$  所张开的子空间中的可能性很小。因此,人们把它归结为求解线性回归问题,

$$\|Y - WX\|^2 = \min \quad (5)$$

本文中,我们用奇异值分解的方法来求解这个线性回归问题。奇异值分解方法主要分两步,首先,在  $W$  所张开的子空间中寻找正交矩阵  $U$  和  $V$

$$V^T W U = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}$$

其中,  $\Sigma$  是对角矩阵。

再由式求得线性最小二乘问题的唯一解

$$W = U \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} V^T Y$$

此处,  $\Sigma^{-1}$  是  $\Sigma$  的逆。

在神经网络的训练模式的内部表示已经确定的情况下,我们可以把求解输出权重问题转化为线性回归问题。隐层神经元与输出层的连接权矩阵就变成式中的  $W$ ,训练模式的内部表示即为  $X$ ,训练模式的理想输出的反  $\sigma$  函数为  $Y$ 。通过这样的转换,就可以用奇异值分解(SVD)方法求解神经网络的输出权重。这样确定的输出权重是最优的,因为它使得训练模式的内部表示到输出之间的误差为最小。

### 5 学习算法

前面我们已经详细地介绍了神经网络拓扑结构和初始权重优化的几何策略,为了获得更满意的结果,我们采用 EBP 算法对已经获得的网络进一步地训练。整个网络的学习过程归纳如下:

1. 在训练模式的表示空间中任选一点  $p_0$  作为参考点,令  $k=1$ 。
2. 在剩余样本集中,挑选  $p_k = \max_{x_k} |x_k - p_0|$ ,在与  $p_0$  异类的样本中,选  $p_r = \max_{x_m} (x_m - p_0)^T (p_k - p_0)$ ,以  $p_k - p_0$  作为超平面的法线,并让该超平面通过点  $p_r$ ,完成超平面  $H_{w,\theta}$  的结构过程。
3. 让第  $k$  个隐节点与输入节点的连接权  $w_k = p_k - p_0$ ,阈值  $\theta = -p_r^T (p_k - p_0)$ ,删除属于正侧的样本。若剩下的样本集中只有一类样本,则转4,否则,令  $k=k+1$ ,转2。
4. 求得训练样本的内部表示,与相应的理想输出的反函数  $\sigma^{-1}(Y)$ 。运用 SVD 方法求输出权重。
5. 用 EBP 算法训练已经建立的神经网络。如满足要求,退出。

### 6 模拟实验

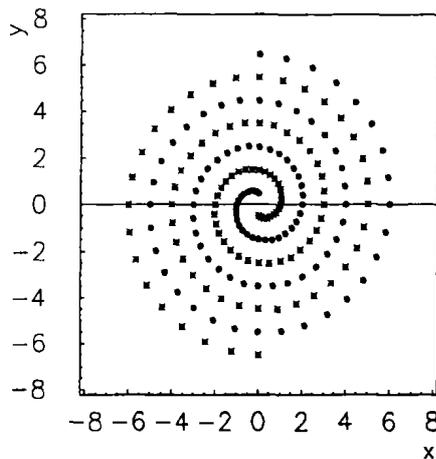


图1

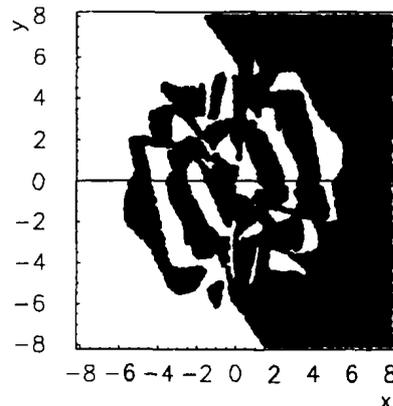


图2

(下转第47页)

鼠基因: AF121351)。在两个基因序列中,使用 RepeatMasker 的默认设置 (<http://ftp.genome.washington.edu/RM/RepeatMasker.html>) 找出重复元素并将其从序列中剔除。虚线矩形框是利用程序 SIM (<http://globin.cse.psu.edu>) 计算所得的结果, SIM 采用的是 Smith-Waterman 的局部联配算法, 它所报告的最长的局部联配(位置在人类基因的 6817-53081) 将非保守区域和高保守区域(图5中的阴影矩形框)连在了一起。实线矩形框是利用本文的改进算法的计算结果,  $X=100, c=1, b=1$ 。

**结论** Smith-Waterman 算法是目前最著名的一种局部联配算法, 但将此算法用于同源长序列的局部联配时, 经常会出现马赛克问题, 该问题也一直困扰着人们。本文提出利用动态扣分策略(对特定情况采用动态扣分以切断异常联配)来解决马赛克问题。实验结果表明, 在基本不改变原 Smith-Waterman 算法的前提下, 在没有显著增加算法的时间复杂度和空间复杂度的情况下, 顺利解决了序列局部联配中的马赛克问题。

## 参考文献

1 Rota P A, et al. Characterization of a Novel Coronavirus

(上接第37页)

双螺旋问题是公认的神经网络基准测试。它由平面上两条互相缠绕三圈的螺旋线组成, 每一圈有32个点。加上一个结束点, 每类共有97个点。其平面上的分布如图1。Lang 和 Witbrock 用有直接连接且结构为2-5-5-5-1的 BP 网络来分类双螺旋问题, 获得成功<sup>[7]</sup>。但我们在实验中用同样的网络结构和学习参数, 未能获得成功。可见 BP 网络在学习复杂任务时, 成功具有很大的偶然性。Falham 和 Lebiere 把级联相关学习网络运用于求解双螺旋问题, 最终得到的神经网络的隐节点12到19个之间, 平均隐节点数为15.2个<sup>[8]</sup>。我们运用 Falham 和 Lebiere 发展的级联相关学习网络于双螺旋分类问题, 最终生成的网络具有17个隐含节点, 共有207个连接权。其泛化结果如图2。从中可以看出, 级联相关学习网络的过拟合现象严重, 泛化能力差。

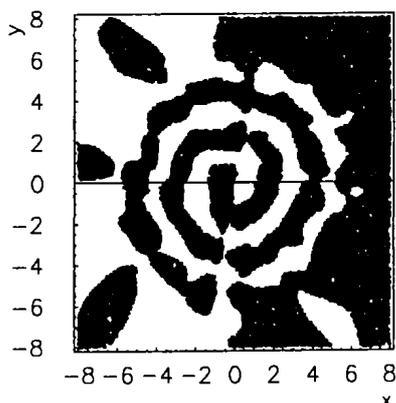


图3

我们设计的方法也能快速地学会所有的训练样本, 形成一具有48个隐节点的网络, 共有193个权参数。其泛化结果见图3。比较两种方法生成的网络结果, 不难看出, 两种网络在结构紧致性方面相当, 但是, 在泛化能力方面, 我们设计的网络却优于级联相关网络。

**结论** 我们已经详细地介绍了结构神经网络的几何策

- Associated with Severe Acute Respiratory Syndrome. Science, 2003, 300(5624)
- 2 Fujiyama A, et al. Construction and Analysis of a Human-Chimpanzee Comparative Clone Map. Science, 2002, 295(5552): 131~134
  - 3 Wood R D, et al. Apoptotic Molecular Machinery: Vastly Increased Complexity in Vertebrates Revealed by Genome Comparisons. Science, 2001, 291(5507): 1279~1284
  - 4 Smith T F, Waterman M. S. Identification of common molecular Sequence. J. Mol. Biol., 1981, 97: 723~728
  - 5 Vingron M, Waterman M. S. Sequence Alignment and penalty choice. J. Mol. Biol., 1984, 235: 1~12
  - 6 Huang X, et al. Parametric recomputing in alignment graph. In: Proc. of the 5th Annual Symposium on Combinatorial Pattern Matching, Asilomar, California 1994. 87~101
  - 7 Zhang Z, et al. Post processing long pairwise alignments. Bioinformatics, 1999, 15: 1012~1019
  - 8 Arslan A N, et al. A New Approach to Sequence Comparison: Normalized Sequence Alignment. Bioinformatics, 2001, 17: 327~337

略, 并把它运用于双螺旋分类问题的求解。由于在结构网络的过程中没有用到误差极小的方法, 所以, 形成网络的速度很快。它不仅克服了人为指定的拓扑结构的不适应性的困难, 而且, 克服了随机选择初始权重所引起的学习过程缓慢的困难。此外, 结构神经网络的几何策略还说明了如何有效地划分特征表示空间是神经网络学习过程中至关重要的问题, 也是神经网络面对愈来愈复杂的问题取得成功的关键所在。

结构神经网络的几何策略还存在着值得深入探讨的问题, 包括如何更为有效地划分特征表示空间、如何修剪已经形成的网络使网络得以进一步优化地及引入容错功能于特征表示空间划分的过程等。

本文仅给出结构两层前置网络的策略, 但并不是它只适用于两层, 它完全可以推广到多层网络情况。

## 参考文献

- 1 Rumelhart D E, McClelland J L. Parallel Distributed Processing. Cambridge, MA: MIT press, 1987
- 2 Teng C C, Wab B W. Automated learning for reducing the configuration of a feedforward neural network. IEEE Trans. Neural Networks, 1994, 7(5): 1072~1085
- 3 LeCun Y, Denker J S, Solla S A. Optimal Brain Damage. in Advances in Neural Information Processing System 2, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, 1990. 598~605
- 4 Hwang J-N, You S-S, Lay S-R. The cascade-correlation learning: a projection pursuit perspective. IEEE Trans. NN, 1999, 7(2): 278~289
- 5 孙功星, 朱科军, 戴长江, 等. 多层次多子网级联神经网络. 电子学报, 1999, 27(8): 49~51
- 6 Ji Chuanji, Ma Sheng. Combination of weak classifiers. IEEE Trans. Neural Networks, 1997, 8(1): 32~42
- 7 孙功星, 戴贵亮. 改进 CAS 性能的多网络表决模型. 小型微型计算机系统, 2001, 22(2): 168~170
- 8 Lang K I, Witbrock M J. Learning To Tell Two Spiral Apart. In: Proc. 1988 Connectionist Models Summer School, 1989. 52~59
- 9 Fahlman S E, Lebiere C. The Cascade-Correlation Learning Architecture. in Advances in Neural Information Processing Systems 2, D. S. Touretzky, Ed. Los Attos, CA: Morgan Kaufmann, 1990. 524~532