

[7] 杨顺安,曹剑芬,语言研究, 1(1984),15—22.
 [8] 贺宁基,“北京话二合元音感知中的时间因素”,《北京语言实验录》,北京大学出版社,1985.
 [9] Ren Hongmo, UCLA working papers in phonetics, 65(1986).
 [10] Morton J., Marcus S., and Frankish C., *Psychological Review*, 83(1976), 405—448.

[11] Fowler C. A., *Perception and Psychophysics*, 25 (1979), 375—388.
 [12] 梁之安,声学学报, 2(1965),20—23.
 [13] Bladon A., *Speech Communication*, 4(1985), 145—154.

34-38

人工神经网络, 语音识别 8
声图象识别

自组织人工神经网络用于声图象识别的研究

桑恩方 乔晓宇 李 瑞

(哈尔滨船舶工程学院水声研究所 哈尔滨 150001)

1992年11月30日收到

TP 391.4
TP18

本文通过实验研究了自组织人工神经网络用于声图象识别的步骤和方法,讨论了在水声、超声医学等声图象识别中所遇到的一些关键性技术问题。

一、引言

八十年代以来,人工神经网络的研究取得了新的重要进展。由于它们的并行性、分布式存储、自学习、自组织的结构特点,对传统人工智能有了较大突破。因而在许多领域已展现了广阔的应用前景。

在声学领域中,随着海洋开发、海底沉物探测、智能机器人声视觉及超声医学等事业的发展,人们已发展了多种二维和三维高分辨率声成象技术。随之必然地提出如何根据声图象进行被探测物体的自动识别和理解的任务要求。借助人工神经网络来较好地完成这一任务,是本文研究的主要目的。

已有许多重要的神经网络模型被提出。^[1-3]由 Kohonen 提出的自组织算法模型^[4,7,8]是其中代表研究之一。由于它所模拟的是人脑神经对外界刺激具有自动排列和顺次响应的功能,因而具有很强的分类性。已用这种模型设计出矢量量化器。用于语音识别和图象编码。

完成计算机“识别”和“理解”的基础是实现

一个自动分类器。本文主要研究如何根据声图象的几何特征进行自动分类,从而为目标的进一步识别和理解打下基础。

二、自组织人工神经网络模型

以 Kohonen 网络算法为代表的自组织算法是一种无“教师”学习的方法。输出可看成是输入样本特征的一种表达。它是一个双层网络,其输出节点是在平面上按顺序排列的。其算法为:

若网络有 N 个输入节点, M 个输出节点,给出一组初权向量 W_{ij} 后,对于在时刻 t 给定的一个样本 $X_i(t)$,计算距离

$$d_i = \sum_{i=0}^{N-1} [X_i(t) - W_{ij}(t)]^2$$

$$0 \leq j \leq M - 1 \quad 0 \leq i \leq N - 1 \quad (1)$$

选择距离最小的输出节点为响应节点 j^* ,然后修正 j^* 及其领域内输出节点连接的权值:

$$W_{ij}(t+1) = W_{ij}(t) + k(t)[X_i(t) - W_{ij}(t)] \quad (2)$$

其中 $0 \leq j < M - 1, 0 < i < N - 1; 0 <$

$k(x) < 1$ 是随时间降低的增益。经过训练后,网络把输入映射为输出平面上的一个点。对于相近的输入它的输出响应结点在输出平面上也是拓朴定义下相近的。

在声图象识别中,我们把它分为两个主要过程,即学习过程和识别过程。前者是用大量图象特征样本进行训练学习的过程。后者则是利用学习过程得到的各类权值判断出某图象样本类别的过程。基本自组织网络模型及其学习训练过程和识别过程的流程图分别示于图 1、图 2 和图 3。

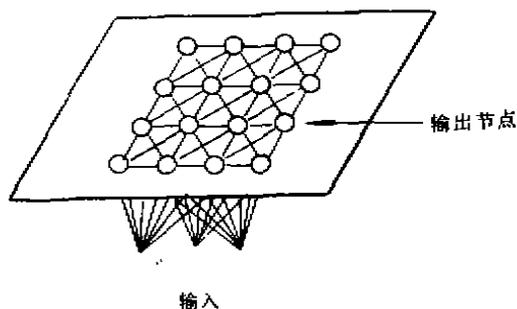


图 1 自组织神经网络的一般结构图

声图象识别的实验研究发现,用作为分类判据的特征集和特征量的选取是一个首要问题,是决定自组织算法分类成功与否的关键。

1. 数据处理技术

由于声成象孔径的限制和声学目标及环境复杂的散射规律,使声图象分辨率和对比度都很差,并存在较强的噪声和混响背景(可参见后面附图中的原始图象)。这使得声图象特征集的提取要比一般光学或电视图象更困难。解决的办法是针对声图象的特点合理地选用和开发数字图象处理技术,^[5,6,9]作好预处理。本研究有效地运用和开发了如下一些数字处理技术于声图象的预处理中:

(1) 线性变换, $g(i, j) = LNT[f(i, j)]$, 其中 $g(i, j)$ 为变换后的图象, $f(i, j)$ 为原图象。LNT 为视图象情况而定的线性函数。

(2) 非线性变换: $g(i, j) = T[f(i, j)]$, T 为非线性变换函数。它往往产生更明显的增

应用声学

强效果。但容易使相对较暗的图象信息受到抑制。本文采用求各点象素第 n 次 (n 一般取 2 或 3) 幂的非线性变换法。变换后再进行归一化处理。

(3) 平滑: 除抑制点噪声,局部干扰和毛刺外,对图象的局部断裂和缺损有一定拟合作用。但平滑次数越多,窗口越大,图象变得越模糊。

(4) 分割: 本文采用按不同灰度特征的区域生成方法将目标和背景区分开来。其中较多地采用了门限化方法。由于声图象干扰背景严重,有时一次门限化处理还达不到所需的分割效果,需平滑和分割交替进行两次或多次,视图象具体情况而定。

(5) 边缘检测,我们选用了抗干扰能力较强的 Sobel 算子法。但检出后的边缘较粗,还需要进行细化处理。

(6) 细化: 本文根据声图象特点开发了一个新的有效细化方法。它最后给出的是图象边缘的中心线和骨架形状,同时也保留了连通性和孤立的端点。该方法与一般细化方法的效果对比见附图 1。在附图 1 中,左上角为原始图象。右上角和左下角为一般通用细化方法的结果,右下角为本文采用的新的细化方法结果。

(7) 边界跟踪: 采用了经典的数字图象边界跟踪方法,^[9]给出了细化后边界点的坐标。

在进行了这些预处理的基础上,进行了几何特征集的提取。开发了声图象中目标物体的面积、周长、形状因子、曲率、区域长度和宽度、内径等检测算法和程序。限于篇幅,这些内容在此不做详细介绍。

显然,参加训练的各类特征集越多,收敛后的权值越能反映各类概率分布的中心,训练后网络结构越具有适应能力。然而工程中所能得到的训练样本总是有限的,我们在尽可能多地搜集训练样本的前提下,对所有样本采用循环输入的训练方法,收到了较好的效果。

2. 初始权值

当各类差别较大时,初始权值可设定为任意随机小数。而当各类差别不大时,初始权值

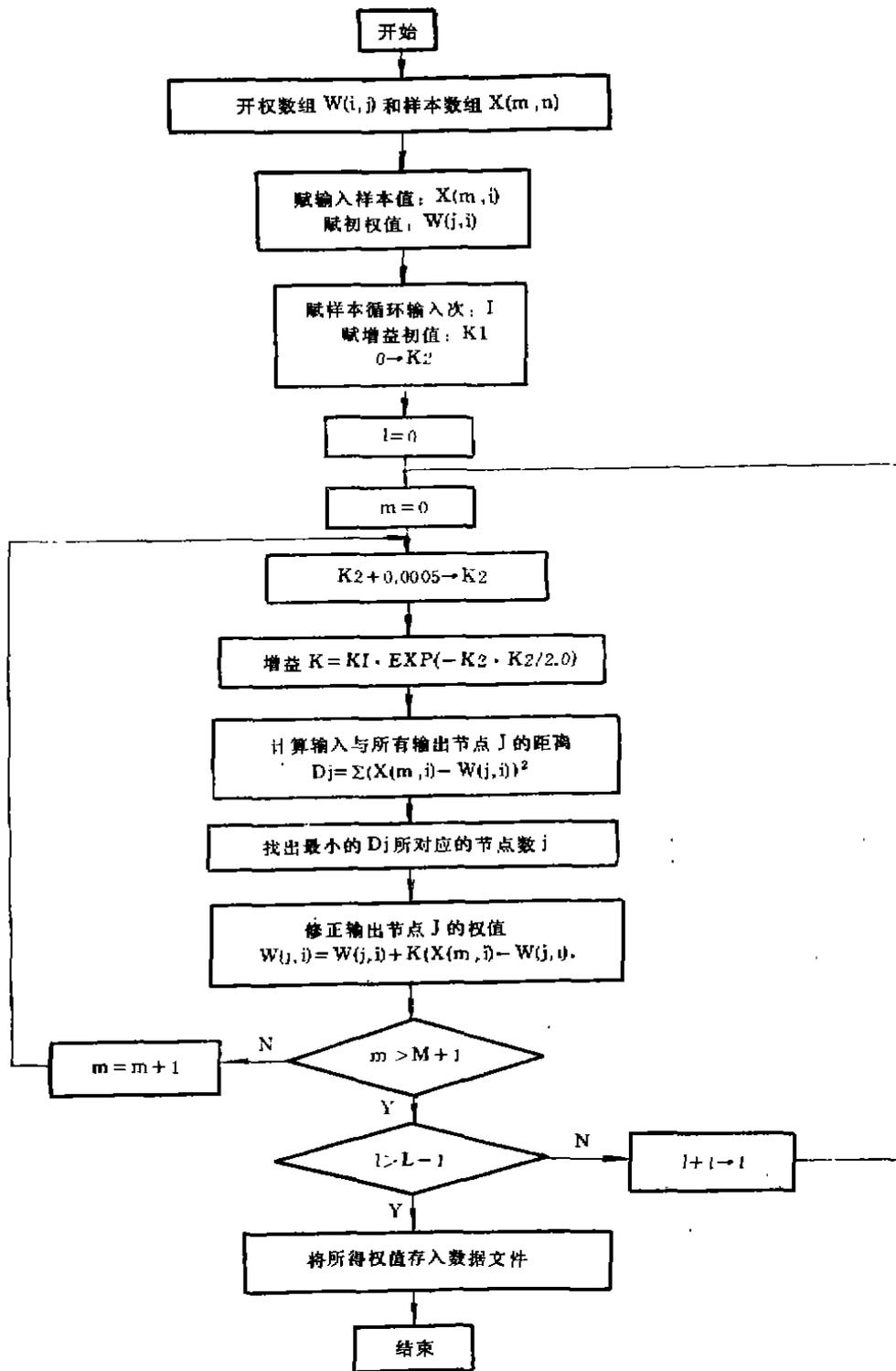


图2 自组织算法学习训练仿真流程图

应设定为参加训练样本的代数平均值或典型值, 否则将出现并类情况。

3. 各类特征向量(集)的距离
由于距离判据

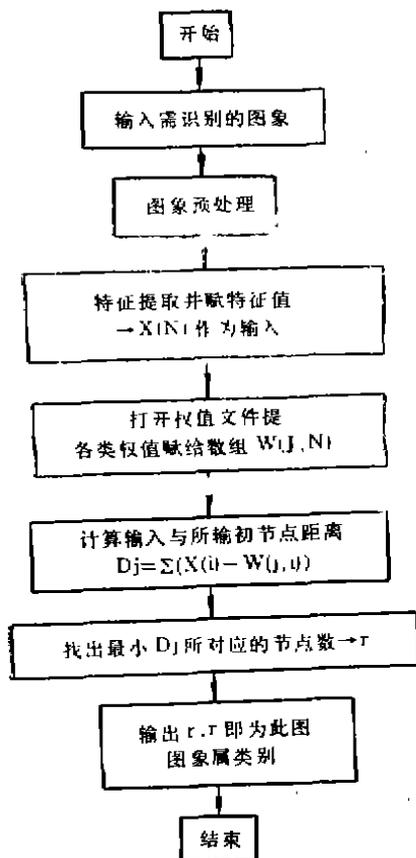


图3 自组织算法识别流程

$$d_j = \sum_{i=0}^{N-1} (X_i(t) - W_{ji}(t))^2$$

是基于最小方差原理，因而各特征矢量维数的多少及各矢量之间的距离将对分类有决定性意义。可采用变换的方法，人为地增大这一距离。此外应注意各维样本值对 d_j 贡献的均衡。对于那些绝对数值较小而更具特征的样本应赋予大的权值。

4. 增益 k 的选择

增益 k 是一个时间递减函数。由于权值修正公式为：

$$W_{ji}(t+1) = W_{ji}(t) + k(t)(X_i(t) - W_{ji}(t)),$$

因而 $k(t)$ 的函数形式将影响收敛速度和收敛精度。本文根据实验结果选择了

$$k(t) = k_0 \exp(-t^2/2).$$

应用声学

三、实验研究结果举例

1. B 超图片的病理诊断研究

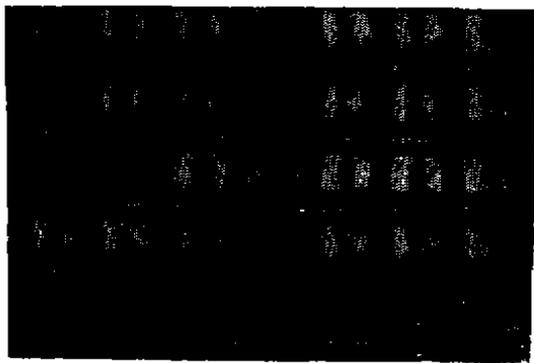
B 超图象是迄今分辨率最高的声图象，我们摄取了 150 例患者腰椎 B 超图片，每个患者取四节腰椎。从中选取 36 例典型图片。医生对腰椎病人诊断的主要依据是腰椎骨路的宽度特征来定某一节腰椎为正常或异常。我们按医生提供的特征判据，对腰椎 B 超图片病理的自动识别作了一定的研究，首先设自组织网络的输出节点个数为 16，每一输出节点代表骨骼情况中的一种，例如对某一患者而言，四节腰椎骨骼全部正常则为 1111，若第一节腰椎有病变为 0111。这样对每个病人可有 16 种情况。其次把 B 超图片分为训练集和识别集两部份。开始经过较简单预处理后进行识别发现有 25% 左右的并类现象，后进行样本特征的非线性变换等相应的一系列图象预处理，那么对未参加训练的识别集中 B 超图片病理自动诊断结果与医生诊断结果相吻合，因此可说识别率为百分之百。B 超图片的处理过程见附图 2。

2. 水声图象目标识别研究

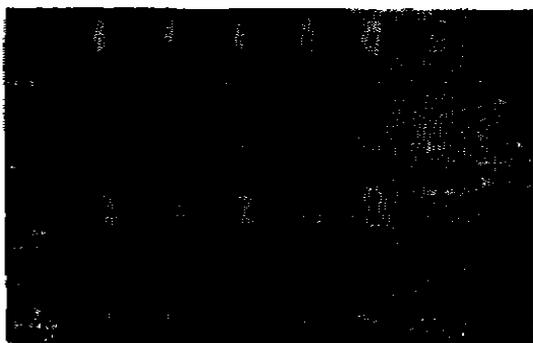
选取了两幅水雷水声图象见附图 3，和圆筒水声图象见附图 4。在对水雷声图象特征集的提取过程中注意选用了水雷目标亮区和阴影区的联合特征集提取。其中包括各自的面积、周长、圆度、长度、宽度等十个简单的几何特征。对于水下圆筒声图象的特征提取也结合其横断



附图1 左上角为原始图象，右上角及左下角为常用方法细化后的图象。右下角为本文所改进的细化方法生成的图象。



附图2 以一个患者的四节腰椎B超图片为例,进行一系列图象处理(原始、线性变换、二值化、平滑、边缘提取、二值化、细化)后所得图象。



附图3 两幅水雷声图象的图象处理过程。提取每幅水雷声图象的亮区部分及阴影部分的图象特征,进行相应处理



附图4 水下圆筒声图象的图象处理及特征提取过程。

面和侧面的一系列相应的几何特征。以上的声图象经过一系列的图象处理,诸如图象的线性变换、二值化、平滑、边缘提取、细化等,再进行特征集提取。由于可用的声图象种类很少,训练时增加了一些模拟样本。自组织网络经过对上述特征集的学习过后,能正确分出水雷,圆筒和模拟样本。

四、结 论

在针对声图象的特点进行数字预处理,正确地选取特征集并配以适当的变换和加权以后,采用自组织人工神经网络算法对声图象进行分类识别是可行的。人工神经网络的自组织方法具有较强的聚类功能,计算简单明确,易于实现,可在声呐图象、超声医学图象及其他声学图象分类识别中应用。

参 考 文 献

- [1] Lippmann R. P. "Introduction to computing with Neural Nets", *IEEE Assp Mag*, April 1987.
- [2] Wasserman P. "Neural Computing Theory and Practice", Vannoststrand-Reinhold, New York, 1989.
- [3] 严红凤,戴汝为,模式识别与人工智能, 13-3 (1990), 3-10.
- [4] Kohonen T, "Self-organization and Associative Memory", Springer-Verlag, Berlin, 1984.
- [5] Davis L. S. "A Surrey of Edge Detection Techniques", *Computer Graphics and Image Processing*, 4 (1975).
- [6] P.K. Sahoo et al, "A Surrey of Thresholding Techniques" *CVGIP* 41, 1988, 233-260.
- [7] Von der Malsberg C., "Self-organization of Orientation Sensitive Cell in the Strato Cortex", *Kybernetik* 14.
- [8] Kohonen T., *Associative Memory*, Springer 1977.
- [9] 周新伦、柳建、刘志华编,数字图象处理,国防工业出版社,1986.
- [10] A mari S., *IEE Trans. on Computer Vision, Graphics and Image Processing Vol. 2* (1986), 48-56.