

图形、图像、模式识别

回声状态网络及其在图像边缘检测中的应用

裴承丹

PEI Cheng-dan

中南民族大学 工商学院 计算机系,武汉 430223

Department of Computer Science, School of Industry and Merchandise, South-Central University for Nationalities, Wuhan 430223, China

E-mail: dawnpcd_2006@163.com

PEI Cheng-dan. Echo state networks and its application on image edge detection. *Computer Engineering and Applications*, 2008, 44(19): 172-174.

Abstract: Recurrent Neural Networks (RNN) is a kind of important artificial neural networks with better ability for learning and rate of convergence in comparison with forward neural networks, however, the design of the structure of the hidden-layer is a difficult problem all the time. Echo State Networks has no such problems with special construction, good stability, short-cut learning process. Application of ESN to the edge detection of images has been introduced after the presentation of the structure and method of learning of ESN, resulting well.

Key words: Echo State Networks (ESN); edge detection; statistical vector

摘要: 循环神经网络(RNN, 也称反馈神经网络)是一种重要的人工神经网络,与前馈神经网络相比具有更好的学习能力和更快的收敛速度,但其隐层结构的设计一直是个难点问题。回声状态网络(ESN)有效地解决了上述问题,相比于以前的循环神经网络,其具有结构独特、稳定性好、学习过程简单快捷等特点。介绍了回声状态网络及其学习方法,将其用于图像的边缘检测中,取得了良好的效果。

关键词: 回声状态网络; 边界检测; 统计向量

DOI: 10.3778/j.issn.1002-8331.2008.19.052 **文章编号:** 1002-8331(2008)19-0172-03 **文献标识码:** A **中图分类号:** TP391

1 引言

边缘检测是一种重要的图像预处理技术,是图像处理和计算机视觉领域中最重要研究课题之一,是进行模式识别和图像信息提取的基本手段,一直受到研究者的广泛重视。虽然传统的算子算法对边缘的检测速度快,但实际处理的图像一般都混有噪声,其得到的往往是断续的、不完整的边缘信息,且这类检测方法都是基于微分的思想,而图像边缘和噪声在频域都表现为高频分量,因此这类方法对噪声比较敏感,在检测噪声污染图像时会得到许多虚假的边缘。由于人工神经网络(Artificial Neural Network)固有的自学习和自适应能力,决定了它可以很好地完成数字图像的边缘检测。该领域成为近几年来人们关注的焦点。

目前,用人工神经网络来做边缘检测的算法大多是选取BP神经网络,BP神经网络具有较强的学习能力,能从已知信息中得到有利于解决最优化问题的结论,但由于BP神经网络是基于梯度下降的误差反传算法进行学习的,所以网络训练速度通常很慢,且容易陷入局部极小点^[1]。尽管采用一些改进的快速学习算法可以较好地解决某些实际问题,但是在设计过程中往往都要经过反复的试凑和训练过程,无法严格保证每次训练时BP算法的收敛性和全局最优性。此外,BP网络隐层神经元的

作用机理及其个数选择已成为BP网络研究的一个难点问题。

竞争Hopfield神经网络具有收敛速度快的特点,可用于图像边缘检测^[2],但是实际操作时也存在如何选择适合于具体问题的拓扑结构以及如何设计训练的方法的问题。且随着所要解决问题规模的增大,Hopfield网络的收敛速度会大大降低。

作为一种新型递归神经网络,回声状态网络^[3-5](Echo State Network)较传统的递归神经网络有较大改进^[6],能够很好地解决上述问题。首先,在稳定性方面,可以通过预先设定动态记忆库矩阵的谱半径来保证递归网络的稳定性;其次,在网络训练方面,输出权值的确定是唯一且是全局最优的,因此没有传统神经网络普遍存在的局部最小问题;再次,ESN避免了传统递归神经网络求取时序偏微分的过程,因此网络的训练过程变得特别简单。

2 回声状态网络

2.1 回声状态网络的结构和特点

回声状态网络是一种全新的分析和训练循环神经网络的方法,其训练过程快速、简单。回声状态网络中的隐含层神经元之间的权值在学习过程中是不进行调整的。这样的学习方法使得回声状态网络最后通过线性回归的方式得到训练,所以非常

作者简介: 裴承丹(1983-),女,助教,主要研究方向为图像处理与分析。

收稿日期: 2007-09-07 **修回日期:** 2007-12-14

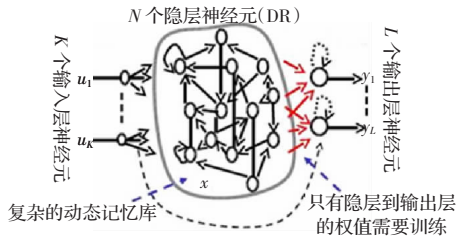


图1 回声状态网络示意图

容易实现与理解。

回声状态网络的基本思想是将隐含层设计成一个巨大的动态记忆库(Dynamic Reservoir, DR),其典型结构如图1所示,分为三层:输入层、隐含层和输出层。DR的规模一般需要上百至上千个神经元,而每个神经元在内部的连接是随机且有向的。输入与输出神经元是在动态记忆库已经构建完成后再与其隐含层中的任何一个神经元相联。整个网络可用连接矩阵来表示:假设选定了动态记忆库的规模为 N ,即有 N 个神经元组成动态记忆库,则这 N 个神经元相互连接的所有权值可以利用一个 $N \times N$ 的矩阵 W 来表示, a_{ij} 表示这个连接矩阵的第 i 行第 j 列的元素,表示从第 i 个神经元到第 j 个神经元的有向权值的数值。同样,输入与输出神经元与DR相连时也可以用连接矩阵来表示。假设有 K 个输入神经元,则其与DR内部所有的神经元相连时可以产生 $N \times K$ 的矩阵 W^{in} ;而假设有 L 个输出神经元,则其与DR内部所有的神经元相连时可以产生 $L \times N$ 的矩阵 W^{out} 。另外有些情况下需要输出神经元与DR有反馈的连接,这时也可以产生一个 $L \times (L+K+N)$ 反馈连接矩阵 W^{back} 。在训练过程中,只需训练输出权值 W^{out} 即可, W^{in} 、 W 、 W^{back} 三个矩阵在学习之前建立起来就不再改变,这就解决了循环神经网络对训练方法的要求。回声状态网络须使隐含层具有动态记忆能力,因此要求 W 是一个稀疏矩阵,其密度一般为5%~20%,且谱半径的绝对值小于1。 W^{in} 、 W^{back} 也需要一个绝对值小于1的谱半径。

通过DR连接输出节点的权值是主要学习对象,而通过输入节点连向输出节点与输出节点自己连接自己的权值部分是在某些情况下才进行连接与学习的。

2.2 回声状态网络的学习过程^[7]

循环神经网络在学习的时候一般需要通过信息反馈来调整权值,而回声状态网络的学习思路是通过对学习数据的采集使DR记忆相应的信息,最后通过线性回归过程来最小化平均误差平方值。假设有一个样本序列对 $(u(1), d(1)), \dots, (u(n), d(n))$,其中 u 为输入时间序列矢量,维数为 K ; d 为输出时间序列矢量,维数为 L ,则从集合 $U=(u(1), u(2), u(3), \dots, u(n))$ 到集合 $D=(d(1), d(2), d(3), \dots, d(n))$ 形成了一种对应的映射关系。将 U 当作神经网络的输入, D 当作输出,两个集合的数据就能通过 W^{in} 、 W^{back} 传输进DR中,DR内部的神经元也随之产生相应的震荡,每输入一对 $(u(n), d(n))$,网络按式(1)的运行规则进行:

$$x(n+1) = f(W^{in}u(n+1) + Wx(n) + W^{back}d(n)) \quad (1)$$

式中 n 为网络运行的步进, $x(n)$ 是在先前第 n 步的时候DR的内部神经元状态矢量,维数为DR的规模数, $u(n+1)$ 是新的输入量, $d(n)$ 是先前第 n 步的输出量, f 是DR的节点激活函数。而网络输出按公式(2)进行:

$$y(n+1) = f^{out}(W^{out}(u(n+1), x(n+1), y(n))) \quad (2)$$

式中 $u(n+1)$ 、 $x(n+1)$ 、 $y(n)$ 分别代表当前的输入、隐含层以及

先前第 n 步的输出矢量。通过观察公式(1)和公式(2)可以发现,虽然我们希望隐含层神经元的连接路径存在,但是对此并没有做要求,而且对拓扑结构也没有要求。这就解决了循环神经网络对拓扑结构的要求。

回声状态网络的基本数学基础可以归纳如下:以一个限制最大元素值(或者多数时候是谱半径)的稀疏矩阵作为内部相互连接的拓扑结构;只有输出权值需要训练。因此,它很好地解决了循环神经网络的拓扑结构和训练方法的问题。回声状态网络具有比较低的算法复杂度,允许快速计算。当然,要训练回声状态网络,设计者还需提供一些参数,比如矩阵的维数,密度及谱半径等。

3 边缘点的统计特征向量

边缘可定义为两个具有不同灰度的均匀图像区域的边界,即边界反映局部的灰度变化。边缘点邻域内的灰度分布和非边缘点邻域内的灰度分布是不同的。边缘点的邻域内至少有两种不相同的灰度分布,而非边缘点的邻域内只有一种灰度分布。因此,可以基于边缘的方向性,通过构造统计向量定量描述边缘点邻域灰度的分布特征。

图像的边缘信息在频域中表现为高频分量,边缘检测过程是一个高频增强过程,但噪声在这一点上非常相似,同样表现为高频分量。因此,所构造的作为神经网络的输入序列^[8]的特征向量要既能准确传达图像边缘信息,又能克服噪声的干扰。

图像处理中检测边缘点主要依据边缘邻域内灰度的分布的有序性和方向性;在做边缘检测之前我们做如下假设:给定区域的每个像素要么属于边缘像素,要么属于非边缘像素;给定区域的所有像素至少属于两个不同的分类区

在文献^[9,10]的基础上,根据边缘像素的灰度分布特征与噪声点的不同之处来构造了由4个统计量组成的统计向量,并将训练图像的统计向量作为回声状态网络的样本来对网络进行训练,将训练完成的网络直接用来判断像素点是否为边缘点。该方法避免了通常边缘检测方法中设定阈值的问题,且回声状态网络的容错性也进一步增强了抗噪性能。

以点 (i, j) 为中心的邻域记为 $R = \{(x, y) | |x-i| \leq L, |y-j| \leq L\}$, x, y 分别是像素点的行、列索引值, L 为邻域长度的一半。过中心点 (i, j) 以不同的方向将邻域分成两部分,记为 R_1 和 R_2 ,如图2所示,令 N_1, N_2 分别表示两个子区域内的像素个数, f_{xy} 表示像素点 (x, y) 的灰度值。则可定义统计量:

$$S_1^2 = \sum_{(x,y) \in R_1} \left(f_{xy} - \frac{\sum_{(x,y) \in R_1} f_{xy}}{N_1} \right)^2 \quad (3)$$

$$S_2^2 = \sum_{(x,y) \in R_2} \left(f_{xy} - \frac{\sum_{(x,y) \in R_2} f_{xy}}{N_2} \right)^2 \quad (4)$$

$$S^2 = \frac{S_1^2 + S_2^2}{N_1 + N_2} \quad (5)$$

$$M = \left| \frac{\sum_{(x,y) \in R_1} f_{xy}}{N_1} - \frac{\sum_{(x,y) \in R_2} f_{xy}}{N_2} \right| \quad (6)$$

$$T = \frac{M}{\sqrt{S^2}} \quad (7)$$

若当前邻域存在过中心点的边缘且当边缘轨迹与假定边缘方向(图2中的黑色像素)重合时,两个区域 R_1 和 R_2 的灰度分布差别最大,两个区域内部的灰度分布趋于一致,因此 M 取得极大值, S^2 取得极小值, T 取得极大值;当轨迹方向与假定边缘方向垂直时,情况相反。因此,若当前邻域是平滑区,则8种假定边缘方向都是将一个灰度分布趋于均匀的区域分成两个部分, M 和 T 均取值较小,因此 M 和 T 的取值反映了边缘点邻域内灰度分布的差异。而且,无论是边缘区还是平滑区,无论窗口如何划分,区域 R_1 和 R_2 内的噪声分布和噪声强度在概率上都是相同的,不会对 M 、 T 反映边缘点的特征造成太大影响,所以统计量 M 、 T 具有较好的抗噪性能。

所选的模板不能太大也不能太小^[11],如果所选的领域模板太大,网络的规模就比较大,网络的计算量就比较大,增加网络的训练及检测时间,同时边缘的信息也可能被削弱甚至丢失;同时也不能太小,如果模板太小,就不能很充分地对该个像素的周围领域进行很好的学习。本文选用的是 5×5 的模板,即每个模板中包含 25 个像素, $N_1=N_2=10$ 。

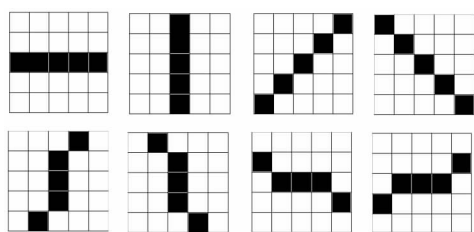


图2 窗口的8种不同划分方法

在以每个方向划分的子区域内,根据(6)、(7)计算 M 、 T ,得到集合 $\{M_k; k=1, 2, \dots, 8\}$ 和 $\{T_k; k=1, 2, \dots, 8\}$,然后在 5×5 窗口内构造中心点 (i, j) 的统计向量 v :

$$v = \{\max(M_k), \max(T_k), \min(M_k), \min(T_k)\} \quad (8)$$

根据前面的分析,如果 v 中的前两个分量较大,而后两个分量较小,那么模板中心点对应的像素是边缘点;否则不是边缘点。

4 基于 ESN 和统计向量的边缘检测

实验步骤:

(1)随机产生一个 N 为 100、密度为 5% 的隐含层稀疏矩阵(动态记忆库),它和 4 个输入神经元及一个输出神经元构成一个循环神经网络,其中输出神经元要以随机权值连接回动态记忆库。

为使得回声状态网络的动态记忆库具有动态记忆特性,其权值矩阵 W 按照以下方法确定:记 $\sigma_{\max}(W)$ 表示矩阵 W 的最大元素的值, $|\lambda_{\max}(W)|$ 表示矩阵 W 的最大特征值的绝对值。则有:

$$\sigma_{\max}(\alpha W) = \alpha \sigma_{\max}(W) \quad (9)$$

$$|\lambda_{\max}(\alpha W)| = \alpha |\lambda_{\max}(W)| \quad (10)$$

实验中 λ_{\max} 取 0.8。随机生成一个稀疏矩阵 \tilde{W} , 利用一个缩放因子 α 将 \tilde{W} 转换为所要求的矩阵 $W = \alpha \tilde{W}$, 其中 $\frac{1}{\sigma_{\max}(\tilde{W})} \leq$

$\alpha \leq \frac{1}{|\lambda_{\max}(\tilde{W})|}$ 。这样就使得矩阵 W 具有动态记忆功能。生成谱半径的绝对值小于 1 的权值矩阵 W^{in} 和 W^{out} 。

(2)对训练图像用经典算子与人工相结合的方法得到边缘图像。具体可以综合各种经典边缘检测算法获得全部的边缘,

而只保留希望得到的边缘;然后通过窗口在训练图像上的滑动计算像素点的统计向量构成回声状态网络的训练样本 $v(1), v(2), \dots, v(n)$ 。

(3)输入组样本 $v(1), v(2), \dots, v(n)$, 激活函数采用双曲正切函数 \tanh 。对网络进行训练,得到隐含层到输出层的连接权值 $w_i (i=1, \dots, 100)$ 。

(4)将训练完成的回声状态网络直接用于边缘检测。以图像的统计向量作为输入,训练的回声状态网络可以直接判断各像素是否为边缘点,从而得到边缘图像。

本文用回声状态网络对图像的统计向量进行边缘检测,一方面对回声状态网络进行有目的的训练使它能够直接判断边缘点,这样就避免了通常边缘检测方法中设定阈值的问题;另一方面基于回声状态网络的容错性可以进一步增强抗噪性能。

5 实验结果

用图3(a)中图像作为训练图像,分别对其采用传统的 Canny 算子、Roberts 算子等进行边缘检测来获取全部的边缘,然后用手工的方法保留希望得到的边缘,结果如图3(c)。实验中分别对 BP 神经网络和 ESN 进行训练,结果网络的收敛速度如图4所示。



(a)训练图像 (b)加噪图像 (c)目标边缘

图3 训练图像及目标边缘

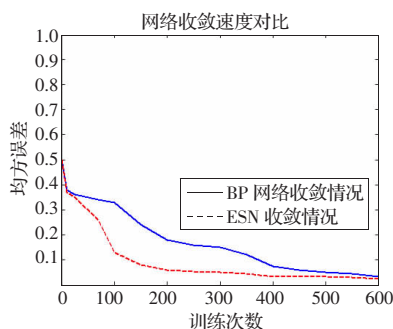


图4 BP 网络和 ESN 的收敛速度

BP 神经网络的收敛情况在图中用实线表示,而 ESN 的收敛情况用虚线表示,从图中可以看出,ESN 的收敛速度明显快于 BP 网络。

用训练好的回声状态网络直接对图5(a)进行边缘检测,其结果如图5(d)所示,作为对比,Sobel 算子和 Prewitt 算子的边缘检测结果分别如图5(b)和图5(c)所示。



(a)原始图像 (b)Roberts 检测结果 (c)Prewitt 检测结果 (d)本文检测结果

图5 边缘检测测试结果

对图5(a)进行加噪,然后分别用上述方法对其进行分割,

(下转 192 页)