

文章编号:1000-582X(2005)01-0042-03

BP 网络研究及其在肺癌诊断系统中的应用*

黄雪梅¹,唐治德¹,赵一凡²,舒志强²

(1. 重庆大学 高电压与电工新技术教育部重点实验室,重庆 400030;
2. 南京长城信息系统有限公司 软件部,江苏 南京 210004)

摘要:以三层 BP 网络为例,讨论了网络的学习算法,并在 VC++6.0 语言环境下编制了 BP 算法程序,在这个基础上建立了基于 BP 网络的肺癌智能诊断系统,深入研究 BP 网络在实际应用中的结构设计、参数选择和样本数据的来源,应用含有动量参数的 BP 算法训练网络,并给出了 BP 算法动态演化过程的训练调试界面和临床测试界面。目的是用训练好的 BP 网络来识别肺癌细胞的病变情况,实验表明神经网络能够对细胞图像进行正确的分类,证明了该系统对肺癌诊断具有可行性。

关键词:BP 神经网络;肺癌诊断;模式识别

中图分类号:TP183

文献标识码:A

从 20 世纪 80 年代开始,国外掀起了一股研究人工神经网络的热潮,至今仍方兴未艾。1986 年,Hagan T M^[1]提出了反向传播(back-propagation)学习算法。该算法在肺癌诊断中得到了广泛的应用。Falchini^[2]报道基于 CAD 联用神经网络技术提高对肿瘤的早期检查,可以使敏感性达到 80%,准确率达到 98%。Nakamura^[3]利用神经网络建立了计算机辅助诊断方案,帮助放射科医生鉴别良性和恶性肿瘤,极大地提高了诊断的准确性。笔者利用三层 BP 网络来诊断肺癌,根据肺癌细胞的临床特征提取特征参数,从而采集大量样本来训练神经网络,再利用训练好的网络来诊断肺癌。

1 BP 算法研究

该系统使用三层前馈网络,即由输入层、隐含层和输出层组成。其网络结构如图 1 所示。

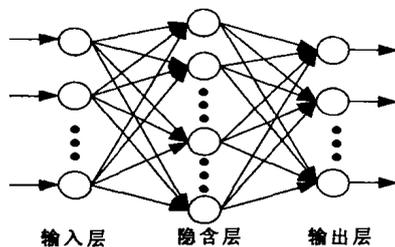


图 1 三层 BP 网络结构

三层前馈网络能以任意精度逼近任意的非线性函数^[4],这已经得到了严格的证明。其算法步骤如图 2 所示。

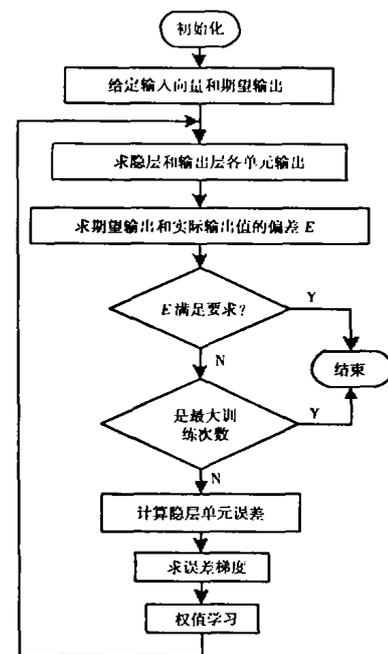


图 2 BP 算法框图

为了改善收敛特性,本系统在学习权值时加入动量项,因此权值修改公式如下:

$$W_{ij}(t+1) = W_{ij}(t) + \eta d_j^k V_j^{k-1} + \alpha \Delta W_{ij}(t) \quad (1)$$

* 收稿日期:2004-09-15

作者简介:黄雪梅(1979-),女,重庆人,重庆大学硕士研究生,从事图像处理与人工智能研究。

式中 η 为学习速度; α 为动量系数。

2 系统的实现

该系统具有两种功能:训练功能和测试功能。其中训练过程的人机交互界面如图 3 所示,其接口如下:

- 1) 各层神经元数。可选,包括输入层、隐含层和输出层;
- 2) 参数设置。可供设置的参数有学习步长、动量参数、训练次数和训练样本数;
- 3) 网络训练结果。全局误差、权值、阈值和识别率。

网络训练好了之后直接用来诊断病人的病情,根据节点输出就可以得出诊断结果。图 4 给出了一幅细胞切片图的测试结果。

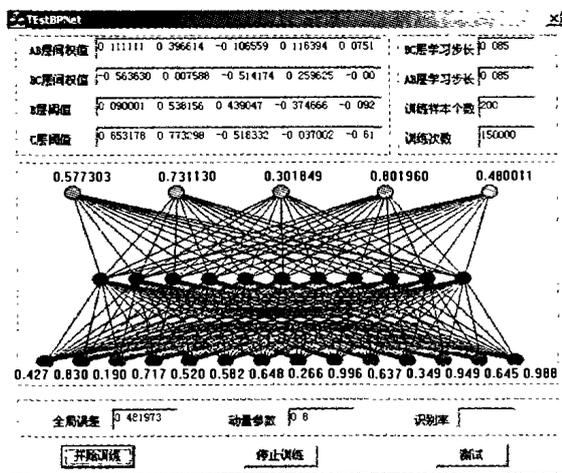


图 3 训练过程的人机交互界面

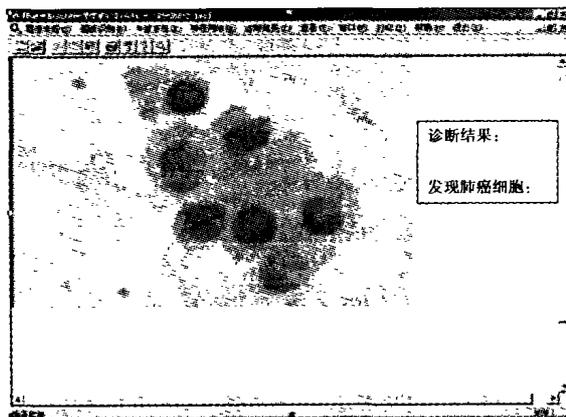


图 4 诊断结果

3 具体应用与结果分析

3.1 训练样本及网络结构的确定

3.1.1 样本数据来源

构造训练样本的过程就是将肺癌患者和正常人的肺部细胞的特征抽象成机器知识的过程。文中的样本

数据是通过采集肺癌穿刺标本涂片或者切片图,先后经过图像预处理、图像分割和特征提取^[5]而得到的(共 14 个特征值)。所有样本输入数据都是经过归一化处理的,取值范围为[0,1]。诊断结果有 5 种:鳞癌、腺癌、小细胞癌、核异型和无癌,对应 5 个数据输出。每个输出取值“0”或者“1”(即对应诊断类为“1”,其余类为“0”)。

3.1.2 网络结构的确定

输入层取 14 个神经元,对应肺部细胞的 14 个特征值——6 个形态特征值、6 个色度特征值、整幅切片图的红色分量平均值和细胞区域的灰度平均值;隐含层根据经验值以及实际情况取 11 个节点;输出层取 5 个神经元,对应 5 种不同的诊断结果——鳞癌、腺癌、小细胞癌、核异型和无癌。

3.2 网络训练过程

在网络初步确定结构后,接下来的工作就是确定各个初始参数的取值。一般而言,学习步长对训练结果的影响在于当其过大时,可能导致系统不稳定,过小时会导致较长的训练时间,本系统中取学习步长为 0.085。另外,各单元的权值和阈值的初始选取 [-0.5, +0.5] 之间的随机值,最大训练次数定为 15 000 次,并以固定目标函数(均方误差)0.000 1 来进行训练,采用的是 Pentium 4 的品牌机。

由于动量参数对网络收敛特性有很大的影响,笔者取用 0.7 到 0.9 之间的数作为动量大小来训练网络,通过比较最后选定 0.8 作为动量参数。在其他参数相同的条件下,加入动量使得均方差更小,说明动量项加快了收敛速度。并且在误差变化过程中,加入动量后误差变化更稳定,出现的局部极小次数 n 明显降低,这表明使用动量项,可以抑制振荡,能够在维持算法稳定的前提下达到更快的学习速度。

选定参数下,网络训练误差 E 随着训练次数 n 变化的曲线如图 5 所示。

从图 5 可以看出的训练参数的选择是合理的,训

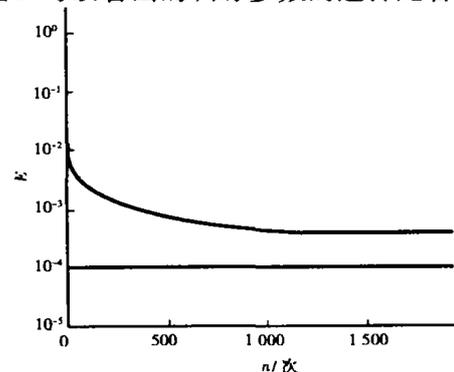


图 5 神经网络训练误差曲线

训练的收敛速度快,下降趋势平稳。

3.3 训练策略与分类能力

经过多次实验以后,笔者用不同的样本数分别进行训练,得到的权值和阈值存储在文本中,以供测试。本训练中一共有500个样本,选取部分样本作为训练集来训练网络,剩下的样本作为测试集来测试网络,并计算出识别率,记录见表1。测试过程中,如果对应类的输出单元产生高值(≥ 0.95),而同时其它节点的输出产生低值(≤ 0.05),则认为是一正确识别^[6]。

表1 以不同样本集训练后的诊断结果

训练样本数	测试样本数	识别率/%
100	400	99.5
150	350	97.3
200	300	95.8
250	250	94.1
300	200	92.5
400	100	80.9

从表2可以看出以相同的训练次数、不同的训练集来训练网络时,训练集样本数越多越难收敛,识别率越低。

3.4 应用实例说明

用上一节中训练好的BP网络来测试网络的分类能力。首先对肺癌穿刺标本涂片进行采集,然后进行染色处理。一共有69张涂片,其中无癌5例,腺癌17例,鳞癌23例,小细胞癌15例,核异型9例,每张涂片采集8个视野,共552幅图像,对每幅图像进行预处理和图像分割,标记细胞并提取特征值,共获得无癌测试样本41个,腺癌测试样本138个,鳞癌测试样本164个,小细胞癌测试样本136个,核异型测试样本73个。试验结果如表2所示。

笔者所采用的BP网络的肺癌诊断方法的结果较好地符合了已知数据,具有较高的准确性,误判率较低,有利于肺癌的早期发现和治理。

表2 肺癌识别结果

细胞类别	测试样本数	正确识别	误判
无癌	41	39	2
腺癌	138	134	4
鳞癌	164	156	8
小细胞癌	136	129	7
核异型	73	70	3

4 结束语

笔者在原有BP理论的基础上稍作改进,加入了权值和阈值调整动量参数,大大加快了收敛速度。而且写出了三层BP网络训练的通用程序,给出了人机交互界面。将设计好的BP网络用于肺癌智能诊断系统中,提高了系统的强壮性,为肺癌智能诊断系统的研制开辟了一条新途径。实验表明,这个方法是可行的,其总体识别率达92%以上,该结果与病理专家的诊断精度相当接近。

参考文献:

- [1] HAGAN T M, DEMUTH B H, BEALE H M 著. 神经网络设计[M]. 戴葵等译. 北京:机械工业出版社,2002.
- [2] FALCHINI M, STECCO A, CARMIGNANI L. Neural Network Based Detection of Pulmonary Nodules on Chest Radiographs[J]. Radio Med(Torino), 1999, 98(4): 259 - 263.
- [3] NAKAMURA K, YOSHIDA H, ENGELMANN R. Computerized Analysis of the Likelihood of Malignancy in Solitary Pulmonary Nodules with Use of Artificial Neural Networks[J]. Radiology, 2000, 214(3): 823 - 830.
- [4] 高大启. 有教师信号的线性基本函数前向三层神经网络结构研究[J]. 计算机学报, 1998, 21(1): 80 - 86.
- [5] 夏良正. 数字图像处理[M]. 南京:东南大学出版社,2001.
- [6] 郑南宁. 计算机视觉与模式识别[M]. 北京:国防工业出版社,1998.

BP Neural Networks and Its Application in Lung Cancer Intelligent Diagnosis System

HUANG Xue-mei¹, TANG Zhi-de¹, ZHAO Yi-fan², SHU Zhi-qiang²

(1. Key Laboratory of High Voltage Engineering and Electrical New Technology Under the State Ministry of Education, Chongqing University, Chongqing 400030, China;

2. Software Department, Nanjing Great Wall Information System Limited Company, Nanjing 210004, China)

Abstract: The learning algorithm of networks is discussed. The programming example of 3 layer BP networks is given with Visual C++ 6.0 program language. Based on this model, a lung cancer intelligent diagnosis system is successfully implemented. Furthermore, the paper introduces network's structure design, preferences and the source of stylebook datum in factual applications. The ameliorative arithmetic is applied to the study of networks and BP dynamic evolving process is designed. The experiments indicate cell images are recognized and classified by the trained neural network. The study illustrates the system has feasibility and clinical value in lung cancer diagnosis.

Key words: BP neural networks; lung cancer diagnosis; pattern recognition

(编辑 李胜春)