中国农业大学学报 2001, 6(3): 6~10
Journal of China Agricultural University

① 人工神经网络在恶性肿瘤诊断中的应用

李晓薇[1①] 刘文凯[2] 李军会[1]
(1 中国农业大学基础学院; 2 中国医学科学院肿瘤医院)

摘 要 本研究将人工神经网络方法用于人血清荧光光谱法诊断癌症。采用反向传播神经网络方法, 建立识别人血清荧光光谱的网络模型。参数经过训练, 平均识别率高达 94%。

关键词 人工神经网络; 血清; 荧光光谱; 癌; 肿瘤诊断
中图分类号 R318

Application of Artificial Neural Networks in Cancer Diagnosis

Li Xiaowei[1] Liu Wenkai[2] Li Junhui[1]
(1 College of Basic Science and Technology, CAU; 2 Tumour Hospital, Chinese Academy of Medical Sciences)

Abstract The serum fluorescence spectrum mothed of cancer diagnosis is developed by using Artificial Neural Network. The Back-Propagation Network is adopeted to set up the network model for serum fluorescence spectrum recognition. After training, the averaged rate of recognition is high to 94%.

Key words artificial neural networks; serum; fluorescence spectrum; cancer; tumour diagnosis

近年来有关癌症的检查与诊治越来越引起重视。目前较先进的诊断方法多为: 微量元素的比例失调与癌症诊断的关系[1], 人血清荧光光谱中叶啉代谢与肿瘤诊断的关系, 并根据荧光光谱的特征峰判断人血清是否含有恶性组织[2,3]。另外, 利用光学法诊断癌症还有其他方法[4~7]。但光谱方法由于仪器来源, 物理来源和化学来源导致的非线性使得光谱有: ①浓度响应函数的曲线化, ②吸收带位置的漂移, ③吸收带宽度的改变。这 3 方面的变化。给分析工作者对血清中癌变组织固有荧光光谱特征峰的判别造成困难, 使得人血清光谱的人工判断率较低, 一般临床的判别率为 80%。

有关神经网络的应用在国内外已有诸多报道[7~18]。本研究提出一种新的人工神经网络方法, 可用于大批量数据的非线性分析, 并采用反向传播神经网技术对人血清荧光光谱进行判别, 以提高癌症诊断的判别率。

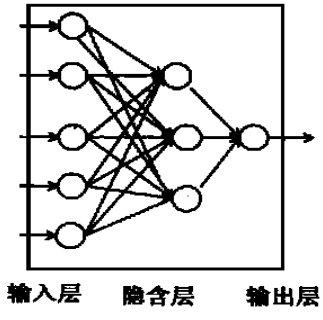

图 1 人工神经网络模型

图 1 是人工神经网络(artificial neural networks, ANN)模型: 误差反传算法BP(back propagation)为 3 层ANN 系统, 它包括一定数量结点(神经元), 图中圆圈表示神经元, 而且上一层的每一个结点都同下层的每一

收稿日期: 2000-09-22
①李晓薇, 北京圆明园西路 2 号中国农业大学(西校区), 100094

个结点连接。同外界的联系通过输入, 输出层的结点进行, 隐含于内部的中间层决定系统的主要计算能力。在ANN 的每条连接线上是权重因子, 代表系统记忆能力的重要部分, 当网络学习时, 权重的数值可随着正在网络中流通的新信息而改变[19~20], 一般应用时先用某种类型的一组已知数据输入ANN 模型, 经过训练后ANN 就能用学过的已知数据推测出新输入的未知数据。

1　材料与方法

1.1　材料

已知有癌、无癌的人血清荧光光谱值(北方交大和中国医学科学院肿瘤医院提供)。人工神经网络分析模型软件, 本课题组编制。

1.2　方法

将人血清光谱数据转变为通用格式, 便于微机采集。神经网络采集数据时选择包含有健康样品和癌变样品(人血清荧光光谱值)共 42 例, 作为训练集; 另选一批包含有健康样品和癌变样品共 68 例, 作为预测集, 然后用"1"表示健康数据,"0"表示有癌变数据, 输入ANN, 便于训练模型。

对光谱进行预处理, 以降低燥声、减小分析样品误差。可供选择预处理方法有矢量归一化, 极差归一化, 附加点散射校正, 一阶导数和二阶导数。

提取光谱的主成分, 即将数据降维, 将变量进行转换, 使少数几个变量成为原变量的线性组合, 同时这些变量尽可能多表征原变量的数据结构特征而不丢失信息。本研究将人血清荧光光谱数据的前五维主成分信息作为神经网络的 5 个输入层结点, 其中每个主成分都是所有波长点的线性组合。

算法参数优化: 光谱预处理方法与主成分确定后, 即可对人工神经网络模型进行训练。在训练过程中, 通过不断调整网络各层之间的权重因子来调节输出层的目标输出值, 使含癌变样品的数据逼近于 0, 健康样品的数据逼近于 1, 直到收敛精度满足条件为止。训练结束后, 将ANN 参数保存在模型中, 用该模型来预测未知人血清荧光光谱数据, 预测结果可作为肿瘤诊断依据。

需要考查训练的参数为学习速率 η, 范围 0.1~0.9; 动量系数 α, 范围 0.1~0.9; 训练次数 t, 范围 1.0~10^6 次。

2　结果与分析

2.1　神经网络拓扑学结构

$$j=(i+1)/2 \qquad i=1, 2, 3, \cdots$$

(1) 当神经网络输入层结点数为 $i=1$~4 时, 运算时间 $t<5$ s, 网络输入信息过少, 不具代表性。

(2) 当神经网络输入结点数为 $i=6$~10 时, 运算时间 $t>30$ s, 网络输入信息增多, 运算较慢。无论选择何种变量, 对测试结果影响均不大。经反复优化最优ANN 输入层结点数 $i=5$, 隐含层结点数 $j=3$, 输出层结点数 $k=1$。因此解决本问题的神经网络拓扑学结构为BPN (5-1-

3), 训练效果最佳。

2.2 光谱预处理方法的选择

对人血清荧光光谱数据进行预处理, 发现选用矢量归一化与极差归一化后, 模型训练时误判率较低, 平均正确识别率达 98% (表 1)。本研究选择矢量归一化作为光谱的预处理方法, 在所用 42 例样品中, 健康血清光谱 16 例, 癌变血清光谱 26 例。

表 1 光谱预处理方法的选择

方 法	健康血清样品量				癌变血清样品量			
	误判	不确定	正确	φ/%	误判	不确定	正确	φ/%
矢量归一化	0	0	16	100	1	0	25	96
极差归一化	0	0	16	100	1	0	25	96
附加点散射校正	11	2	3	19	1	0	25	96
一阶导数	6	3	7	44	0	0	26	100
二阶导数	13	0	3	19	0	0	26	100

人工神经网络模型判别方式: 当训练集样品为健康者, 将误报定为< 0.5, 不确定为 0.4～0.5, 正确为≥0.5。当训练集样品中有癌变, 将误报定为> 0.5, 不确定为 0.4～0.6, 正确为≤0.4。

2.3 训练集样本平均识别正确率

光谱预处理方法确定后, 开始优化参数结果如下。

2.3.1 学习效率训练 此时其他参数: 训练次数 $t=10^4$ 次, 动量系数 $a=0.1$。

当学习效率 $\eta=0.9$～0.8 时, 人工神经网络模型发散, 不适宜预测样品。如 $\eta=0.3$～0.1 时, 网络训练太少模型误判率较高, 也不适宜预测样品。

当学习效率 $\eta=0.7, 0.6, 0.5, 0.4$ 时, 人工神经网络模型收敛, 训练集样品平均识别正确率分别为 85%, 90%, 85%, 80%, 可预测样品。而最佳选择为 $\eta=0.7$ 和 $\eta=0.6$ 两项。

由以上训练结果可知当其他参数不改变时, 网络学习效率可使用 $\eta=0.6$～0.7。

2.3.2 动量系数训练 此时其他参数: 学习效率 $\eta=0.7$～0.6, 训练次数 $t=10^4$ 次。

当动量系数 $a\geq0.6$ 时模型发散, 不适宜预测样品。不断调整动量系数 a 使训练集样品平均识别正确率不断上升, 当动量系数 $a=0.5, 0.4, 0.3, 0.2, 0.1$ 时模型收敛, 训练集样品平均识别正确率为 69%, 78%, 80%, 85%, 88%, 可预测样品。而其中最佳选择 $a=0.1, 0.2, 0.3$。利用上面 2 项结果对其他参数进行进一步训练。

2.3.3 训练次数调试 此时其他参数: 学习效率 $\eta=0.7$～0.6, 动量系数 $a=0.1$～0.3。

将训练次数设定在 $t=10^4$～5×10^5 次时, 人工神经网络模型收敛, 训练集样品平均识别正确率为 90%, 95%, 95%, 95%, 可预测样品, 其中 $t\geq3\times10^4$ 次时训练效果最佳。

综合以上训练结果可知学习效率 η, 动量系数 a 较为关键, 当其他参数不改变时, 训练次数 $t\geq3\times10^4$ 次时, 网络运算不受影响。

将上述几项训练样品平均识别正确率较高参数优化预处理, 得到最佳人工神经网络模型反传算法参数如下: 学习效率 $\eta=0.6$, 训练次数 $t=5\times10^4$ 次, 动量系数 $a=0.1$。然后将上述优化的 ANN 对训练集 41 例样品进行校正, 其平均识别正确率高达 95%, 适合样品检测(表 2)。

2.4　预测集样本平均识别正确率

本研究采用优化训练过的ANN 对预测集 68 例样本(其中健康 28 例, 癌变 40 例)进行预测, 即对病人血清荧光光谱值进行的鉴别诊断说明, 本网络模型平均预测正确率高达 94% (表 3), 结果优于人工判别。

人工神经网络模型判别方式: 当预测集样品为健康者, 将误报定为< 0.5, 不确定为 0.4～0.5, 正确为≥0.5。当预测集样品中有癌变, 将误报定为> 0.5, 不确定为 0.4～0.6, 正确为≤0.4。

表 2　训练集模型报告

样本类型	数量	误报	不确定	正确
健康	16	0	0	16
乳腺癌	10	1	0	9
肺癌	4	0	0	4
胰腺癌	5	0	0	5
胃癌	2	0	0	2
结肠癌	5	0	0	5

表 3　预测集模型报告

样本类型	数量	误报	不确定	正确
健康	28	0	1	27
乳腺癌	15	0	0	15
肺癌	9	0	0	9
胰腺癌	8	0	0	8
胃癌	3	0	0	3
结肠癌	3	0	0	3
脑癌	2	1	0	1

3　结论

综上所述, 本研究所提供的人工神经网络对图形处理有较强的优越性, 尤其对于批量较大的样品。采用人工神经网络的方法进行总体测定, 采集信息量多, 就人血清荧光光谱样品而言, 除卟啉成分外, 光谱中与肿瘤有关其他信息也被采集, 而且计算方法是全谱测定, 因此具备较强的抗噪声和抗干扰能力, 克服了上述因素的影响。又由于所选样品包括大部分恶性肿瘤的信息如: 肠癌、肺癌、胃癌、胰腺癌、乳腺癌, 因此所建模型不受癌症类型局限, 通用性较强。另外, 本方法分析速度快, 预测癌症准确率较高, 是一种新的临床肿瘤诊断方法。

参　考　文　献

1　孙抒, 李龙山, 孙东植, 金昌范　乳腺癌及其癌前病变中细胞凋亡与细胞增殖的原位观察　中华肿瘤杂志, 1999, 21(6): 447～449

2　孟继武, 西坂刚, 深海隆明　肿瘤发展过程中卟啉代谢的特点及在肿瘤诊断上的意义　科学通报, 1995, 6(12): 11144～11147

3　连少辉, 杨士珍　原卟啉发光特性与癌发光关系的研究　生物化学杂志, 1996, 4(2): 225～228

4　Yang Y, et al. Exciation spectroscopy neaveal changes of proteins and the degree in malignant. Tissue SPIE, 1999, 3597: 511～513

5　Li X, Wang Q, et al. Laser-incluced blood serum fluorescene and raman spectroscopy for cancer diagnosis. Proceeding of SPIE, 1999, 3863: 301～304

6　Alfano R R, et al. Optical spectroscopic diagnosis of cancer and normal breast eissue. J Opt Soc Am B, 1989, 6(5): 1015～1023

7 Bimman K P. Rnle-based learning for more accurate ECG analysis. IEEE Pattern And Intell, 1982, 4: 369～375

8 Xuc Q Z, Hu Y H, Tompkins W J. Neural net work-based adaptive matched filtering for QRS detection. IEEE Trans Biomed Eng, 1992, 39: 317～329

9 Cheung J Y, Hull S J. Detection of abnormal eleitrocardiograms using a neural network approach. In: Proc Ann Int Conf. IEEE Eng Med Biol Soc, 1989, 2015, 2016

10 王继成, 吕维雪. 一个基于符号神经网络结的心电图分类系统. 中国生物医学工程学报, 1996, 9: 202～297

11 蔡煜东, 宫家文, 甘骏人. 人工神经网络方法在乳腺癌死亡率研究中的应用. 中国生物医学工程学报, 1994, 12: 364～366

12 刁晓娣, 江志斌, 刘瑾. 根据孕妇参数预测胎儿体重的神经网络方法. 中国生物医学工程学报, 1999, 6: 155～158

13 Ercal F. 基于肿瘤彩图像的恶性黑瘤的神经网络诊断法. 国外医学生物医学工程分册, 1995, 18(5): 302 (摘自 IEEE Trabs BM E, 1994, 41(9): 837)

14 Hassoum M H. 肌电描记忆的重复矢量的神经网络的提取法: Ⅱ. 性能分析. 国外医学生物医学工程分册, 1995, 18(5): 302(摘自 IEEE Trans BM E, 1994, 41(11): 1053)

15 Park B, Chen Y R, Whittaker A D, et al. Neural network modeling for beef sensory evaluation. ASA E, 1994, 37(5): 1547～1553

16 李志良, 曾鸽鸣, 胡芳. 神经网络在肝硬化病因鉴别诊断中的应用. 中国生物医学工程学报, 1997, 3(1): 92, 93

17 吉海彦, 严衍录. 用人工神经网络处理谷物成分分析. 高等学校化学学报, 1993, 5: 38～42

18 任劲松, 任超世. 人工神经网络在生物医学技术中的应用. 国外医学生物医学工程分册, 1996, 19(10): 40～45

19 黄德双. 神经网络模式识别系统理论, 北京: 电子工业出版社, 1996, 4～68

20 袁曾任. 人工神经网络及其应用. 北京: 清华大学出版社, 1999, 66～131