

证的研究

前馈反向传播网络在抑郁症中医证型分类中的初步研究

胡随瑜¹ 唐凤英¹ 喻长远² 张海男¹ 陈泽奇¹

摘要 目的:研究前馈反向传播网络(BP网络)在抑郁症中医证型分类中的作用。方法:将1731例抑郁症患者随机分成两组,轮流进行训练和测试。结果:抑郁症BP网络训练总体阳性率为97.9%,测试总体阳性率为72.5%。结论:BP网络能较好地地区分抑郁症5类中医证型,在中医证型分类识别中有一定价值。

主题词 抑郁症/诊断 @前馈反向传播网络

抑郁症是一种情感、精神障碍性疾病,其中中医证型分类各家看法不一,尚无统一标准。本研究旨在引入一种新的方法——人工神经网络中的前馈反向传播网络(BP网络),对抑郁症中医证型进行分类,为抑郁症的辨证论治提供依据。

1 一般资料

本项目研究数据来源于临床流行病学调查。本课题组与协作单位于2001年12月~2003年5月在湖南、天津、北京、哈尔滨、深圳、广西、福建、贵州等8个调查点对1731例抑郁症患者进行中医辨证及证候指标调查^[1],用人工神经网络对5个常见证型进行分析,其中肝郁气滞证588例,男243例,女345例,平均年龄 36.8 ± 12.1 岁;肝郁脾虚证487例,男195例,女292例,平均年龄 37.7 ± 12.6 岁;肝郁痰阻证264例,男110例,女154例,平均年龄 38.3 ± 12.1 岁;心脾两虚证254例,男94例,女160例,平均年龄 38.3 ± 12.1 岁;肝肾阴虚证138例,男62例,女76例,平均年龄 38.3 ± 12.1 岁。均具有可比性。

2 研究方法

2.1 BP网络原理:BP网络是对非线性可微分函数进行权值训练的多层前向网络,它通过连续不断地在相对于误差函数斜率下降的方向上计算网络权值和偏差变化而逐渐逼近目标。每一次权值和偏差变化都与网络误差的影响成正比,并以反向传播的方式传递到每一层。BP算法由两部分组成,即信息的正向传递与误

差的反向传播。在正向传播过程中,输入信息从输入层经隐含层逐层计算传向输出层,每一层神经元的输出作用于下一层神经元的输入。如果在输出层没有得到期望的输出,则计算输出层的误差变化值,然后转入反向传播,通过网络将误差信号沿原来的连接通路反传回来修改各层神经元的权值直至达到期望目标^[2]。BP网络工作可分为训练与测试两个过程,在训练过程中,通过对样本的学习掌握样本规律,提取样本特征,从而利用其产生的某些参数对未知样本进行测试,达到模式识别、分类及数据压缩等目的。

2.2 BP网络结构设计:本研究采用的BP网络是在Borland C++ Builder 5上编译执行的,本设计采用3层结构网络,即输入层、隐含层、输出层。BP网络拓扑结构见附图。(1)输入层:在抑郁症常见中医证型临床流行病学调查的96个症状基础上,去其含义重复及罕见的症状,余75个症状,每一症状对应一个节点,网络设计为75个节点。以每例病例作为输入向量,则输入向量 X_i 满足:相应节点上的症状不出现时,标记为“0”,轻度症状标记为“0.25”,中度症状标记为“0.5”,重度症状标记为“0.75”。(2)隐含层:人工神经网络隐含层的层数及节点数是经过大量的训练之后得出的最佳网络。本研究训练好的网络隐含层为1层,节点数30。网络隐含层节点数太多,网络将记住输入模式而不是输入特征,网络隐含层节点数太少,网络不能识别输入模式中的所有特征,从而使训练误差加大。(3)输出层:根据临床流行病学调查,对1977例抑郁症患者不同中医证候构成比分析,依据其构成比大小,排在前5位的证型是肝郁气滞证、肝郁脾虚证、心脾两虚证、肝郁痰阻证、肝肾阴虚证^[1]。每一证型以五维数字标记为理想输出。本研究人工神经网络输出层设计为5个节点,每一节点对应一种证型,即肝郁气滞证→10000、肝郁脾虚证→01000、心脾两虚证→00100、肝郁痰阻证→00010、肝肾阴虚证→00001。网络训练达到理想输出时,即视为训练好的网络。

2.3 训练及测试方法:不同的网络训练会产生不同的结果。本研究采用交叉训练测试法。把1731例抑郁症患者随机分为两组,分别为865例及866例。训练测试

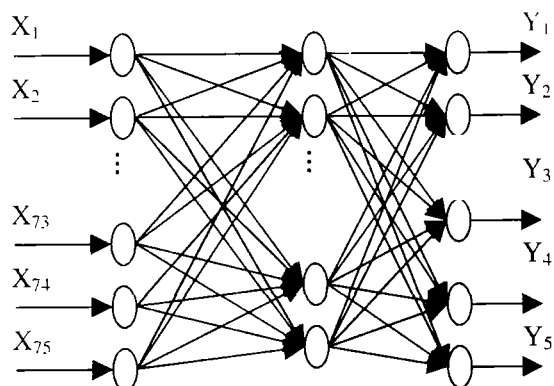
基金项目:国家“十五”科技攻关项目(2001BA701A22)

1 中南大学湘雅医院中西医结合研究所,湖南省长沙市湘雅路141号(410008)

2 中国中医研究院基础理论研究所博士后

收稿日期:2003-07-22;修回日期:2003-09-20

分两轮进行。第一轮以一组作为训练组,另一组测试。第二轮顺序相反。以两组训练测试结果合计表示为最终结果。网络训练开始时,首先设置好输入层、隐含层、输出层节点数、隐含层层数、学习率、理想输出及控制误差。当网络训练误差小于控制误差时,即认为网络已训练好。以其权值、阈值代入测试未知数据。



附图 抑郁症中医证型分类 BP 网络结构

3 结果

3.1 BP 网络训练结果:抑郁症 BP 网络训练肝郁气滞证 99.3% 不越界,肝郁脾虚证不越界率为 96.9%,心脾两虚证不越界为 94.5%,肝郁痰阻证不越界为 99.2%,肝肾阴虚证不越界为 99.3%,总体阳性率为 97.9%。

3.2 BP 网络测试结果:从附表中可以看出,对 1731 例抑郁症进行测试,肝郁气滞证测试阳性率为 78.4%,肝郁脾虚证阳性率为 73.5%,心脾两虚证阳性率为 66.5%,肝郁痰阻证阳性率为 64.0%,肝肾阴虚证阳性率为 73.9%,1731 例抑郁症总体阳性率为 72.5%。其中以肝郁气滞证阳性率最高,肝郁痰阻证、心脾两虚证阳性率偏低。

附表 BP 网络测试结果

中医证型	例数	误差越界例数	不越界例数	阳性率
肝郁气滞证	588	127	461	78.4%
肝郁脾虚证	487	129	358	73.5%
心脾两虚证	254	85	169	66.5%
肝郁痰阻证	264	95	169	64.0%
肝肾阴虚证	138	36	102	73.9%
合计	1731	472	1259	72.5%

4 讨论

人工神经网络是多学科交叉的前沿高技术领域,涉及信息学、计算机、生物学、数学等诸多领域。它是由大量简单的处理单元组成的非线性、自适应、自组织系统,是在现代神经科学研究成果基础上,试图通过模拟人类神经系统对信息进行加工、记忆和处理的方式设

计出的一种具有人脑风格的信息处理系统。它具有感知、识别、学习、联想、记忆、推理等功能,目前已广泛应用于经济、金融等领域。在医学领域中已用于疾病预测^[1]、疾病诊断分类^[2]、疾病耐药性分析^[3]等方面。

本研究数据来源于流行病学调查,采用大量的样本进行分析,以期展示研究对象的全部规律。充分的样本集可表达充分正确的规则,这一规律早已被证实和应用。而来自流行病学调查的大量样本,其全部特征无法用常规的统计学方法进行分析归纳,必须借助于先进的数理方法,对此,人工神经网络就是最佳方法之一。人工神经网络按其信息流向可分为前馈网络和反馈网络,本文采用前馈反向传播网络。

本研究采用 BP 网络对 1731 例抑郁症 5 种中医证型进行训练,仅 2.1% 不收敛,其中以心脾两虚证不收敛率较高,为 5.5%,总阳性率为 97.9%。这说明了以下两点:其一,本研究采用的 BP 网络性能良好,相对于普通统计学方法来说,具有非线性、容错性、函数逼近等优势特征。其二,来源于流行病学调查的数据具有较高的准确性与客观性,具有非线性可分的特征。对 1731 例抑郁症 5 种中医证型进行网络测试,以肝郁气滞证准确率最高,测试准确度为 78.4%,肝郁痰阻证、心脾两虚证准确率较低,测试准确度分别为 64.0%、66.5%,说明 BP 网络有很好的自学习、自适应及联想推理能力,在处理复杂的医学统计问题时可弥补常规统计学的不足。

在本研究训练和测试的过程中发现,肝郁气滞证识别的准确率最高,为最具典型特征的证型。心脾两虚证网络识别的准确率偏低,回顾原始资料发现,其中一部分数据接近肝郁脾虚,一部分无法判别证型,分析心脾两虚证易出现兼夹证,特征性不明显,与本文结果相符。从本研究初步结果可以看出,BP 网络能较好区分抑郁症 5 类中医证型,在中医证型分类识别中有一定价值,值得进一步研究。

参考文献

- 1 胡随瑜,张宏耕,郑林,等. 1977 例抑郁症患者中医不同证候构成分析. 中国医师杂志,2003,5(10):1312-1314.
- 2 丛爽主编. 神经网络、模糊系统及其在运动控制中的应用. 合肥:中国科学技术大学出版社,2001:13-15.
- 3 Orr RK. Use of an artificial neural network to quantitate risk of malignancy for abnormal mammograms. Surgery, 2001, 129(4):459-465.
- 4 Lo JY, Markey MK, Baker JA, et al. Cross-institutional evaluation of BI-RADS predictive model for mammographic diagnosis of breast cancer. AJR Am J Roentgenol, 2002, 178(2):457-463.
- 5 Draghici S, Potter RB. Predicting HIV drug resistance with neural networks. Bioinformatics, 2003, 19(1):98-107.