

基于RBF神经网络的老年痴呆症智能诊断研究

张会敏^① 叶明全^{*} 罗永钱^② 孟婷玮^① 陈玥珠^③

摘要 为验证单RBF神经网络更适用于老年痴呆症的预测诊断,通过仿真实验将单BP神经网络、单RBF神经网络、遗传算法优化BP神经网络及遗传算法优化RBF神经网络分别应用于老年痴呆症的预测诊断,建立这四种网络模型,并对四种网络模型的预测结果进行分析比较。仿真实验在Matlab软件平台上进行。结果表明:在老年痴呆症的预测诊断中,单RBF神经网络比单BP神经网络预测结果更好,建模时间更短。此外,单RBF神经网络与遗传算法优化的BP神经网络预测结果相同,但单RBF神经网络建模较为简单,预测结果更为稳定。而遗传算法对RBF神经网络优化作用不明显。因此,单RBF神经网络更适用于老年痴呆症的预测诊断,实际应用时可以此结论作为理论指导。

关键词 遗传算法 BP神经网络 RBF神经网络 老年痴呆症预测 数据挖掘

Doi:10.3969/j.issn.1673-7571.2015.06.011

[中图分类号] R741;TP391 [文献标识码] A


Study on Intelligent Diagnosis of Senile Dementia Based on RBF Neural Network / ZHANG Hui-min, YE Ming-quan, LUO Yong-qian, et al//China Digital Medicine.-2015 10(6): 38 to 41

Abstract In order to verify single RBF neural network is more suitable for the predictive diagnosis of senile dementia, through the simulation experiment, a single BP neural network, a single RBF neural network, a genetic algorithm to optimize BP neural network and a genetic algorithm to optimize RBF neural network are used to predict senile dementia, establishing of these four kinds of network model, then analyzing and comparing the forecasted results of these four kinds of network model. The simulation experiments were carried out on the platform of Matlab software, the results show that: in the predictive diagnosis of senile dementia, the single RBF neural network predictive results is higher than the single BP neural network, and the modeling time is shorter. Furthermore, the prediction results of the single RBF neural network is as the same as the genetic algorithm to optimize BP neural network, but the single RBF neural network model is relatively simple, and the prediction results are more stable. Therefore, diagnosis and prediction of the single RBF neural network is more suitable for senile dementia, and this conclusion can be used as a theoretical guide to the actual application.

Keywords genetic algorithm, BP neural network, RBF neural network, dementia disease prediction, data mining

Fund project National-level College Students' Innovative Training Program (No. 201310368027); Provincial-level College Students' Innovative Training Program (No. AH201310368027, AH201410368072); Key Project of Provincial-level Natural Science Research of Colleges and Universities in Anhui Province (No. KJ2014A266)

Corresponding author Doctor and Professor, Computer Teaching and Research Office, Wannan Medical College, Wuhu 241002, Anhui Province, P.R.C.

 **基金项目:** 国家级大学生创新训练项目(编号:201310368027);省级大学生创新训练项目(编号:AH201310368027, AH201410368072);安徽省高校省级自然科学研究重点项目(编号:KJ2014A266)

*通讯作者:皖南医学院计算机教研室,博士,教授,241002,安徽省芜湖市高教园区文昌西路22号

①皖南医学院2011级信息管理与信息系统专业本科生,241002,安徽省芜湖市高教园区文昌西路22号

②皖南医学院2012级信息管理与信息系统专业本科生,241002,安徽省芜湖市高教园区文昌西路22号

③皖南医学院2011级临床医学专业本科生,241002,安徽省芜湖市高教园区文昌西路22号

随着我国人口进入老龄化,老年痴呆症^[1]患者数不断增加,成为严峻的社会性问题。老年痴呆症的传统诊断方法主要有某些量表及CT、核磁共振等技术,这些方法多存在不足^[2]。因此,要求更为有效和方便的智能诊断方法。

近年来,人工神经网络在医学领域开始获得应用^[2-3],这为老年痴呆症的预测诊断提供了新思路。此外,医院电子病历储存了大量患者数据,从这些数据中挖掘出有用的信息成为医疗数据挖掘的热点^[4]。本文首先将单BP和单RBF网络分别应用于老年痴呆症的预测诊断,再利用遗传算法对电子病历中患者的相关数据进行特征值约简,将约简后的数据再分别应用于这两种网络。在Matlab软件平台上建立上述四种网络模型并进行仿真实验,然后对四种模型的预测结果和性能等进行分析比较,为后续研究提供理论准备和指导。

1 BP神经网络与RBF神经网络

人工神经网络(Artificial Neural Network, ANN)是模拟人脑的结构和功能,采用数学和物理方法进行研究形成的一种自适性信息处理系统。人工神经网络由大量的人工神经元^[1]组成(见图1)。

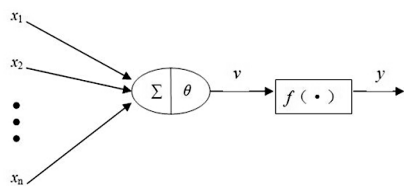


图1 人工神经元结构

神经元可按照任意形式进行连接,不同的连接形式构成不同拓扑结构的神经网络,具有不同的特性和功能。

1.1 BP神经网络 BP (Back

Propagation)神经网络是一种采用误差反向传播学习算法训练的多层感知机模,能在不确定映射关系的情况下,进行大量输入-输出模式的学习和存储。BP网络的拓扑结构由输入层、隐含层和输出层组成,层间神经元采用全连接方式,层内神经元之间无连接(见图2)。

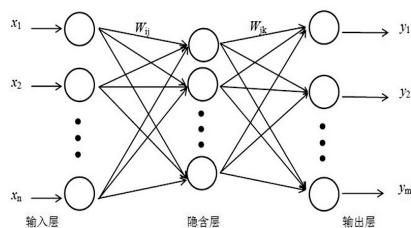


图2 三层BP神经网络拓扑结构

BP神经网络算法步骤如下^[5-6]。

Step 1: 随机初始化网络的权值和阈值; 初始化训练误差 $E=0$; 最小训练误差 E_{min} 初始化为一个很小的正数; 训练样本数总数设置为 P ; 学习率 η 初始化满足 $0 < \eta < 1$ 。

Step 2: 将训练样本从输入层输入,依次计算各隐含层和输出层神经元的输入与输出。

Step 3: 计算输出层误差 E 和各隐含层神经元的误差。

Step 4: 调整隐含层至输出层之间连接的权值以及输出层各神经元的阈值。

Step 5: 调整输入层至隐含层之间连接的权值以及隐含层各神经元的阈值。

Step 6: 检查是否完成一次训练,若辅助参数 $p < P$,则 $p=p+1$,转向步骤7; 否则,返回Step 2。

Step 7: 比较网络总体误差和最小训练误差,若 $E < E_{min}$,则训练结束; 否则, E 置为0, p 置为1,返回Step 2。

1.2 RBF神经网络 RBF (Radial Basis Function)神经网络采用的激活函数

是径向基函数。该网络输入至输出的变换是非线性的,而隐含空间至输出层空间的变换是线性的,因此学习速度很快,能够避免局部极小现象的出现,具有最佳逼近和全局最优的特点,适合解决分类问题。RBF网络的拓扑结构也包括输入层、隐含层和输出层三层(见图3)。

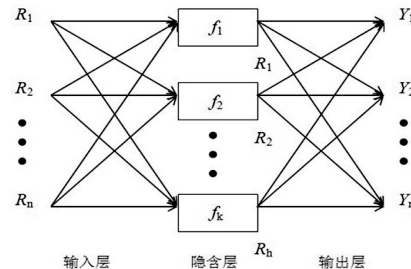


图3 RBF神经网络拓扑结构

RBF网络算法主要有三个参数要求解^[7]: 基函数的中心、方差以及隐含层到输出层的权值。常用的径向基函数是高斯函数,此时RBF网络的激活函数可以表示为:

$$R(x_p - c_i) = \exp\left\{-\frac{1}{2\sigma^2} \|x_p - c_i\|^2\right\} \quad (1)$$

式中, c 表示高斯函数的中心, σ 表示方差, $\|x_p - c_i\|$ 表示欧式范数。

RBF网络的输出公式为:

$$y_j = \sum_{i=1}^h w_{ij} \exp\left\{-\frac{1}{2\sigma^2} \|x_p - c_i\|^2\right\} \quad (2)$$

式中, x_p 表示第 p 个输入样本, c_i 为网络隐含层神经元的数据中心, w_{ij} 表示隐含层至输出层连接的权值, $i=1,2,\dots,h$ 表示隐含层的节点数, y_j 表示与输入样本对应的第 j 个输出神经元的实际输出值。

基函数的方差计算公式为:

$$\sigma = \frac{1}{p} \sum_j \|d_j - y_j c_i\|^2 \quad (3)$$

式中, d 表示样本的期望值。

RBF神经网络算法步骤(基于K-均值聚类方法求基函数中心)^[7]如下。

Step 1: 初始化网络,随机选取 h 个训练样本作为网络的聚类中心 $c_i(i=1,2,\dots,h)$ 。

Step 2: 根据 x_p 与聚类中心 c_i

间的欧氏距离,采用最近邻原则,将输入样本 x_p 分配到各个聚类域 ϕ_p ($p=1,2,\dots,P$)中。

Step 3: 计算各个聚类域 ϕ_p 中训练样本的平均值,作为新的聚类中心 c_i ,若 c_i 没有发生变化,则所得的 c_i 即为RBF神经网络最终的基函数中心,否则返回Step 1。

Step 4: 求解方差 σ_i :

$$\sigma_i = \frac{c_{\max}}{\sqrt{2h}} \quad i=1,2,\dots,h \quad (4)$$

式中 c_{\max} 为聚类中心之间的最大距离。

Step 5: 用最小二乘法计算隐含层到输出节点的连接权值:

$$w = \exp\left\{\frac{h}{c_{\max}} \|x_p - c_i\|^2\right\} \quad p=1,2,\dots,P; i=1,2,\dots,h \quad (5)$$

2 遗传算法优化BP与RBF神经网络

遗传算法 (Genetic Algorithms, GA) 是模拟生物界适者生存、优胜劣汰机制随机化搜索最优解的全局优化算法。GA首先采用随机化法初始化种群,然后根据适应度函数,采用选择、交叉及变异三种基本操作对个体进行筛选。个体中适应度较好的会被保留下来,适应度较差的就会被淘汰,这样产生的后代种群会更加适应环境。

GA优化BP/RBF神经网络算法GABP/GARBF步骤^[8-9]如下。

Step 1: 建立单BP/RBF模型。将所有可能的影响因素作为输入变量构建单BP/RBF网络模型。

Step 2: 随机初始化种群。随机取得值为“1”或“0”的N个串结构数据,每个数据为一个个体,N个数据构成一个群体,作为GA的初始种群。

Step 3: 计算个体的适应度。用适应度来衡量群体中各个体的优良程度,适应度越高的个体保留下来的可

能越大。本文个体的适应度为测试集数据输出误差的平方和倒数。

Step 4: 选择。采用比例选择算子计算群体中所有个体的适应度,使个体的适应度与其遗传到下一代种群的概率成正比。

Step 5: 交叉。对所有个体进行随机两两配对,采用单点交叉算子对于BP/RBF神经网络的输入变量进行降维,根据给定的概率重组一对个体作为后代。

Step 6: 变异。采用单点变异算子,随机选取变异点,交换每对个体变异点位置的基因。

Step 7: 优化输出。反复执行以上迭代过程,当输出结果满足终止条件时,末代输出的种变量就是相关性最大的最佳输入变量组合。

Step 8: 建立GA优化BP/RBF网络模型。利用Step 7得到的输入变量组合,重新建立BP/RBF网络模型,并进行仿真实验。

3 仿真实验

为比较单BP、单RBF、GABP及GARBF这四种网络在老年痴呆症预测诊断中的有效性,在Matlab R2010a环境下,利用Matlab软件建立这四种网络模型^[9-10],以电子病历中患者血常规数据的29项生理指标(GA优化后为29项中部分指标)作为网络输入,诊断结果作为网络输出(对诊断结果进行编码,用0表示诊断结果为非老年痴呆症,用1表示诊断结果为老年痴呆症),用已经确诊的多组历史数据对四种网络分别进行训练,使得网络通过权值记忆临床特征与老年痴呆症与否之间的对应关系。

实验样本为141例某医院电子病历系统中收集的患者数据(其中老年痴

呆症患者数据86例,非老年痴呆症患者数据55例),每一例数据都包含患者血常规中29项生理指标。将这141例数据中其中90例作为训练样本(其中老年痴呆症患者数据54例,非老年痴呆症患者数据36例),其余51例作为测试样本(其中老年痴呆症患者数据32例,非老年痴呆症患者数据19例)。

首先,利用Matlab神经网络工具箱中newff()函数建立29-15-1的3层BP网络模型,隐含层和输出层的传递函数分别为tansig和purelin,训练函数为trainlm,并设置网络的训练目标net.trainParam.goal=0.001,网络的最大训练次数net.trainParam.epochs=1000,网络的学习率net.trainParam.lr=0.1,其余参数采用默认值。对BP网络进行训练和测试,记录预测效果。然后利用遗传算法工具箱中GA对实验数据进行优化,优化过程适应度函数进化曲线如图4所示。根据优化后的数据重新建立BP网络并进行训练和测试,记录测试结果。

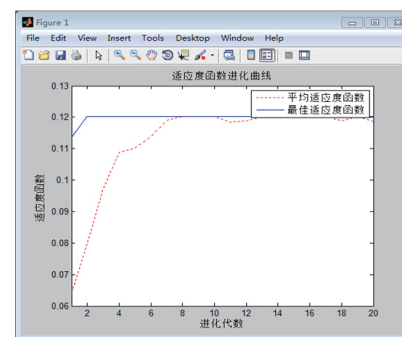


图4 GABP适应度函数进化曲线

其次,利用newrb()函数建立RBF网络模型。本文设置该网络的训练精度为0.001,其余参数采用默认值。对RBF网络进行训练和测试,记录测试结果。然后利用GA对实验数据进行优化,优化过程适应度函数进化曲线如图5所示。根据优化后的数据重新建立

RBF网络并进行训练和测试,记录测试结果。

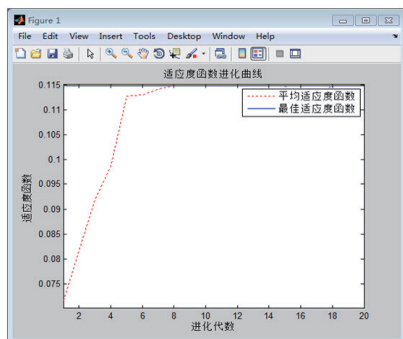


图5 GARBF适应度函数进化曲线

表1给出了四种模型对老年痴呆症及非老年痴呆症预测诊断的结果。由表1可以看出,单RBF网络预测结果较单BP网络更准确,其中老年痴呆症的确诊率由68.7500%上升为84.3750%,非老年痴呆症的确诊率由68.4211%上升为73.6842%。此外,单RBF网络与GABP网络对老年痴呆症的确诊率同为84.3750%,对非老年痴呆症的确诊率同为73.6842%。其次,GA对RBF网络的优化结果不明显,其中对老年痴呆症的预测确诊率由84.3750%下降为71.8750%,对非老年痴呆症的预测确诊率由73.6842%上升为89.4737%,建模时间由1.9656s上升为86.2636s。同时,在单BP、单RBF、GABP及GARBF这四种网络中,单RBF网络与GABP网络的分类准确率相同且最高,为80.3922%。但单RBF建模时间最短,相比GABP,单RBF建模时间为1.9656s,GABP为86.2636s。

表1 BP、RBF、GABP及GARBF四种模型预测结果比较

	诊断结果	确诊数	误诊数	确诊率(%)	误诊率(%)	分类准确率(%)	建模时间(s)
BP 模型	老年痴呆症	22	10	68.7500	31.2500	68.6275	2.3712
	非老年痴呆症	13	6	68.4211	31.5789		
RBF 模型	老年痴呆症	27	5	84.3750	15.6250	80.3922	1.9656
	非老年痴呆症	14	5	73.6842	26.3158		
GABP 模型	老年痴呆症	27	5	84.3750	15.6250	80.3922	86.2636
	非老年痴呆症	14	5	73.6842	26.3158		
GARBF 模型	老年痴呆症	23	9	71.8750	28.1250	78.4314	91.2636
	非老年痴呆症	17	2	89.4737	10.5263		

4 结语

为验证单RBF网络更适用于老年痴呆症的预测诊断,将单RBF网络与单BP网络、GABP网络及GARBF网络预测结果的确诊率、误诊率、分类准确率以及建模时间等进行比较。结果表明,单RBF网络比单BP网络分类准确率更高,确诊率更高,误诊率更低,建模时间更短。未经优化的RBF网络与优化后的BP网络预测效果相同,但单RBF网络模型的建立比GABP网络模型简单。同时,GA对RBF网络优化作用不明显。所以,在老年痴呆症的预测诊断中,单RBF网络的性能优于BP网络、GABP网络、GARBF网络,在实际应用中应优先考虑单RBF网络。本文结论可为基于人工神经网络的老年痴呆症智能预测诊断研究提供理论准备。

参考文献

- [1] 贾伟华,马颖,陈若陵,等.中国部分城乡社区老年痴呆症患病率及其影响因素研究[J].安徽医科大学学报,2012,47(8):944-947.
- [2] 张会敏,叶明全,孟婷玮,等.遗传算法优化BP神经网络的老年痴呆症智能诊断[J].中国数字医学,2014,9(7):81-84.
- [3] 郝涛,张智.基于BP神经网络的原发性肝癌CT图像纹理分析[J].中国数字医学,2013,8(8):73-76.
- [4] 吴信东,叶明全,胡东辉,等.普适医疗信息管理与服务的关键技术与挑战[J].计算机学报,2012,35(5):1-19.
- [5] 韩敏.神经网络基础[M].大连:大连理工大学出版社,2014.

工大学出版社,2014.

- [6] 马锐.神经网络原理[M].北京:机械工业出版社,2010.
- [7] 刘冰,郭海霞.MATLAB神经网络超级学习手册[M].北京:人民邮电出版社,2014.
- [8] 陈守平,董瑞,罗晓莉.MATLAB神经网络30个案例分析[M].北京:北京航空航天大学出版社,2010.
- [9] 龙辉,戚文艳.MATLAB遗传算法工具箱及应用[M].西安:西安电子科技大学出版社,2004.
- [10] 钟志芳.MATLAB神经网络设计与应用[M].北京:清华大学出版社,2013.

【收稿日期:2015-04-13】

【修回日期:2015-04-21】

(责任编辑:赵士洁)

业界观察

国内首个远程移动医疗系统在沪上线

5月30日,一场围绕“分级诊疗模式创新”的智慧健康论坛传出消息,国内首个远程移动医疗系统已在浦东南片区域正式上线,形成医医、医患智慧健康工程新模式。

2012年,浦东新区卫生计生委就提出新农合按人头支付,基于此项改革举措,以浦东医院为主导、11家社区卫生服务中心参与的医疗卫生协同网平台搭建初步完成。经过近3年实践,完成了1个平台即医疗卫生协同网平台,2个联动即医—医、医—患实时联动,3个协作即国际协作、系统协作、平台协作,4个共享即专家共享、临床共享、科研共享和教学共享的建设,带动社区提高技术。

浦东南片医疗卫生协同网总负责人余波介绍,区域内启用的国内首个远程移动医疗系统协同网,已经形成了医医、医患智慧健康工程新模式;同时,首个“云病理”疑难病理诊断平台,以平均50例/月运行至今,大数据下的一体化区域影像诊断中心,平均诊断5000例/月。

(来源:新华网)

OBSERVATION