

# 基于径向基神经网络用于中医舌诊诊断的研究

谢铮桂, 韦玉科, 钟少丹

(广东工业大学 计算机学院, 广东 广州 510090)

**摘 要:**研究基于聚类分析的径向基神经网络用于中医舌诊诊断, 构建一个中医舌诊智能诊断的神经网络模型。先对输入样本集进行数据归一化处理, 采用 K-means 聚类算法对样本集进行聚类分析, 得到聚类中心的位置和数量, 再构建 RBF 神经网络进行学习、训练。以肝病病证诊断进行仿真。实验结果表明: 该中医舌诊智能诊断模型具有诊断能力强、收敛速度快、泛化能力强等特点。因此, 基于聚类分析的径向基神经网络用于中医舌诊诊断的研究是可行的、有效的。

**关键词:**聚类分析; 径向基函数神经网络; 舌诊

**中图分类号:** TP183

**文献标识码:** A

**文章编号:** 1673-629X(2008)09-0242-02

## Research of RBF Neural Networks Based on Clustering Analysis in TCM Inspection of Tongue Diagnosis

XIE Zheng-gui, WEI Yu-ke, ZHONG Shao-dan

(Faculty of Computer, Guangdong University of Technology, Guangzhou 510090, China)

**Abstract:** Presents the RBF neural networks based on clustering analysis is applied in TCM inspection of tongue diagnosis, to construct a neural networks model of TCM inspection of tongue intelligent diagnosis. Put the input sample data set into normalization treatment. It uses K-means clustering algorithm to analyze the input sample data set and gains the clustering center's position and number and constructs the RBF networks to learn and train. It uses the hepatic disease symptom as simulation. The experimental result demonstrates that TCM inspection of the tongue diagnosis model has good diagnostic ability, fast convergence rate and good generalization ability. So the research of RBF neural networks based on clustering analysis in TCM inspection of tongue diagnosis is feasible and valid.

**Key words:** clustering analysis; Radial-Basis Function (RBF) neural networks; inspection of tongue

### 0 引言

中医舌诊因其在疾病的诊断和治疗中的显著作用, 在临床医学和生物医学上引起了广泛的关注和重视。舌诊是根据相关诊断规则进行诊断推理, 最终得到相应病证的过程。对于病证的诊断可以通过构建基于症状信息进行病证分类的人工神经网络来解决<sup>[1]</sup>。

径向基函数神经网络具有全局逼近性质, 且不存在局部最小问题, 计算量小、学习速度快等优点, 已经广泛应用于模式分类、系统辨识、函数逼近等<sup>[2]</sup>。基于聚类分析的 RBF 神经网络能够克服以往 RBF 网络存在的学习训练的难题: 高维数据问题。

文中构建的基于聚类分析 RBF 神经网络用于中

医舌诊诊断的建模过程由两个阶段组成: 第一阶段利用 K-means 聚类算法确定 RBF 网络隐层的聚类中心; 第二阶段建立 RBF 神经网络, 对输入舌象数据进行分析、学习、训练, 求输出层的权值矩阵, 实现舌诊自动诊断。

### 1 基于聚类分析的径向基神经网络

#### 1.1 径向基神经网络原理

RBF 网络即为径向基函数网络, 它是一种三层前向网络: 第一层为数据输入层, 第二层是隐层, 第三层是输出层<sup>[3]</sup>。如图 1 所示。

设输入层、隐层和输出层的节点数分别为  $M, N, J$ , 隐层单元选用高斯函数, 即

$$r_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{2a_i^2}\right) \quad 1 \leq i \leq N \quad (1)$$

其中,  $x = (x_1, x_2, \dots, x_M)$ ,  $C_i = (c_{i1}, c_{i2}, \dots, c_{iM})$  是隐层第  $i$  个单元的变换中心,  $a_i$  是对应第  $i$  个中心的形状参数, 它决定了该基函数围绕中心点的宽度。

收稿日期: 2007-12-16

基金项目: 国家自然科学基金资助项目(30472122)

作者简介: 谢铮桂(1980-), 女, 广东潮州人, 硕士研究生, 韩山师范学院助理实验师, 研究方向为神经网络; 韦玉科, 副教授, 博士, 研究方向为人工智能和计算机测控技术; 钟少丹, 硕士研究生, 韩山师范学院实验师, 研究方向为神经网络。

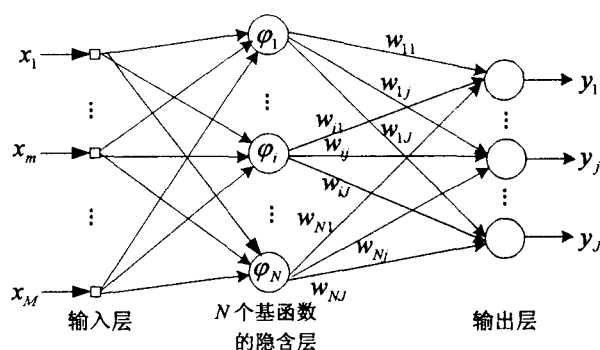


图1 RBF神经网络结构图

$$a_1 = a_2 = \dots = a_i = \dots = a_N$$

其中,  $a_i = \frac{d_{\max}}{\sqrt{2N}}$ ,  $N$  为隐层节点数,  $d_{\max}$  为取中心之间的最大距离。

RBF网络的输出层节点为线性处理单元, 其中第  $j$  个单元对应的输出为:

$$y_j = \sum_{i=1}^N w_{ji} r_i(x) + b_j \quad 1 \leq j \leq J$$

对于权值  $w$  和  $b$  的更新采用 LMS 算法:

$$w(k+1) = w(k) + 2ae(k)r(k)$$

$$b(k+1) = b(k) + 2ae(k)$$

其中,  $k$  为第  $k$  次迭代,  $a$  为学习率,  $e(k)$  为第  $k$  次迭代的误差。

在 RBF 神经网络中, 隐层中心的数量和宽度直接影响网络的逼近能力, 因此, 建立 RBF 网络的关键在于选择合适的 RBF 中心。

文中采用基于 K-means 聚类算法来选取聚类中心, 该算法具有通用性强、计算简单、性能良好等特点。

## 1.2 K-means 聚类算法

基于 K-means 聚类算法具体步骤如下<sup>[3]</sup>:

第一步: 初始化聚类中心, 即根据经验从训练样本集中随机选取  $I$  个不同的样本作为初始中心  $t_i(0)$  ( $i = 1, 2, \dots, I$ )。

第二步: 随机输入训练样本  $X_k$ 。

第三步: 寻找训练样本  $X_k$  离哪个中心最近, 即找到  $i(X_k)$  使其满足

$$i(X_k) = \arg \min_i \|X_k - t_i(n)\|$$

第四步: 调整基函数的中心。

$$t_i(n+1) = \begin{cases} t_i(n) + \eta[X_k(n) - t_i(n)] & i = i(X_k) \\ t_i(n) & \text{其他} \end{cases}$$

第五步:  $n = n + 1$ , 转到第二步, 直到学完所有的训练样本且中心的分布不再变化。

## 2 实验

以肝病病证诊断进行仿真, 建立基于聚类分析的 RBF 网络的中医舌诊诊断模型。如图 2 所示。模型前端部分为输入样本的“数据归一化”处理。后端部分为给出识别结果部分。它是根据输出层各节点的输出值, 利用最大值准则给出识别结果, 即哪类节点的输出值最大则相应的输出为 1, 其它节点相应的输出为 0<sup>[4]</sup>。模型中的中间部分是基于聚类分析的 RBF 神经网络。

将每一个病例的舌象特征参数作为一个输入样本。选取 4 类病证共 120 例舌象特征参数的样本数据作样本集。每个样本包括证型、舌象、主干形态、主干充盈度、质红、质绿、质蓝、舌下 R、舌下 G、舌下 B 等 33 项舌象特征参数作为输入样本的属性。样本数据分四类: 正常 30 例, 肝癌 30 例, 肝硬化 30 例, 乙肝 30 例。选取其中 80 例用于作训练集, 40 例用于作测试集。表 1 给出了部分训练集及其相应期望诊断结果。

表1 训练集及其相应诊断结果的片段

序号	诊断结果	证型	舌象	主干形态	主干充盈度	质红 R	绿 G	蓝 B	...
1	正常	正常组	淡红舌	单支干	下隆上平	189.00	110.00	113.00	
2	肝癌	气滞血瘀	绛紫舌	单支干	下隆上平	85.00	43.00	55.00	
3	肝硬化	肝瘀痰阻	淡紫舌	双支干	饱满弯曲	155.00	118.00	102.00	
4	乙肝	肝瘀痰阻	瘀点舌	双支干	下隆上平	103.00	57.00	103.00	
...									

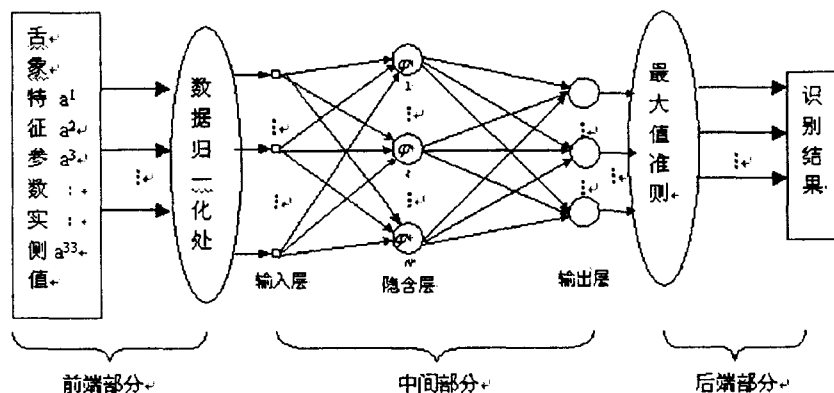


图2 舌象识别的神经网络模型

先对 120 例数据进行数据归一化处理, 将所有数据的值转化在  $[-1, 1]$  之间。数据归一化的目的是可避免出现“奇异输入样本”, 可使网络的权值和阈值在较短的时间内易于修改, 并有足够的变化量<sup>[5]</sup>。

将 80 例数据采用 K-means 算法进行聚类, 得到 16 个聚类中心, 因此, 建立一个输入层节点数为 33, 隐层节点数为 16, 输出层节点数为 4 的 RBF 神经网络模型。学习过程的误差曲线如图 3 所示。将 40 例样本

(下转第 247 页)

无信标的星型配置网络。对于有信标的网络配置也是接下来要着手解决的问题。

参考文献:

[1] ZigBee Specification v1.0[S].[s.l.]:Zigbee Alliance,2004.  
[2] IEEE Std 802.15.4[S].2003.  
[3] Freescale. MC13192RM Rev1.4 [S].2005.  
[4] 蒋 廷,赵成林.紫蜂技术及其应用[M].北京:北京邮电大学出版社,2006.  
[5] 周立功,张 华.深入浅出 ARM7[M].北京:北京航空航天大学出版社,2005.  
[6] 许 森,黄 东.基于 MC13192 的无线传感网络节点设计[J].单片机与嵌入式系统应用,2007 (3):48-50.  
[7] 吴晓然.基于 MC13192 的 Zigbee 系统设计[J].工业控制计算机,2007(4):19-20.  
[8] 王 东,张金荣,魏 延.利用 Zigbee 技术构建无线传感器网络[J].重庆大学学报,2006(8): 95-97.

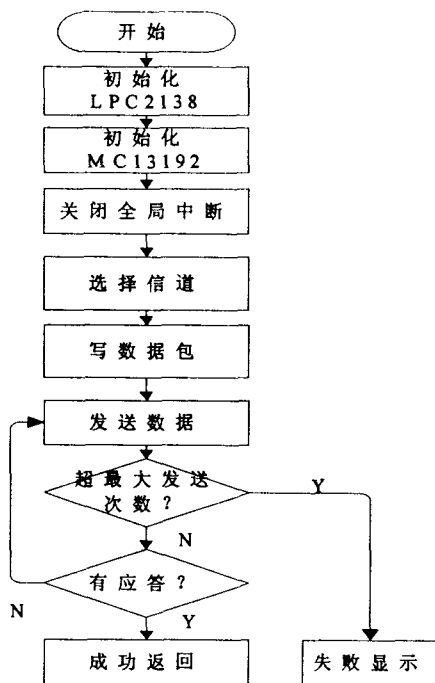


图 6 发送数据流程图

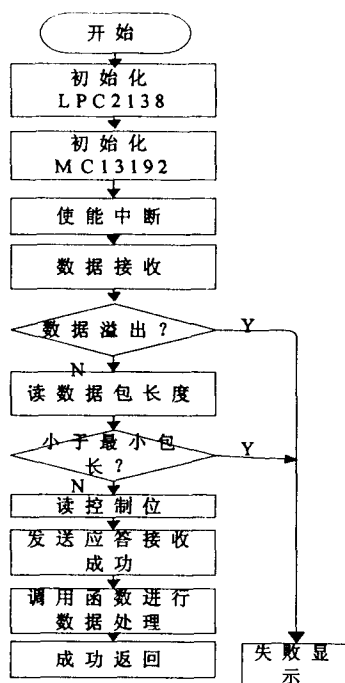


图 7 接收数据流程图

(上接第 243 页)

用于测试,识别率为 90%。表 2 为对应表 1 样本集的

实际诊断结果。

3 结束语

用基于聚类分析的 RBF 神经网络构建中医舌诊诊断模型。旨在对大数据、高维数据样本能快速聚类,以提高收敛速度,保证整个 RBF 网络具有良好的泛化能力。

实验结果表明该模型能有效地学习输入样本数据,具有诊断能力较强、收敛速度较快等特点。

参与文献:

[1] 吴 芸,周 昌,张志枫.中医舌诊八纲辨证神经网络知识库构建[J].计算机应用研究,2006,23(6):188-189.  
[2] 岳彩青,常青美,庞学民,等.基于聚类分析的 RBF 网络建模方法及应用的研究[J].计算机仿真,2006,23(1):120-123.  
[3] 高 隽.神经网络原理及仿真实例[M].第 2 版.北京:机械工业出版社,2007:55-62.  
[4] 王炳和,相敬林.基于神经网络方法的人体脉象识别研究[J].西北工业大学学报,2002,20(3):454-457.  
[5] 冯 定.神经网络专家系统[M].北京:科学出版社,2006: 66-69.

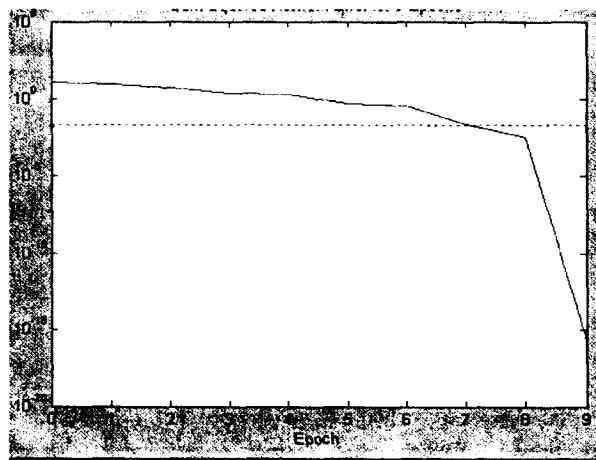


图 3 误差曲线图

表 2 部分样本的测试结果

序号	病证	期望输出结果	实际诊断结果
1	正常	[1,0,0,0]	[1,0,0,0]
2	肝癌	[0,1,0,0]	[0,1,0,0]
3	肝硬化	[0,0,1,0]	[0,0,1,0]
4	乙肝	[0,0,0,1]	[0,0,0,1]