

改进的 RBFNN 在运动员竞技状态预测中的应用

张乐¹,魏振钢¹,姚晓晓¹,杨红云²

ZHANG Le¹,WEI Zhen-gang¹,YAO Xiao-xiao¹,YANG Hong-yun²

1.中国海洋大学 计算机系,山东 青岛 266100

2.青岛酒店管理学院,山东 青岛 266100

1.Department of Computer Science and Technology,Ocean University of China,Qingdao,Shandong 266100,China

2.Qingdao Hismile College,Qingdao,Shandong 266100,China

E-mail:xiaole2001033536@163.com.cn

ZHANG Le,WEI Zhen-gang,YAO Xiao-xiao,et al.Application of RBF neural network based on mixture clustering in forecasting of athlete's sports state.Computer Engineering and Applications,2008,44(9):217-219.

Abstract: This paper presents an improved RBF neural network which based on fuzzy system model.The first determine the centers' number of RBF using subtractive clustering method,the second optimize the position of the centers and centers' width in RBF according to fuzzy C-mean algorithm,the last design and train the RBF neural network depending on the result of samples clustering.The neural network is used to forecast the sports state of the tennis athletes.The result of this algorithm is effective and has higher precision,and this model can be available to the domain.

Key words: radial basis function neural network;subtractive clustering;fuzzy C-mean algorithm;sports state;forecasting

摘要:提出了一种改进的径向基函数(RBF)神经网络,该神经网络以模糊系统模型为基础。首先利用减法聚类算法确定径向基函数的中心数,然后通过模糊C均值聚类算法优化基函数中心与宽度,最后依据样本数据的聚类结果设计RBF神经网络并进行训练。将该神经网络应用于网球队运动员的竞技状态的预测。仿真结果表明:该算法先进有效、具有较高的精度,用其建立的模型具有较强的实用性。

关键词:径向基神经网络(RBFNN);减聚类算法;模糊C均值算法;竞技状态;预测

文章编号:1002-8331(2008)09-0217-03 **文献标识码:**A **中图分类号:**TP18;TP183

1 引言

网球运动是一项技巧性极强的比赛项目,运动员的技、战术特点是取得比赛胜利的关键因素。由于在评判运动员的竞技状态时要综合考虑运动员的诸多技战术指标,因而仅仅依靠教练员主观的观察和判断很难实现准确、合理的评判、预测运动员的竞技状态。随着人工智能技术的不断发展,基于神经网络的预测模型逐渐被提出。径向基函数(RBF)网络是一种新颖有效的前馈神经网络,由于该网络具有结构自适应确定且输出层是对中间层的线性加权,使得该网络避免了像BP网络那样繁琐冗长的计算,同时该网络还具有较高的运算速度和外推能力,使得网络有较强的非线性映射功能^[1]。因此广泛应用于模式识别、函数逼近、自适应滤波、非线性时间序列预测等方面。

在RBF神经网络的训练过程中,确定径向基函数的数据中心和宽度是一个关键问题^[2]。常见的选取RBF中心的方法有:随机选取中心法、自组织选取中心法、有监督选取中心法和正交最小二乘法^[3]。但这些方法都根据先验知识来事先确定网络的中心个数,虽然根据Kolomogorov定理表明前馈神经网络的

中间层节点数理论计算值是 $2n+1$ (n 为输入层节点数)。但此定理在径向基函数的中心选取中却达不到预期效果,因为中心点的个数虽然知道,但在大样本集合中选取哪几个样本作为中心点却无法得知,同时中心数不仅与网络输入层的节点数有关,而且与输入输出样本集的信息有关。

本文引入一种混合聚类算法即减法聚类和模糊C均值聚类相结合的算法用于指导聚类学习,从而优化了径向基函数的中心位置及其宽度参数。最后将改进的RBF神经网络构造模型,对网球运动员的竞技状态进行了预测。

2 RBF神经网络

RBF神经网络是由输入层、中间层和输出层构成的三层前向网络,其结构如图1所示^[4]。

中间层采用径向基函数作为激励函数,其输出表达式为:

$$f_n(X)=w_0 + \sum_{i=1}^n w_i + \sum_{i=1}^n w_i \phi(\|X-c_i\|) \quad (1)$$

基金项目:奥运科技攻关项目。

作者简介:张乐(1982-),男,在读研究生,主要研究方向软件工程、数据挖掘;魏振钢(1962-),男,教授,研究生导师,主要研究方向:软件工程,数据挖掘,计算机应用开发;姚晓晓(1982-),女,在读研究生,主研方向:软件工程,数据挖掘,计算机应用等。

收稿日期:2007-07-12 **修回日期:**2007-09-17

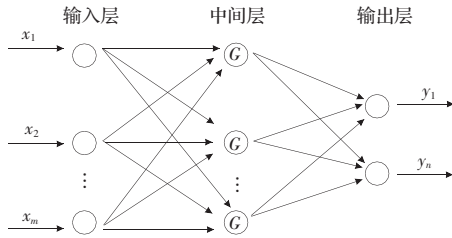


图1 径向基(RBF)神经网络结构图

式(1)中: w_0 表示的为输出节点的阈值, w_i 表示输出节点的权值, $\| \cdot \|$ 表示欧式空间距 $\phi(x)$ 为径向基函数, 一般为高斯函数:

$$\phi(\|X - c_i\|) = \exp\left(-\frac{\|X - c_i\|^2}{2\sigma_i^2}\right) \quad (2)$$

其中, c_i 与 σ_i 分别表示径向基函数的中心和宽度参数。因此采用高斯函数作为径向基函数的 RBF 网络需要确定参数中心 c_i 、宽度 σ_i 、中心的数目 n 以及输出单元的权值。由于输出单元是线性单元, 所以其权值可以通过最小二乘法求得。所以参数 c_i 、 σ_i 、 n 的确定是建立 RBF 神经网络的关键。

3 减法聚类

减法聚类是用来自动估计数据中的聚类个数及其位置的单次算法^[5]。首先将 n 个 m 维的样本数据 $\{x_1, x_2, \dots, x_n\}$ 作为聚类中心的候选点。然后, 求出每个聚类中心的密度指标。 x_i 处的密度指标定义为:

$$D_i = \sum_{j=1}^n \exp\left(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2}\right) \quad (3)$$

r_a 表示为 x_i 的一个邻域半径。当计算完每个样本数据密度指标后, 选择具有最高密度指标的数据作为第一个聚类中心, 假设 x_{c_1} 为选中的数据, D_{c_1} 为其密度指标。则每个数据 x_i 的密度指标需用公式:

$$D_i = D_i - D_{c_1} \exp\left(-\frac{\|x_i - x_{c_1}\|^2}{(r_b/2)^2}\right) \quad (4)$$

进行修正。其中 r_b 表示为一个密度指标函数显著减小的邻域。为了避免出现相距很近的聚类中心, 一般取 $r_b = 1.5r_a$ 。修正了每个数据的密度指标后, 选定下一个聚类中心 x_{c_2} , 再次修正数据所有密度指标。该过程不断重复, 直至 $\frac{D_{c_i}}{D_{c_1}} < \delta (\delta = 0.5)$ ^[6] 停止操作, 即可确定聚类的个数。

4 模糊 C 均值聚类原理

在各种聚类算法中, 模糊 C 均值(FCM)聚类算法的应用较为广泛。它的思想就是使得被划分到同一类的对象之间的相似度最大, 而不同类之间的相似度最小。模糊 C 均值算法是用隶属度确定每个数据点属于某个聚类的程度的一种聚类算法, 是普通 C 均值算法的改进^[7]。

FCM 把 N 个 n 维特征矢量 $x_i (i=1, 2, \dots, N)$ 分为 c 类, 并求每组的聚类中心, 使得非相似性指标的价值函数达到最小。使得每个给定数据点用值在 0, 1 间的隶属度来确定其属于各个组的程度。与引入模糊划分相适应, 隶属矩阵 U 允许有取值在 0, 1 间的元素。不过, 加上归一化规定, 一个数据集的隶属度

的和总等于 1。

设 $X = \{x_1, x_2, \dots, x_N\}$ 是 N 个 n 维特征矢量作为待测的数据集, 采用公式(5)作为聚类的目标函数, 在迭代寻优过程中逼近该目标函数的最小值

$$J_m(U, V) = \sum_{j=1}^N \sum_{i=1}^c u_{ij}^m d_{ij}^2 \quad (5)$$

式中, U 为隶属度矩阵; V 为 X 的聚类中心集合; u_{ij} 为样本 x_j 与第 i 类的聚类中心 v_i 的隶属度; d_{ij} 为欧氏距离 $\|x_j - v_i\|$ 。

5 基于混合聚类的 RBF 神经网络的学习算法

在 RBF 神经网络训练中, 中心的数目 n 、中心 c_i 及宽度 σ_i 的确定是一个关键问题, 传统的做法是将网络中间层的节点数等于训练样本的数目, 每个训练样本分别作为基函数的中心, 当输入样本的数目很大时, 数目庞大的中间层节点会极大地影响网络的学习效率和预测的精度。本文采用减法聚类与模糊 C 均值聚类相结合的算法将特征相近的训练样本进行归类处理, 确定样本的最佳聚类数目作为中间层的节点数, 并且获取了各个聚类的中心作为基函数的中心。该方法的优点在于无先验数据集经验的基础上, 经计算能自动地提供网络中间层的节点数, 大大减少了中间层的节点数目, 从而降低了网络的复杂度。

5.1 算法的基本步骤

步骤 1 将所有的样本数据进行归一化处理;

步骤 2 依据公式(3)计算出 n 个样本数据的密度指标, 之后在所有的密度指标中找出密度指标最高的一个作为第一个聚类中心 x_{c_1} ;

步骤 3 找出第一个聚类中心 x_{c_1} 后, 根据公式(4)进一步计算剩下的 $n-1$ 个数据的密度指标, 再找出其中的最高指标值作为第 2 个聚类中心 x_{c_2} , 依此类推, 一直找到 c 个聚类中心;

步骤 4 设定初始模糊分类矩阵 $U^{(0)}$;

步骤 5 利用 $U^{(s)}$ 结合式(6)计算 c 个聚类中心集合 $\{v_i^{(s)}\} (i=1, 2, \dots, c)$, 取 $1 < m \leq 5$:

$$v_i^{(s)} = \frac{\sum_{j=1}^N u_{ij}^m x_j}{\sum_{j=1}^N u_{ij}^m} \quad (6)$$

步骤 6 调整 $U^{(s)}$ 为 $U^{(s+1)}$ 。对 $j=1$ to N

如果 $I_j = \Phi$, 那么 $u_{ij}^{(s+1)} = \sum_{i=1}^c \left(\frac{d_{ij}^{(s)}}{d_{ij}^{(s+1)}}\right)^{\frac{2}{m-1}}$, 否则, $u_{ij}^{(s+1)} = 0, \forall i \in \bar{I}_j$

并取 $\sum_{i \in I_j} u_{ij}^{(s+1)} = 1$, 其中 $I_j = \{i | d_{ij} = 0\}, \bar{I}_j = \{1, 2, \dots, c\} - I_j$;

步骤 7 如果 $\|U^{(s)} - U^{(s+1)}\| < \varepsilon (\varepsilon > 0)$, 停止否则 $s = s + 1$ 返到步骤 4;

步骤 8 通过模糊 C 均值优化好的中心位置 $\{x_{c_1}, x_{c_2}, \dots, x_{c_c}\}$ 结合 $r_i = \frac{1}{|I_i|} \sum_{x \in I_i} \|x - x_{c_i}\|$ 求出网络中心的宽度。其中 $I_i = \{x | \|x - x_{c_i}\| < \|x - x_{c_k}\|, \forall k \neq i\}$;

步骤 9 采用最小二乘法求出输出单元的权值;

步骤 10 整个学习算法结束。

6 基于改进的RBF神经网络在网球运动员竞技状态中的预测

网球运动员的竞技状态主要反映在运动员训练、比赛中的技、战术指标的信息上。如网球单打项目中运动员的发球得分率、接发球成功率、底线击球得失分比率、中前场拦截得失分比率这4项技、战术指标综合在一起即可反映出该运动员的竞技状态。本文选取某位运动员在一段时间内训练中记录的指标数据作为实验的样本数据。(如表1所示)其中运动员状态是通过有多年经验的教练员根据该运动员的平时训练比赛分析所得,分为3种状态:较好、一般、较差分别以1,0,-1表示。

表1 某运动员各项技、战术指标及专家评判状态

序号	发球/%	接发球/%	底线/%	中前场/%	状态
1	21.3	91.1	17.6	4.3	1
2	16.7	54.3	15.4	3.8	-1
3	17.8	89.1	14.3	2.7	0
4	22.1	83.9	13.2	4.1	1
5	19.3	67.1	13.6	2.8	0
6	15.6	61.1	14.6	4.7	-1
7	17.9	77.3	12.3	3.2	-1
8	19.7	45.7	16.7	4.8	0
9	20.6	81.4	15.4	3.5	0
10	26.3	78.3	13.7	2.9	0
11	25.4	78.5	13.2	3.7	0
12	22.1	81.2	13.9	3.2	0
13	14.2	87.9	13.4	4.0	0
14	22.2	78.4	17.8	4.2	1
15	24.7	91.2	15.7	4.5	1
16	20.8	90.2	15.4	4.0	1
17	21.2	87.5	16.9	3.5	1

首先通过文中介绍的混合聚类算法求出表1中17个样本数据的6个聚类中心,其结果如表2所示。

表2 由聚类算法得到的神经网络中心表

第一类中心	21.023	88.899	15.301	4.0045
第二类中心	24.737	79.043	13.491	3.4516
第三类中心	21.791	84.484	17.465	4.2184
第四类中心	20.493	81.748	15.211	3.4825
第五类中心	18.259	76.76	12.965	3.1318
第六类中心	16.933	56.375	15.289	4.4664

然后将学习样本的集合通过公式: $r_i = \frac{1}{|I_i|} \sum_{x \in I_i} \|x - x_{ci}\|$ 求出网络中心的宽度即径向基函数的参数值。聚类的中心数即为中间层网络的节点数。

最后设计RBF神经网络,该网络出入层节点数为4个,中间层节点数为6个,输出节点数为1个。将设计好的RBF网络

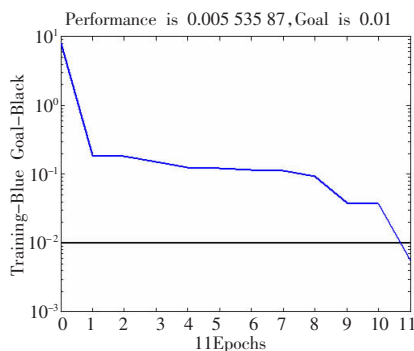


图2 神经网络训练误差曲线图

对表1中的17个样本数据进行训练。在误差为0.01的前提下只需迭代11步就可实现目标。

为了验证该神经网络的预测的正确性,选取表1中的5项数据作为检测数据,将其预测结果与传统的RBF预测结果进行比较(表3所示)得到图3。预测结果令人满意。其预测值与实际值在数值上几乎完全吻合,平均预测误差率由0.083降到0.0432。其拟合精度和预报精度为网球教练员对运动员竞技状态的科学评判和预测提供了有效的方法和手段。

表3 改进训练算法后的RBF预测结果

序号	实际值	传统RBF预测	改进RBF预测
1	1	0.982	0.994
3	0	-0.078	-0.092
7	-1	-1.142	-1.021
13	0	0.012	0.008
16	1	1.165	1.089

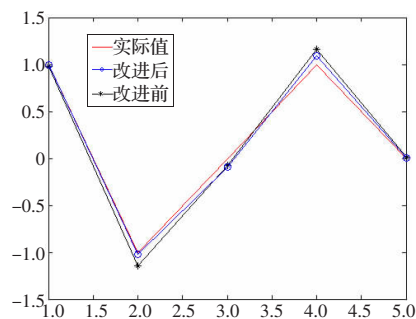


图3 预测结果与实际结果比较图

7 结束语

在减法聚类将样本数据进行分类处理,获得样本数据的聚类数目及其聚类的基础上通过模糊C均值聚类方法将聚类中心作进一步的优化,从而得到RBF神经网络中所需的网络中心宽度及中间层的参数。最后应用改进后的RBF神经网络对网球运动员的竞技状态进行预测,其结果与传统的RBF网络进行比较,结果表明:该方法训练时间短,收敛速度快、分类精度高、易于得到最小结构、在学习过程中不易陷入局部极小点等优点,有利于实现实时分析,该方法在运动竞技领域预测有良好的应用前景。

参考文献:

- [1] Moody J, Darken C. Fast learning in networks of locally-tuned processing units[J]. Neural Computation, 1989, 1: 281-294.
- [2] 王旭东, 邵惠鹤. RBF神经网络在非线形系统建模中的应用[J]. 控制理论与应用, 1997, 14(1): 59-66.
- [3] 朱明星, 张德龙. RBF网络基函数中心选取算法的研究[J]. 安徽大学学报: 自然科学版, 2000, 24(1).
- [4] Hahn M T, Beale M H. 神经网络设计[M]. 北京: 机械工业出版社, 2002: 285-310.
- [5] 王立新. 模糊系统与模糊控制教程[M]. 北京: 清华大学出版社, 2003.
- [6] Pal N R, Chakraborty D. Mountain and subtractive clustering method: improvements and generalizations[J]. International Journal of Intelligent Systems, 2000, 15(4): 329-341.
- [7] Dunn J C. A fuzzy relative of the ISODATA process and its use in detecting compact well separated cluster[J]. J Cybernet, 1974, 3: 32-57.
- [8] 孙即祥. 现代模式识别[M]. 长沙: 国防科技大学出版社, 2003.