

文章编号: 1001-0920(2005)12-1411-04

一种新的基于构造型 RBF 神经网络分类算法

黄国宏, 邵惠鹤

(上海交通大学 自动化研究所, 上海 200030)

摘 要: 依据 RBF 神经元模型的几何解释, 提出一种新的构造型神经网络分类算法. 首先从样本数据本身入手, 通过引入一个密度估计函数来对样本数据进行聚类分析; 然后在特征空间里构造超球面, 以逼近样本点分布的几何轮廓, 从而将神经网络训练问题转化为点集“包含”问题. 该算法有效克服了传统神经网络训练时间长、学习复杂的缺陷, 同时也考虑了神经网络规模的优化问题. 实验证明了该算法的有效性.

关键词: 模式识别; 神经网络; 最大密度覆盖; M-P 神经元; 构造型神经网络

中图分类号: TP18 **文献标识码:** A

Classification Algorithm Based on Constructive RBF Neuron Networks

HUANG Guo-hong, SHAO Hui-he

(Institute of Automation, Shanghai Jiaotong University, Shanghai 200030, China. Correspondent: HUANG Guo-hong, E-mail: h. guohong@163.com)

Abstract According to the geometrical representation of RBF neural model, a classification algorithm is proposed. Starting with the sample data directly, clustering analysis is proceeded by introducing a density function. And then hyperspheres are constructed to draw up the distribution of the sample data in feature space. The training problem of neural networks can be transformed into the “including” problem of a point set. The proposed algorithm can reduce the long training time and learning complexity of traditional neural networks. At the same time, the optimization of the neural network is also considered and computer simulation results show that the proposed neural network is quite efficient.

Key words: Pattern recognition; Neural networks; Max density covering; M-P neuron; Constructive neural networks

1 引 言

人工神经网络因其具有极强的自学习和分类能力, 在模式识别领域中得到了广泛的应用^[1-5]. 但由于每个神经元行为的数学描述都是一个多变量的非线性函数方程, 组成网络后的庞大多变量非线性方程组难以进行数学分析^[6]. 虽然很多学者提出了多种数学分析方法^[7], 但复杂多变量非线性方程组的处理与分析始终是个难题, 它限制了人工神经网络从分析角度的深入发展. 目前, 用数学解析的方法对神经网络行为进行精确分析还难以实现.

为改善对神经网络行为的认识和研究中的“黑

匣子”式难以处理的状态, 许多学者从几何角度对人工神经网络的行为进行了分析和研究^[8-11]. 文献[8, 9]试图利用 M-P 神经元的几何解释在多维空间里构造一组超平面来分析神经网络的行为, 遗憾的是, 随着超平面个数和维数的增加, 超平面之间的相互交叉变得异常复杂, 使得对神经网络的分析几乎无法进行. 文献[10]也利用 M-P 神经元的几何解释, 巧妙地构造出一个“球领域”模型, 从而将神经网络的训练问题转化为点集的覆盖问题, 克服了神经网络训练时间长和收敛性不确定的难题. 在此基础上, 文献[11]进一步提出了基于最大密度覆盖的神

收稿日期: 2004-11-05; 修回日期: 2005-01-24

作者简介: 黄国宏(1975-), 男, 内蒙古鄂尔多斯人, 博士生, 从事计算机视觉、图像识别等研究; 邵惠鹤(1936-), 男, 浙江宁波人, 教授, 博士生导师, 从事多变量约束控制、现场总线等研究.

神经网络训练算法,大大降低了学习的复杂度

受以上思想的启发,本文提出一种新的基于RBF神经元的构造型神经网络分类算法.通过引入一个密度估计函数来确定未被“包含”样本的最大密度点,依次构造出一系列具有不同优先级别的超球面.该算法不受先验知识和初始种子样本的选择以及样本分布的影响,算法简单,几何意义明确.通过对低维螺旋曲线和高维图像进行分类实验,结果表明,本文算法相对于传统的神经网络,具有训练时间短、正确率高的优点.

2 神经元模型的几何解释

广义地看待一个神经元,其数学模型可描述为^[6]

$$y = f[\Phi(x_1, x_2, \dots, x_n) - \theta], \quad (1)$$

其中: $X = (x_1, x_2, \dots, x_n)^T$ 为输入向量, θ 为阈值, $f(\cdot)$ 为神经元激励函数, y 为神经元输出. 神经元的基本运算规则,如对于BP网络的神经元

$$\Phi(x_1, x_2, \dots, x_n) = \sum_{i=1}^n w_i x_i, \quad (2)$$

其中 w_i 为神经元的权值.不难看出,激励函数的基 $\left[\sum_{i=1}^n w_i x_i - \theta \right]$ 就表示输入空间中输入点离一个超平面的距离,该超平面的方程为

$$\sum_{i=1}^n w_i x_i - \theta = 0 \quad (3)$$

如果激励函数是阶跃函数,则神经元的功能就是在多维空间中作一个超平面,输入位于此超平面的一侧时其输出为1,而另一侧时输出为0.

对于RBF网络的神经元,其数学模型可表示为

$$y = f \left[\sum_{i=1}^n (x_i - w_i)^2 - \theta \right]. \quad (4)$$

如果式(4)RBF神经元的激励函数 $f(\cdot)$ 为阶跃函数,则该神经元相当于在输入空间中以核心 $W = (w_1, w_2, \dots, w_n)^T$ 为球心,以 θ 为半径作一个超球面,当输入点在此超球面内时输出为0,在超球面以外时输出为1.可以看出,激励函数的基就表示输入空间中输入点离一个核心点的距离.

显然,超平面或超球面的几何概念对于帮助人们对神经网络行为的认识与分析是十分有效的.在模式识别中,将每个神经元看作是在多维空间中作的一个超平面或超球面,依次来划分输入空间的样本点.而如果把先分离的某些样本点挖掉不再参加后面的划分,就像一只苹果有许多不同颜色的污点,先靠边切一刀把若干个同一颜色的污点切掉,并将这切下的一片苹果拿走,则依次一刀刀地切下去,总能把所有污点都分离在许多片切下的苹果中.每切

一片至少能切去一个或几个同色的污点.如果每切一片都尽量切下最多的同颜色的污点,那么便可实现以最少的切片把所有的污点都除去.

正是基于这一几何常识,本文提出一种基于RBF神经元模型的构造型前向神经网络分类算法.首先在样本空间中找到在同类样本中具有最大密度的样本点,并以该样本点为球心 W 构造超球面,使之尽量多地“包含”同类的样本,而不包含异类样本.然后,将落入该超球面内的训练样本去掉,不参加以后的运算.如此重复,每次在剩余的样本中寻找同类样本的密度最大的样本点,构造超球面,并去掉落入其中的样本点,直到所有的样本点都被去掉为止.每个超球面都给予相应的类别标号,同时,给先得到的超球面赋予较高的优先级别.因此,对于训练样本,只要构造若干超球面,将它们分别“包含”,然后在判断某一样本的类别时,只要看它被哪个优先权较高的超球面“包含”,则该超球面对应的类别就是该样本的类别.

3 构造型RBF神经网络分类算法

给定一个训练样本集 $K = \{(X_{i,j}, y_i), 1 \leq i \leq c, 1 \leq j \leq \text{num}_i\}$, 其中 c 为训练样本的类别数, num_i 为第 i 类样本的样本个数, $X_{i,j} \in R^n$ 为属于第 i 类的第 j 个样本向量, y_i 为其输出的类别标志.

对于任一训练样本 $X_{i,j}$, 它在同类样本中的密度估计函数的定义为

$$D_{i,j} = \frac{\text{num}_i}{n} \exp \left[- \frac{X_{i,j} - X_{i,n}}{(r_{i,j}/2)^2} \right], \quad (5)$$

$$r_{i,j} = \min_k (X_{i,j} - X_{k,n}). \quad (6)$$

其中正数 $r_{i,j}$ 为样本 $X_{i,j}$ 到异类样本的最小欧氏距离,它定义了一个以 $X_{i,j}$ 为中心的区域,这样,在半径 $r_{i,j}$ 以外的样本向量对 $X_{i,j}$ 的密度值 $D_{i,j}$ 几乎没有影响.显然,如果一个训练样本有较大的密度值,则其周围一定有较多的相同类别的样本点.

下面给出神经网络分类器的训练学习算法:

1) 计算所有未被“包含”的样本数据点在同类样本中的密度 $D_{i,j}$, 找出其最大值.这里仍用 $X_{i,j}$ 表示具有最大密度的样本数据点,它对应的到异类样本的最小距离为 $r_{i,j}$.

2) 以样本点 $X_{i,j}$ 作为球心 $W = (w_1, w_2, \dots, w_n)^T$, 以 $\alpha r_{i,j}$ 作为阈值 θ 作超球面

$$\sum_{i=1}^n (x_i - w_i)^2 - \theta = 0,$$

其中: $X = (x_1, x_2, \dots, x_n)^T$ 为球面上任意一点, $0 < \alpha < 1$ 为一可调参数.

3) 求出满足 $\sum_{i=1}^n (x_i - w_i)^2 - \theta < 0$ 的样本点,

并将其从训练样本中去掉,不参加以后的运算,同时设置该超球面的类别标号和优先级别,原则是越先得到的超球面对应的优先级别越高

4) 返回 1) 继续重复执行,直到所有的样本都被去掉

综上,可将神经网络分类器的设计问题转化为训练样本的“包含”问题,即用一组超球面去“包含”具有相同类别的训练样本,而不同类别的样本点被不同组的超球面“包含”;也就是一个神经元(w, θ)在几何意义上对应一个超球面。取神经元的激励函数 $f(\bullet)$ 为阶跃函数,当一个样本落在超球面里,则神经元的输出为 0, 否则为 1。一个神经元就充当一个类别的鉴别器。在判断某一样本的类别时,只要看它被哪个优先权较高的超球面“包含”即可。而且,超球面的个数就是神经网络分类器第 1 层(隐含层)所需神经元的个数,所有输入数据被第 1 层神经元分成不同的类别,而第 2 层神经元则给出输入数据的特定输出

4 性能分析

对于传统的神经网络而言,并没有一个一般的理论来保证其一定收敛,学习复杂度很高。而且选择恰当的隐层节点数也是一个难题。本文提出的训练算法则有效地克服了以上不足。首先,本文算法是一种构造的方法,因此不存在任何收敛问题,而且学习时间完全由构造超球面所用的时间决定。其次,隐层节点数自动由超球面个数决定,如果得到最少的超球面数,那么神经网络就具有最少的隐层节点数

5 实验及结果

通过两个实验来说明本文提出的分类算法的有效性

5.1 实验 1

双螺旋曲线的识别是一个典型的非线性分类问题^[12],目前已成为人们验证分类算法的重要标准。本次实验数据取自于卡内基梅隆人工智能库,由两个数据集组成,每个集合有 97 个数据点。螺旋形状参数变化如下:

$$\begin{aligned} \text{angle} &= (i * \text{PI}) / (16 * \text{density}), \\ \text{radius} &= \text{maxRadius} * \frac{104 * \text{density} - i}{104 * \text{density}}, \quad (7) \\ x &= \text{radius} * \cos(\text{angle}), \\ y &= \text{radius} * \sin(\text{angle}). \end{aligned}$$

文献[13]试图利用BP算法求解双螺旋问题,但没有成功;文献[14]提出“生成—收缩”法,经过3000次迭代运算,正确率只有89.6%;文献[15]提出“级联”神经网络,使用超过10层神经网络来处理同一问题,但结果也不甚理想

本次实验采用交叉验证法来检验本文所提算法的分类性能。所谓交叉验证法就是从样本数据集中拿走一些样本作为检测样本,剩余的样本作为训练样本。然而,任何交叉验证法对于本文螺旋分类问题都不是很适合。因为为了检测目的而拿走一些样本,螺旋曲线必然会产生许多“裂口”,任何分类器都会很容易作出误判。所以,为减小这些样本对分类结果的影响,本文采用“留一法”即采用循环的方式,每回依次仅拿走一个样本作检测,剩余 $N - 1$ 个作为训练样本 ($N = 194$),实验结果如表 1 所示

表 1 实验结果

| | 训练 样本数 | 平均训练 时间/s | 隐层 节点数 | 识别率/(%) |
|------|-----------|--------------|-----------|---------|
| 本文算法 | 193 | 25.23 | 42 | 100 |

注:所有 Matlab 程序都是在 P4 普通 PC 机上运行,参数

$$a = 0.7, \text{density} = 1, \text{maxRadius} = 6.5$$

5.2 实验 2

实验采用 COIL 图像数据库,该数据库由 100 个物体,且每个物体水平旋转一周采集 72 幅图像所组成,每幅图像的分辨率为 128×128 。图 1 为数据库中部分物体的图像

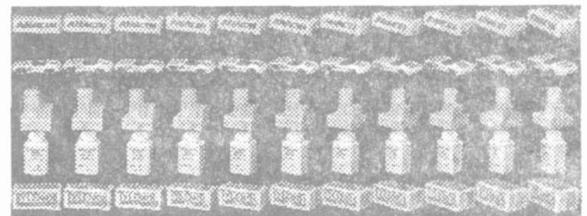


图 1 部分物体图像

任意选取其中的 20 个物体进行实验,选取每个物体相间隔的 36 幅图像作为训练样本,其余的图像作为检测样本。为减少运算量,首先采用主元分析对高维样本进行特征提取,然后利用本文算法进行识别,识别结果如图 2 所示

由图 2 可以看出,随着特征维数的增加,分类错误率迅速下降。当特征维数大于 30 时,错误识别率趋于稳定,且错误率最低可达 1.3889%。在每次实验

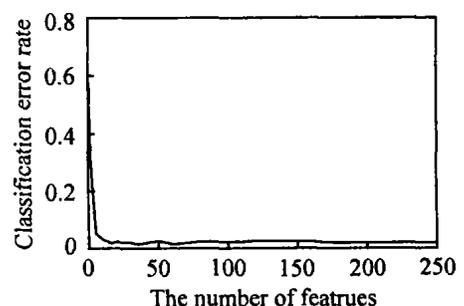


图 2 分类错误率随特征维数增加的变化曲线

中,神经网络的平均训练时间仅为120 s,这在传统神经网络训练中是不可能实现的,而且所构造的神经网络的正确识别率也是可接受的.因此,本文算法同样也适合于高维模式识别问题

6 结 语

本文利用RBF神经元模型的几何解释,提出一种新的基于构造型神经网络分类学习算法,算法几何意义明确,实现容易,不但对于低维的双螺旋曲线的识别率为100%,而且对于高维的图像识别也表现出良好的识别能力,这与传统的神经网络算法相比,具有明显的优势. BP算法不仅不能保证对于训练样本的高识别率,而且当训练到一定程度时,经常会遇到“Over-Fitting”的问题.本文算法之所以表现出良好的分类能力,关键在于它是从数据本身出发去逼近其分布的几何轮廓,而且实验也证明了本文算法的有效性

参考文献(References)

- [1] Mao J, Jain K. Artificial Neural Networks for Feature Extraction and Multivariate Data Projection [J]. *IEEE Trans on Neural Networks*, 1995, 6(2): 296-317.
- [2] Song H H, Lee S W. A Self-organizing Neural Tree for Large-set Pattern Classification [J]. *IEEE Trans on Neural Networks*, 1998, 9(5): 369-380
- [3] Yuan J L, Fine T L. Neural-network Design for Small Training Sets of High Dimension [J]. *IEEE Trans on Neural Networks*, 1998, 9(1): 266-280
- [4] Mukhopadhyay S, Roy A, Kim L S. A Polynomial Time Algorithm for Generating Neural Networks for Pattern Classification: Its Stability Properties and Some Test Results [J]. *Neural Comput*, 1993, 5(2): 317-330
- [5] Chen Q C. Generating-shrinking Algorithm for Learning Arbitrary Classification [J]. *Neural Networks*, 1994, 7(9): 1477-1489
- [6] 王守觉,王柏南.人工神经网络的多维空间几何分析及其理论[J].*电子学报*,2002,30(1): 1-4
- (Wang S J, Wang B N. A Analysis and Theory of High-dimensional Space Geometry for Artificial Neural Networks [J]. *Acta Electronica Sinica*, 2002, 30(1): 1-4)
- [7] Hopfield J J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities [A]. *Proc Natl Acad Sci [C]*. USA, 1982, 79(8): 2554-2558
- [8] Rujan P, Marchand M. A Geometric Approach to Learning in Neural Networks [A]. *Proc IJCNN '89 [C]*. Washington DC, 1989, 2: 105-110
- [9] Ramacher U, Wesseling M. A Geometric Approach to Neural Network Design [A]. *Proc IJCNN '89 [C]*. Washington DC, 1989, 2: 147-1154
- [10] Zhang L, Zhang B. A Geometrical Representation of M-P Neural Model and Its Applications [J]. *J of Software*, 1998, 9(5): 334-338
- [11] 黄国宏,邵惠鹤.一种新的基于神经网络覆盖分类算法[J].*中国图象图形学报*,2004,9(10): 1165-1168
(Huang G H, Shao H H. A New Classification Method Based on Neural Network Covering Algorithm [J]. *J of Image and Graphics*, 2004, 9(10): 1165-1168)
- [12] Singh S. 2D Spiral Pattern Recognition with Possibilistic Measures [J]. *Pattern Recognition Letters*, 1998, 19(2): 141-147.
- [13] Baum E B, Lang K J. Constructing Hidden Units Using Examples and Queries [A]. *Neural Information Processing [C]*. San Mateo: Morgan Kaufmann Publishers, Inc, 1991: 904-910
- [14] Chen Q C, David W Thomas, Mark S Nixon, et al. Generating-shrinking Algorithm for Learning Arbitrary Classification [J]. *Neural Networks*, 1994, 7(9): 1477-1489.
- [15] Fahman S E, Lebiere C. The Cascade-correlation Learning Architecture [J]. *Advances in Neural Information Processing Systems*, 1990, 2: 524-532

(上接第1410页)

- [6] 毕春长,李柠,黄道.钢坯加热炉仿真系统研究[J].*系统仿真学报*,2004,16(4): 663-666.
(Bi C C, Li N, Huang D. Simulation System of Billet Reheating Furnace [J]. *J of System Simulation*, 2004, 16(4): 663-666)
- [7] 青格勒,程素森,杨天钧,等.步进梁式加热炉内的板坯温度场数值模拟[J].*北京科技大学学报*,2004,26(2):

165-168

(Qing G L, Cheng S S, Yang T J, et al. Numerical Simulation for the Temperature Field of Slabs Heated in a Walking Beam Type of Reheating Furnace [J]. *J of University of Science and Technology Beijing*, 2004, 26(2): 165-168)