

基于概率型径向基神经网络(P-RBFN)的人脸识别^{*}

於东军 赵海涛 杨静宇

(南京理工大学计算机系 南京210094)

Probabilistic Radial Basis Function Network Based Face Recognition

YU Dong-Jun ZHAO Hai-Tao YANG Jing-Yu

(Department of Computer Science, Nanjing University of Science and Technology, 210094)

Abstract A probabilistic radial basis function network (P-RBFN) is presented for face recognition. Each P-RBFN is only responsible for the recognition of one class. To a given face pattern to be identified, the probability of the given face pattern belonging to every class is calculated by corresponding P-RBFN and the final recognition result is the fusion of all P-RBFN's outputs. This method combines the statistic theory and neural network technology and can easily be applied in distributed mode. Experiments are implemented on ORL, and an error rate 4% has been got. Comparison between P-RBFN and other methodologies such as Eigenface, SOM+CN and HMM has been done and the advantages of the P-RBFN are demonstrated.

Keywords Face recognition, Radial basis function, Neural network, Pattern recognition, Genetic algorithm

1 概述

近年来,人体生物特征(人脸、指纹、虹膜等)识别技术得到了越来越广泛的研究与应用,其中人脸识别技术尤其是一个热点^[1,2],和其它人体生物特征进行个人身份鉴别的方法相比,人脸识别具有直接、友好、方便的特点。已经有许多学者提出了各种人脸识别技术如:基于人脸几何特征^[3](双眼距离、头宽、眼嘴距离等)的方法,基于 K-L 变换的特征脸^[4~8],奇异值分解^[7]等;神经网络的方法(如卷积神经网络(SOM+CN)^[8]),还有学者提出了隐型马尔可夫模型^[9](HMM)和伪 2D-HMM 等。

由于神经网络具有自学习自适应及鲁棒性强的特点,因此将神经网络技术用于人脸识别将是可行的^[8,10,11]。Steve Lawrence 等使用卷积神经网络(SOM+CN)^[8]进行了人脸识别的研究,得到了较好的结果,在 ORL 上的错误率是 3.8%。卷积神经网络(SOM+CN)方法的弱点之一是集中式计算,不适合于分布式应用,并且当有新的类别加入后,CN 的结构要调整并要重新学习。另外由于此方法的前端使用的是 SOM,在人类别多时学习非常耗时间。在 ORL 图像库上,识别 40 个人,每个人 5 个样本用于学习,使用该方法学习时间约 4 小时。本文旨在提出一种基于概率型径向基神经网络的方法,此方法能够进行分布式计算,并且当有新类别加入时,不会造成原有网络结构的破坏。

2 RBFN 简介

径向基神经网络(RBFN)最初由 Moody 和 Darken 在 1989 年^[12]提出,RBFN 也是一种 3 层前馈网,见图 1。

输入层和隐层,隐层和输出层之间是全互联的,且输入层和隐层所有联接权值为 1,隐层的激励函数使用径向对称的基函数:

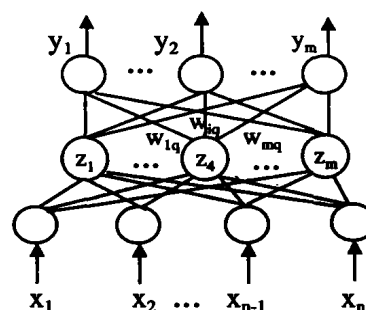


图1 径向基神经网络 RBFN

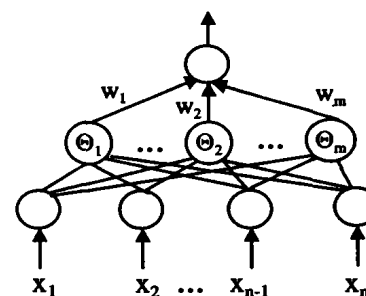


图2 输出为概率密度型的径向基神经网络(P-RBFN)

$$z_q(x) = \lambda_q \exp\left(-\frac{1}{2}(x - \mu_q)^T \sum_q^{-1} (x - \mu_q)\right)$$

其中 λ_q 是第 q 个基函数的系数, $\mu_q = (\mu_{q1}, \mu_{q2}, \dots, \mu_{qn})^T$

$$\sum_q = \begin{bmatrix} \sigma_{q1}^2 & \dots & 0 \\ \vdots & \sigma_{q2}^2 & \vdots \\ 0 & \dots & \sigma_{qn}^2 \end{bmatrix}$$

μ_q 和 $\sum_q$ 称为基函数的中心和宽度参数。这样,RBFN 的最终输出是:

^{*} 本文得到国家自然科学基金资助,编号:60072034。於东军 博士研究生,研究方向为模式识别,神经网络。赵海涛 博士研究生,研究方向为模式识别,图像处理。杨静宇 教授,博士生导师。

$$y_i = \sum_{k=1}^m w_{ik} z_k(x)$$

其中 w_{ik} 是相应的隐节点和输出节点之间的权重。

RBFN 已经在函数逼近、模式识别等领域得到了广泛应用。本文中,我们将使用 RBFN 来估计各类人脸在空间中的概率密度函数,并使用训练样本来优化调整概率密度函数的参数。

3 基于 P-RBFN 的人脸识别

3.1 特征提取

在应用人脸识别系统时,通常分为3个步骤:(1)人脸检测,(2)特征提取,(3)人脸识别。人脸检测就是在给定的图像或是视频序列中定位并分割取得人脸。在得到人脸后就需要进行特征提取,有效的特征提取对于人脸识别系统的成功与否起到很关键的作用。关于特征提取已经有学者提出很多方法,如 K-L 变换、奇异值特征、几何特征、SOM 等等。本文中我们使用一种最简单的方法,即降低人脸的解析度(粗化)。例如原始的 ORL 图像是 92×112 ,我们可以将其粗化为 46×56 或是 23×28 等等。例如,将 92×112 的图像粗化为 46×56 的图像时,就是依次取原图像中 2×2 的子图像的像素平均值作为新图像中对应的一个像素的值。本文中我们将图像粗化至 6×7 的大小。将图像粗化一方面可以降低计算和空间存储代价,另一方面可以提高容错性。在粗化后我们还可以将图像的像素值归一化。这在一定程度上可以降低光照的影响。图3(a)是 ORL 中的一张原始图片,图3(b)、图3(c)和图3(d)分别是粗化后的图像。

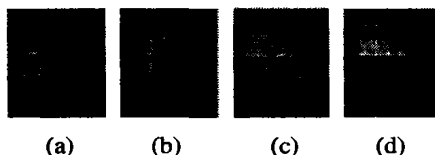


图3 (a)原图 92×112 (b) 23×28 (c) 12×14 (d) 6×7

3.2 PRBFN 的设计

我们假设粗化后得到的 n 维人脸特征 x 在高维空间中是混合正态分布,且各维互不相关,则其概率密度函数可以表示为:

$$p(x|\omega) = \sum_k p(\Theta_k|\omega) p(x|\omega, \Theta_k)$$

其中 $p(x|\omega, \Theta_k)$ 是构成 $p(x|\omega)$ 的正态分布之一,即 Θ_k 是聚类簇之一,且有 $\sum_k p(\Theta_k|\omega) = 1$,

$$p(x|\omega, \Theta_k) = \frac{1}{(2\pi)^{n/2} |\sum_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \sum_k^{-1} (x - \mu_k)\right)$$

其中, $x = (x_1, x_2, \dots, x_n)^T$ 是 n 维的列向量, $\mu_k = (\mu_{k1}, \mu_{k2}, \dots, \mu_{kn})^T$ 是第 k 个正态分布的均值向量, $\sum_k$ 是第 k 个正态分布的 $n \times n$ 维协方差矩阵。前面我们已经假设了各维互不相关,对于正态分布,也就是说各维互相独立,因而 $\sum_k$ 就是对角阵,

$$\sum_k = \begin{bmatrix} \sigma_{k1}^2 & \dots & 0 \\ \vdots & \sigma_{k2}^2 & \vdots \\ 0 & \dots & \sigma_{kn}^2 \end{bmatrix}$$

有了上面的假设,我们可以构造如图2所示的输出为概率密度值的 P-RBFN 如下:

$$\Theta_k = p(x|\omega, \Theta_k)$$

• 44 •

$$= \frac{1}{(2\pi)^{n/2} |\sum_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \sum_k^{-1} (x - \mu_k)\right)$$

其中,基函数 $\Theta_k (1 \leq k \leq m)$, m 表示有 m 个聚类簇。

P-RBFN 的最终输出:

$$y = \frac{\sum_{k=1}^m (\Theta_k \cdot \frac{w_k}{\sum_{i=1}^m w_i})}{\sum_{i=1}^m w_i}$$

$$\text{令 } p(\Theta_k|\omega) = \frac{w_k}{\sum_{i=1}^m w_i}$$

$$\text{显然满足 } \sum_k p(\Theta_k|\omega) = 1$$

所以,上述以基函数 $\Theta_k (1 \leq k \leq m)$ 构造的 RBFN 是以概率密度形式输出的。

P-RBFN 的目的是以各个聚类簇的概率密度函数为基函数来得到整个类的概率密度函数。 $p(\Theta_k|\omega)$ 是聚类簇 Θ_k 出现的概率。

下面我们来考虑一个 c 类 ($\omega_1, \omega_2, \dots, \omega_c$) 的分类问题,决策函数可以是:

$$g_k(x) = \max_i p(\omega_i|x), \text{ 则 } x \in \omega_k.$$

由概率论的知识^[13],当先验概率 $p(\omega_1) = p(\omega_2) = \dots = p(\omega_c)$ 时,上述决策函数等价于:

$$g_k(x) = \max_i p(x|\omega_i), \text{ 则 } x \in \omega_k.$$

也就是说对于一个待识别的模式 x ,在类别的先验概率相等的条件下,我们分别计算它在各个类别上的概率密度函数值,在哪一类上的概率密度函数值最大,就可以判断模式 x 属于哪一类。因此我们使用一个 P-RBFN 只模拟一类人脸的在空间分部的概率密度函数。如果要识别 c 类人,则设计 c 个 P-RBFN。在识别阶段,对于一个待识别的人脸特征 x ,分别计算各个 P-RBFN 的输出,若第 k 个 P-RBFN 输出值最大,则 x 是第 k 类人。见图4。

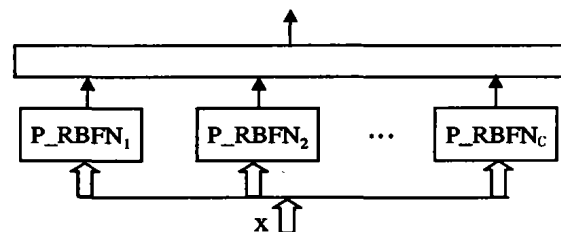


图4 基于 P-RBFN 的人脸识别

使用一个 P-RBFN 只模拟一类人脸的在空间分部的概率密度函数的方法,使得以该模型构建的系统有很强的扩展性。例如要增加识别新的一类人,只要添加一个新的 P-RBFN 负责识别此类人。不再识别某类人时,只要简单地删除该类人对应的 P-RBFN 即可,不会对原来的系统造成破坏。此外,使用此种结构,分布式存储与计算是其与生俱来的性质。

3.3 P-RBFN 的初始参数确定与局部优化

各个 P-RBFN 的基函数个数和初始参数的确定:使用 k 均值方法或是 ISODATA 算法对学习样本进行聚类分析,得到基函数的个数,并根据聚类结果得到各参数的初始值,即各个聚类簇的中心和宽度。使用 k 均值方法的一个缺点就是事先要指定类别个数,当然也可以通过实验的方法来确定 k 的值。ISODATA 算法既能完成聚类分析,还能自动进行类的合并和分裂,从而使得到的类别数较为合理。本文的实验中使用了 k 均值方法来确定基函数个数和初始参数。

P-RBFN 参数的优化:在这里我们使用极大似然估计的

方法先对各个 P-RBFN 进行局部优化。

定义 似然函数:

$$L(\mu, \Sigma) = \prod_{i=1}^N p(x_i | \omega; \mu, \Sigma)$$

其中 N 表示样本个数, x_i 表示第 i 个样本。

两边取对数

$$\ln L(\mu, \Sigma) = \ln \prod_{i=1}^N p(x_i | \omega; \mu, \Sigma) = \sum_{i=1}^N \ln p(x_i | \omega; \mu, \Sigma)$$

将使得 $\ln L(\mu, \Sigma)$ 取得最大值的一组参数 (μ^*, Σ^*) 作为各个 P-RBFN 局部优化的结果。我们使用遗传算法来逼近 (μ^*, Σ^*) , 得到次优解 (μ'', Σ'') 。使用遗传算法时, 就使用似然函数的对数 $\ln L(\mu, \Sigma)$ 作为适应度函数。

3.4 全局学习

在设计并训练好各个 P-RBFN 后就进入了全局学习阶段。全局学习的目的是为了地提高识别的正确率。对于一个给定的学习样本, 如果它应该属于类别 p , 而经过 MAX-Selector 被判别为属于类别 q , 即第 q 个 P-RBFN 的输出大于第 p 个 P-RBFN 的输出, 则第 p 个 P-RBFN 应当进行增强学习, 第 q 个 P-RBFN 进行减弱学习。即

$$\Delta(\mu, \Sigma, w)_p = (\mu, \Sigma, w)_p^{j+1} - (\mu, \Sigma, w)_p^j = \alpha \frac{\partial p(x | \omega_p)}{\partial (\mu, \Sigma, w)_p}$$

$$\Delta(\mu, \Sigma, w)_q = (\mu, \Sigma, w)_q^{j+1} - (\mu, \Sigma, w)_q^j = -\alpha \frac{\partial p(x | \omega_q)}{\partial (\mu, \Sigma, w)_q}$$

其中 α 表示学习效率, j 表示迭代步数。 ω_p, ω_q 分别表示类别 p 和 q 。 $(\mu, \Sigma, w)_p, (\mu, \Sigma, w)_q$ 分别表示网络 p 和 q 的参数。作为例子, 对于某个具体的 P-RBFN, 我们推导一下其参数的学习公式: 因为

$$y = \sum_{i=1}^n (\Theta_i \cdot \frac{w_i}{\sum_{i=1}^n w_i})$$

$$\text{其中 } \Theta_i = p(x | \omega, \Theta_i) = \frac{1}{(2\pi)^{n/2} |\sum_i|^2} \exp(-\frac{1}{2} (x - \mu_i)^T \sum_i^{-1} (x - \mu_i))$$

所以

$$\Delta \mu_{ij} = \alpha \frac{\partial y}{\partial \mu_{ij}} = \alpha \cdot \frac{w_i}{\sum_{i=1}^n w_i} \cdot \frac{\partial \Theta_i}{\partial \mu_{ij}} = \alpha \cdot \frac{w_i}{\sum_{i=1}^n w_i} \cdot \Theta_i \cdot \frac{(x_j - \mu_{ij})}{\sigma_{ij}^2}$$

$$\Delta \sigma_{ij} = \alpha \frac{\partial y}{\partial \sigma_{ij}} = \alpha \cdot \frac{w_i}{\sum_{i=1}^n w_i} \cdot \frac{\partial \Theta_i}{\partial \sigma_{ij}} = \frac{1}{2} \cdot \alpha \cdot \frac{w_i}{\sum_{i=1}^n w_i} \cdot \Theta_i \cdot \frac{(x_j - \mu_{ij})^2}{\sigma_{ij}^3}$$

$$\Delta w_i = \alpha \frac{\partial y}{\partial w_i} = \alpha \cdot [\frac{w_i}{\sum_{i=1}^n w_i} - \sum_{j=1}^n (\Theta_j \cdot \frac{w_j}{(\sum_{i=1}^n w_i)^2})]$$

4 实验结果

在我们的实验中, 使用了 ORL 图像库。在 ORL 图像库中有 40 个人, 每个人有 10 幅图像。这些图像有不同的面部表情(微笑/不笑, 睁眼/闭眼)、面部细节(戴眼镜/不戴眼镜)、比例缩放(不超过 10%)、方向变化(不超过 20 度)。每个人的 10 幅图像, 5 幅作为学习样本, 5 幅作为测试样本。在识别不同类别数时的结果见表 1。

表 1 识别不同类别数时的识别结果

类别数	10	20	30	40
正确率	100%	99.5%	97.4%	96%

使用不同学习/测试比来进行实验, 例如 4/6 表示每个人的 4 幅图像用来学习, 6 幅用来测试。实验结果见表 2。由表 2 我们可以发现, 随着学习/测试比的增大, 识别正确率可以提高。这一点可以做出如下解释: 随着学习样本的增加, 神经网络可以地逼近各类人脸的分布规律, 因而识别正确率提高了。但是当学习/测试比超过一定范围后, 识别率反而下降了, 但是我们可以发现, 导致识别率下降的原因是测试样本的个数减少了, 错误识别的图像个数还是随学习样本增多稳步减少的。

表 3 中我们比较了 P-RBFN 和其它几种方法在 ORL 上的性能。从表 3 我们可以看出, 我们方法的正确识别率已经接近于 SOM+CN 神经网络的方法, 但是在学习时间上明显少于 SOM+CN 方法。

表 2 使用不同学习/测试比时的识别结果, 括号中的数字表示错误识别的图像个数

学习/测试	正确率(错误个数)	
	30	40
4/6	95.5%(10)	94.2%(14)
5/5	97.4%(4)	96%(8)
6/4	98%(4)	97%(5)
7/3	96.7%(3)	96.7%(4)
8/2	96.7%(2)	96.1%(3)

表 3 P-RBFN 方法和其它几种方法在 ORL 上的性能比较

方法	学习时间	正确率
HMM	n/a	87%
特征脸	n/a	90%
Pseudo 2D-HMM	n/a	95%
SOM+CN	4小时	96.2%
P-RBFN	30分钟	96%

结论 本文中我们提出了用概率型的径向基神经网络(P-RBFN)方法来进行人脸识别, 此种方法将统计理论和神经网络技术有机结合起来, 又由于每个 P-RBFN 只负责一类人的识别, 因而可以轻松实现分布式应用。我们使用 ORL 图像库进行了实验, 错误率为 4%, 证明此种方法不失为一种有效的人脸识别方法。

参考文献

- Chellappa R, Wilson C L, Sirohey S. Human and machine recognition of faces: a survey. Proc. IEEE, 1995, 83(5): 705~740
- Rosenfeld A. Survey: Image analysis and computer vision, 1994. Computer Vision and Image Understanding, 1995, 62(1): 90~143
- Harmon L D, et al. Machine identification of human faces. Pattern Recognition, 1981, 13(2): 97~110
- Bischof M, Pentland A. Human face recognition and face image set's topology. CVGIP: Image Understanding, 1994, 59(2): 254~261
- Turk M, Pentland A. Face processing, Models for recognition. Proc. Intelligent Robots and Computer Vision VII, SPIE, 1989, 1, 192: 22~32
- Turk M, Pentland A. Eigenfaces for recognition. J. Cognitive Neuroscience, 1991, 3(1): 71~86

(下转第 123 页)

适应性,额外的存储管理开销可能还会使其性能下降。而且另外一点值得考虑的就是既然该种方案引入了传送给客户的 video 质量的调节,那么它的性能评估参数就不再应该只是传统的字节命中率(byte-hit ratio)了,应该把客户端的播放质量也纳入到考虑因素中统一进行评估。

结论 本文研究了基于 Internet 的视频传送系统中的代理缓存服务器的设计中的若干关键问题。既有对部分缓存策略(partial cache)的详细讨论,又有对针对大体积 video 的特殊置换算法的讨论,还就多个代理服务器群的管理方案进行了一定的研究。最后还针对分层编码的 video 代理缓存方案做了分析。从中可以看到,这些代理缓存的设计分别侧重于某个方面,并没能兼顾其它的考虑,如选择缓存方法就没有考虑到缓存数据的置换问题,基于段的缓存置换算法又是只针对单个代理缓存服务器的,而由多个代理缓存服务器构成代理缓存系统的 MiddleMan 项目又只侧重于系统体系结构的设计,并没有包含针对 video 特别设计的缓存置换算法,最后的针对分层编码的代理缓存系统 Mocha 具有其独到性,同时也有不可避免的局限性。我们认为,基于 Internet 的视频传送系统中的代理缓存应该构成一个具有可扩展性、自适应性的系统,并且针对 video 数据的编码特点、传输特点来设计缓存置换策略,用带宽需求量、客户端播放性能等来作为最终性能测试指标。

参考文献

- 1 Chankhunthod A, et al. A Hierarchical Internet Object Cache. In: Proc. of the 1996 USENIX Technical Conf. Jan. 1996
- 2 <http://squid.nlanr.net/squid>
- 3 <http://netlab1.bu.edu/~staro/546projects/mocha/>
- 4 <http://www.cs.cornell.edu/zeno/Projects/middleman/Default.html>
- 5 Wang Y W, Du D H C. A Network-Conscious Approach to End-to-End Video Delivery over Wide Area Networks Using Proxy Servers. In: Proc. of INFOCOM' 1998
- 6 Balafoutis E, Panagakos A, Laoutaris N, Stavrakakis I. The Impact of Replacement Granularity on Video Caching, IFIP Networking, Pisa, Italy, 2002
- 7 Acharya S, Smith B. An Experiment To Characterize Videos On The World Wide Web. In: Proc. of ACM/SPIE Multimedia Computing and Networking 1998(MMCN'98), San Jose, Jan. 1998
- 8 Salehi J, Zhang Z L, Kurose J, Towsley D. Supporting Stored Video: Reducing Rate Variability and end-to-end Resource Requirements Through Optimal Smoothing. In: ACM Intl. Conf. on Measurement and Modeling of Computer Systems (ACM SIGMETRICS) (Philadelphia, PA, May 1996)
- 9 Miao Z, Ortega A. Proxy Caching for Efficient Video Services Over the Internet. In: 9th Intl. Packet Video Workshop(PVW'99), New York, April 1999
- 10 Miao Z, Ortega A. Scalable Proxy Caching of Video under Storage Constraints. IEEE Journal on Selected Areas in Communications, (Internet Proxy Services) 2002
- 11 Wang J. A Survey of Web Caching Schemes for the Internet. ACM Computer Communication Reviews, 1999, 29(5)
- 12 Aggarwal C, Wolf J, Yu P. Caching on the World Wide Web. IEEE transactions on knowledge and data engineering, 1999, 11(1)
- 13 William S, Abrams M, Standridge C R, Abdulla G, Fox E A. Removal Policies in Network Caches for World Wide Web Documents. In: Proc. of Sigcomm'96
- 14 Abrams M, et al. Caching proxies: Limitations and Potentials. In: Proc. of 4th Intl. WWW Conf. Boston, MA, Dec. 1995
- 15 Wu K, Yu P, Wolf J. Segment-based Proxy Caching of Multimedia Streams. In: Proc. of 10th Intl. Conf. on World Wide Web, Hong Kong, 2001
- 16 Acharya S, Smith B. MiddleMan: A Video Caching Proxy Server. In: Proc. of 10th Intl. Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV), Chapel Hill, NC, June 2000
- 17 Wu K L, Yu P S. Latency-Sensitive Hashing for Collaborative Web Caching, Computer Networks: The International Journal of Computer and Telecommunications Networking, June 2000
- 18 Wu K L, Yu P S. Replication for Load Balancing and Hot-Spot Relief on Proxy Web Caches with Hash Routing, to be appear in Distributed and Parallel Databases, Jan. 2003
- 19 Wu K L, Yu P S. Controlled Replication for Hash Routing-based Web Caching, to be appear in International Journal of Computer Science and Engineering, Sept. 2002
- 20 Rejaie R, Kangasharju J. Mocha: A Quality Adaptive Multimedia Proxy Cache for Internet Streaming. In: Proc. of NOSSDAV
- 21 Cho J, Shin H. Design Issues for Multimedia Information Servers in Mobile Computing Environment. In: Proc. of the 4th Intl. Workshop on Mobile Multimedia Communications, Sep. 1997
- 7 洪子泉, 杨静宇. 基于奇异值特征和统计模型的人像识别算法. 计算机研究与发展, 1994, 31(3): 60~65
- 8 Lawrence S, Giles C L, Tsoi A C, Back A D. Face recognition: a convolutional neural-network approach. IEEE Trans. Neural Network, 1997, 8(1): 98~113
- 9 Yoon K S, et al. Hybrid approaches to frontal view face recognition using the hidden Markov model and neural network. Pattern Recognition, 1998, 31(3): 283~293
- 10 Schofield A J, et al. A system for counting people in video images using neural networks to identify the background scene. Pattern Recognition, 1996, 29(8): 1425~1428
- 11 Intrator N, et al. Face recognition using a hybrid supervised/unsupervised neural network. Pattern Recognition Letters, 1996, 17(1): 67~76
- 12 Lin C T. Neural Fuzzy System. Prentice-Hall Press, U.S.A., 1997. 328~330
- 13 Fukunaga K. Statistical pattern recognition. New York: Academic, 1989. 4
- 14 於东军, 王士同, 吴小俊. 层次径向基神经网络的全局逼近理论. 计算机研究与发展, 1999, 36(11): 1329~1334

(上接第45页)