前馈优先排序神经网络的构造算法及分析

朱世交[1,2]　杨　磊[1,2]　王守觉[2,3]

(同济大学计算机科学与工程系　上海 200092)[1]　(同济大学半导体与信息技术研究所　上海 200092)[2]

(中国科学院半导体研究所神经网络实验室　北京 100083)[3]

摘　要　本文从高维空间特征点覆盖的角度，讨论了优先度排序神经网络(PONN)原理与构造算法，提出了各向同性的随机分割算法(RPA)和中心选择算法(CSA)，最后给出标准测试集的分类测试结果，并对其进行了分析与讨论，实验结果表明 PONN 构造方法优于传统前馈神经网络。

关键词　优先排序网络，模式识别，拓扑空间

Algorithms and Analysis of Construction for Feed-Foreword PONN

ZHU Shi-Jiao[1,2]　YANG Lei[1,2]　WANG Shou-Jue[2,3]

(Department of Computer Science and Engineering, Tongji University, Shanghai 200092)[1]　(Institute of Semiconductors and Information Technology, Tongji University, Shanghai 200092)[2]　(Lab of Artificial Neural Networks, Institute of Semiconductors, CAS, Beijing 100083)[3]

Abstract　From the view-point of coverage in the high-dimension space (HDS), we discuss the theory and analysis of architecture for PONN, and promote RPA and CSA based on isotropy in HDS. Benchmark testing sets are launched in the pager. From the result of experiments, the method of PONN is prior to classic ANN.

Keywords　PONN, Pattern recognition, Topology space

1　引言

在众多类型的神经网络中，前馈网络最受人们的青睐[1]，它已被广泛应用于模式分类、系统识别等领域。但是对于大量的训练样本，传统前馈网络往往需要大量训练时间；此外，对于已训练好的网络，训练过程中的先后学习顺序往往无法得到体现。虽然 SVM[2] 当今得到了广泛的应用，但是其最优划分在大量样本情况下不能在多项式时间范围内得到，属于 NP 难问题[3]。传统神经网络的这些缺陷大大限制了神经网络的应用范围。

传统模式识别只注意“区别”，而没重视事物的“认识”，王守觉院士提出的人工神经网络高维空间几何分析方法[4,5]，把模式识别问题抽象成为一个高维空间的类别划分问题，把特征空间的分布理解为空间形体学习过程，在此基础上提出了与人类认知学习概念模式类似的优先度排序神经网络的概念，这种方式更接近于人类的认识。

本文结合仿生识别中的高维空间覆盖原理，介绍了具有层次结构的优先排序覆盖神经元网络(PONN)，提出了几种网络构造算法，并且通过分类实验对算法进行分析与评估。

2　优先排序神经网络构造的基本理论

2.1　网络基本模型的定义

神经网络对于类别划分问题，本质上讲就是寻找满足条件的未知函数 $f: R^n \to R$ 映射。传统网络没有表示知识层次划分的能力，因此往往不符合人类认知能力的表示。为了下面阐述网络模型的方便，将基本符号定义说明如下：

- R^n：表示 n 维特征欧氏空间；
- S_i：表示 n 维空间中的第 i 个特征类子空间，其中 $S_i \subset R^n, i \in \{1,2,\cdots,m\}$；
- $\{\psi_i^k\}$：k 层第 i 类别的神经元集合，$k \in I, i \in \{1,\cdots,m\}$；单个神经元 $\psi_{i,j}^k$，表示 i 类别的 k 层第 j 个神经元；
- L_k：表示由 k 层神经元集合，定义 $L_k := \{\{\psi_i^k\} | k \in I, i \in \{1,\cdots,m\}\}$；
- O_k：神经元 k 层的输出序列，定义 $O_k := \{0,\cdots,1_q^p,0,0 | q \in \{1,\cdots,m\}, p=0 \text{ 或 } 1\}$；
- Y_i：类别 I 的判别输出。

优先排序网络的基本思想是空间类别的划分具有层次特征(基本符合人类认知过程)，大类别识别由小类逐层判别构成。

概念体现到网络层结构上，如图 1 所示。

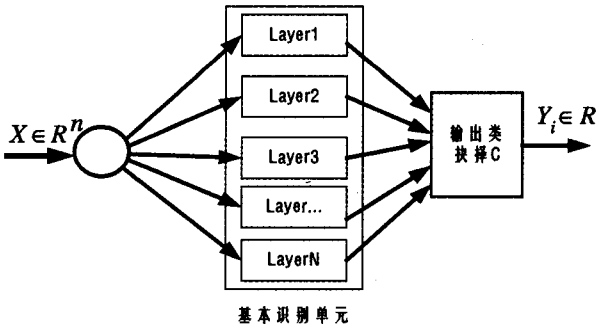

图 1　PONN 网络层次构造单元

网络结构基本单元由各个不同层次的中间层神经网络单元构成，通过不同层次的数据分类，形成一个分类判别的塔式

朱世交　博士研究生，研究方向：优先排序神经网络的理论及其在仿生识别中的应用；杨　磊　硕士研究生，研究方向：模式识别，图像处理；王守觉　中国科学院院士，博士生导师，研究方向：神经网络。

结构网络。数据类别划分依赖于数据本身，而非人为加入先验知识；输出类别判别依赖于前端不同层次神经元并行计算的激活状态。

2.2 网络构造基本原理

对于神经元 $\psi_{i,j}^k$，规定单个神经元 $\psi_{i,j}^k$ 在高维特征空间中是有限空间区域覆盖，同时为了分析方便，认为在高维空间中其覆盖区域具有各向同性性，近邻覆盖空间采用欧式度量空间，通过空间特征点与空间支撑点距离的比较进行测度，公式描述如下。

$$\rho(\psi_{i,j}^k)=\sqrt{\sum_{l=1}^{n}(x_l-x_{l_0})}$$

当 $\rho(\psi_{i,j}^k)$ 测度小于阈值 Γ_0，$\Gamma_0\in R$（阈值由支撑点与异类样本最近距离确定），则认为在区域覆盖范围内；反之，则表示在覆盖区域之外，即不被此神经元 $\psi_{i,j}^k$ 覆盖。形式化描述如下。

$$F(\rho)=\begin{cases}1,\rho<\Gamma_0\\0,\text{其它}\end{cases}$$

对于通过样本构造的神经元 $\psi_{i,j}^k$，同层神经元空间覆盖定义为 $C(\psi_i^k)$，不同神经元的覆盖区域不重叠，$C(\psi_i^{k1})\cap C(\psi_i^{k2})=\phi$，$k1\neq k2$，同类间神经元以及不同层、不同类之间的神经元有可能出现空间重叠现象，此过程类似于一物体在空间区域的递变过程，如看待一个人由远及近，在每个层次上与之类似的物体也不尽相同。另外一个特点是空间神经元覆盖的各向同性，一方面减少高维空间数据处理的复杂性；同时，使用简单模型将使分析更具有普遍意义。

网络构造基本原理。网络训练依据训练样本本身的特征属性，起始阶段网络内部的层次为空，即 $k=0$（不同于传统网络需预先确定神经元数目）；随着训练样本进入网络，网络层 k 逐渐增大，神经元数不断增多，构成层内神经元 $\{\psi_i^k\}$ 集合，要求同层神经元在高维空间覆盖区域不可重叠。

网络识别基本原理。网络接受待判别数据输入 $x\in R^n$，各层神经元输出响应序列 O_k，定义判别：

$$Y_i:=\{i\mid i=\{0\text{ 或 }q:=\{\min(k)\&\{O_k>1\rightarrow p\}\}\}\}$$

通过 Y_i 对待识别样本 $x\in R^n$ 进行类别划分，

$$Y_i=\begin{cases}\text{类别 } i,i>0\\\text{拒绝},i=0\end{cases}$$

从网络对样本类别的划分来看，PONN 很好地描述了异类样本的拒识情况，优于传统网络对不“认识”样本情况下的乱识问题。

3 神经网络构造算法

本部分针对 PONN 网络，提出了三种网络构造算法，方法主要侧重于网络的训练过程。

3.1 随机分割算法（Random Partitioning Algorithm, RPA）

此构造算法为分别随机选取不同测试样本类中的测试点作为支撑点，通过各向同性神经元进行同类覆盖（不包括异类样本），每类构造一次，神经元单元层次数递增一个单位。

Random Partitioning Algorithm (RPA method)

1）定义起始 $i=1,i\in\{1,2,\cdots,m\},k=1$；

2）从当前层次的样本队列中，选取测试样本点 $x\in S_i$，$j=0$；

3）以 $x\in S_i$ 为支撑点，使用 $\psi_{i,j}^k$ 神经元覆盖 S_i，覆盖得到的 S_i 的子集合为 W，$W\subseteq S_t$。$S_t=S_t\backslash W$，如果 $S_i=\phi$，则从样本队列中去除 S_i 类别，否则，S_i 继续进入下一层样本队列；

4）判别当前层次样本队列是否为空，如果为空，则继续步骤 5；否则 $i=i+1$，继续步骤 2；

5）判别下一层次样本队列是否为空，是则构造结束；否则 $k=k+1$，转步骤 2。通常，对于 RPA 算法而言，随机样本的选择不利用任何先验知识；但是由于 RPA 没有利用样本空间本身的先验知识，有可能对紧致型类别会使用较多的神经元个数。

3.2 中心选择算法（Center Selection Algorithm, CSA）

为了克服 RPA 算法没有使用样本空间先验知识的缺陷，CSA 算法基本思想就是同类样本空间在高维空间具有相应程度的密集性和连续性，利用样本本身的特征，神经元在构造时选择样本类中的中心点作为支撑点进行构造。

Center Selection Algorithm (CSA method)

1）定义起始 $i=1,i\in\{1,2,\cdots,m\},k=1$；

2）从当前层次的样本队列中，选取测试样本空间中心点 $x_0=\frac{1}{card(S_i)}\sum_{l}^{card(S_i)}x_l,x_l\in S_i,j=0$；

3）以 x_0 为支撑点，使用 $\psi_{i,j}^k$ 神经元覆盖 S_i，覆盖得到的 S_i 的子集合为 W，$W\subseteq S_t$。$S_t=S_t\backslash W$，如果 $S_i=\phi$，从样本队列中去除 S_i 类别，否则，S_i 继续进入下一层样本队列；

4）判别当前层次样本队列是否为空，如果为空，则继续步骤 5；否则 $i=i+1$，继续步骤 2；

5）判别下一层次的样本队列是否为空，是，则结束；否则 $k=k+1$，转步骤 2。

CSA 很好地利用了同类样本本身在高维特征空间中的点集聚集分布特征，可以使用相对于 RPA 算法较少神经元覆盖特征类空间；对于样本区别很大，中心偏移不重叠的情况效果确实比较好，但是由于 x_0 为虚拟支撑点，在 x_0 空间点周围可能全是异类样本，也就是说在这种情况下使用虚拟支撑点 x_0 无法覆盖同类样本，算法出现振荡，变得不收敛，因此有下述 CASA 算法。

3.3 中心适配选择算法（Center Adaptive Selection Algorithm, CASA）

对于 CASA 算法，把支撑选择点 x_0 做适当调整。调整的基本思路就是当 x_0 不能做同类样本覆盖时，选取同类中与 x_0 偏移最大的点作为支撑点，由 x_0 的计算公式，偏移最大的点对 x_0 的分量贡献也最大，可以很好地打破 x_0 不能划分同类样本的这种振荡平衡状况。

Center Adaptive Selection Algorithm (CASA method)

1）定义起始 $i=1,i\in\{1,2,\cdots,m\},k=1$；

2）从当前层次的样本队列中，选取测试样本空间中心点 $x_0=\frac{1}{card(S_i)}\sum_{l}^{card(S_i)}x_l,x_l\in S_i,j=0$；

3）以 x_0 为支撑点，使用 $\psi_{i,j}^k$ 神经元覆盖 S_i，覆盖得到的 S_i 的子集合为 W，$W\subseteq S_t$；

4）若 $W=\phi$，则选取 $x_0=\{x\mid\max(\rho(x_0,x)),x\in S_i\}$，继续步骤 3（此时同类样本至少有一个点 x_0）；

5）选取 $S_t=S_t\backslash W$，如果 $S_i=\phi$，从样本队列中去除 S_i 类别，否则，S_i 继续进入下一层样本队列；

6）判别当前层次样本队列是否为空，如果为空，则继续步骤 7；否则 $i=i+1$，继续步骤 2；

7）判别下一层次的样本队列是否为空，是，则结束；否则 $k=k+1$，继续步骤 2。CASA 通过适当选择各向同性支撑中

心点，克服了 CSA 构造网络时不收敛的缺陷。

4 试验结果与讨论分析

实验中选择了前馈 BP 网络难以解决的阿基米德螺旋线、iris 数据库、字符识别、图像分割识别，这四种实例对 PONN 进行测试，由于 CSA 算法不稳定，实验只比较 RPA 和 CASA 算法。

· 螺旋线。

螺旋线极坐标公式描述：$R=a\theta$（其中 R 为半径，θ 为绕轴旋转的角，a 为常数）。试验中三条螺旋线定义如下，x-y 平面上的图形如图 2 所示。

$$\begin{cases} R_1=0.4*\theta, \theta\in[-0.4\pi, 4\pi] \\ R_2=0.4*(\theta+\pi), \theta\in[-0.5\pi, 5\pi] \\ R_3=0.4*(\theta+1.7*\pi), \theta\in[-0.6\pi, 6\pi] \end{cases}$$

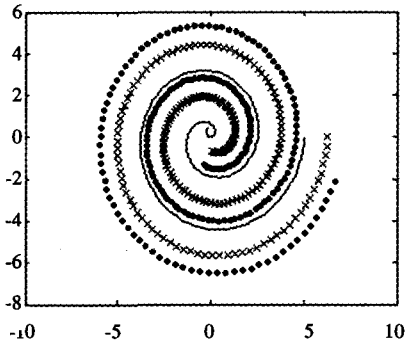

图 2 X-Y 平面三条螺旋线图像

为了观测不同训练样本对网络构造的影响，训练时每类分别取 150～350 个数范围，间隔步长为 5 的训练样本点进行网络覆盖，测试点每类取 1000 个样本点。

· Iris 三分类问题。三类植物长度特征，其中包括萼片长度，宽度，花瓣长度，宽度。共计 150 个样本点，取 2/3 用于训练，剩下 1/3 用于测试。

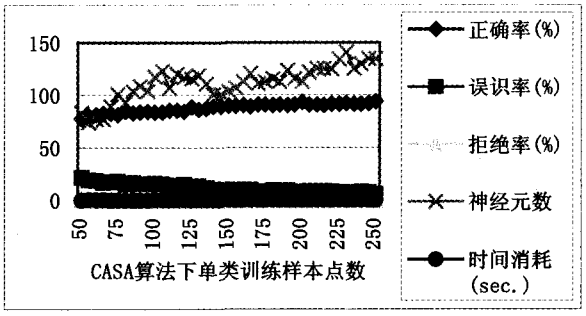

图 3 CASA 算法下三条螺旋线的识别变化

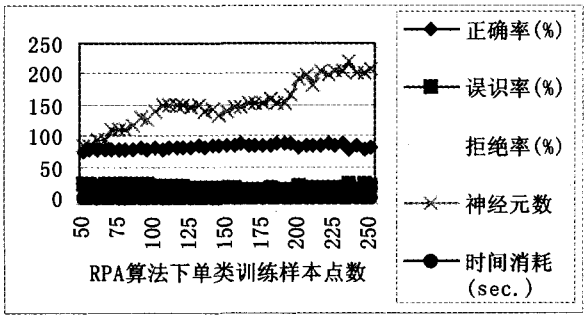

图 4 RPA 算法下三条螺旋线的识别变化

· 图形字符识别。包括 a～z 26 个字母的 20000 个字符实例，每个字符有 20 个不同的字体形状，每个字符特征包括：字符图形起始坐标，长宽等 16 个坐标特征属性（主要为统计边缘、特征），取 26 个字母共计 16000 个点进行训练，把剩余 4000 个样本通过训练后网络进行分类。

· 图形分割。选取户外七种类别图像：砖块、天空、植物、水泥、窗、路、草，图像大小 3 * 3，取颜色、纹理等 19 个特征。每类取 30 个共计 210 个作为训练集，测试集取每类 300 个共计 2100 个样本进行测试。

图 3、4 描述了 RPA 、CASA 算法下 PONN 神经网络的构造和识别情况，表明了网络在不同训练样本点情况下使用神经元数的变化情况，其中 CASA 算法使用的神经元数比 RPA 算法的神经元数量要少，同时其交替变化程度也比较剧烈一些。另外，误识率的变化情况，CASA 算法随着训练样本的不同，误识率变化趋势也比较明显。

表 1 不同测试集下 RPA 算法与 CASA 算法的比较

		RPA 算法	CASA 算法
Iris 数据库	正确率(%)	80.6373	93.9543
	误识率(%)	19.3627	6.0457
	拒绝率(%)	0.000	0.000
	神经元数	17	9
	时间(秒)	0.018	0.010
字符识别	正确率(%)	73.743	77.069
	误识率(%)	25.106	21.730
	拒绝率(%)	1.150	1.200
	神经元数	2271	2448
	时间(秒)	268.802	41.39
图像分割	正确率(%)	70.143	78.095
	误识率(%)	29.619	21.048
	拒绝率(%)	0.238	0.857
	神经元数	59	62
	时间(秒)	0.429	0.491

表 1 中描述了标准测试数据库的测试情况，值得说明的是其中测试结果是通过多次不同条件测试得到的平均结果，主要是为了对比两种不同算法对网络构造和识别的效果。从表 1 的数据来看，CASA 算法的识别率一直比 RPA 算法的识别率要高，从使用神经元覆盖数量以及网络花费的时间来看，除了 Iris 数据库之外，其它两个 CASA 算法使用的神经元数和时间代价都比 RPA 算法要多，其主要原因与训练样本本身特征有关，这些训练样本特征类数据点变化不连续，同时由于其特征在高维空间的表现不是同胚，例如，字符识别中的同字、不同体，而且为了说明方便，算法本身也没有对事物特征本身进行空间特征分析。所以，虽然 CASA 算法的识别率比 RPA 的要高，但在样本特征离散化程度较大的情况下，造成了 CASA 算法在神经元数量和时间花费上比 RPA 算法差一些。值得一提的是，该字符识别集在国际上公布的最好识别率也只是接近 80%[6]，可以看出 CASA 算法构造 PONN 网络的潜力。

小结 本文研究了 PONN 网络的构造原理，并且提出了 RPA、CSA、CASA 算法，对算法进行了详细的描述、分析。对几类分类问题进行了 RPA 与 CASA 算法的比较。通过实验可以看出：

· CASA 算法比 RPA 识别率要高；

· 结合待识别事物特征的高维空间特征分布，有利用于

基于样本分布不平衡的近似支持向量机*)

陶晓燕[1,2] 姬红兵[1] 马志强[2]

(西安电子科技大学电子工程学院 西安 710071)[1] (空军工程大学电讯工程学院 西安 710077)[2]

摘 要 针对标准的近似支持向量机(PSVM)没有考虑样本分布不平衡的问题,提出了一种新的PSVM算法—BPSVM;根据训练样本数量的不平衡对正负样本集分别分配了不同的惩罚因子,并将原始优化问题中的惩罚因子由数值变更为一个对角阵,最后推导出了线性和非线性BPSVM的决策函数。实验结果表明,BPSVM的性能优于PSVM,与SVM方法相比效率更高。

关键词 近似支持向量机,非平衡分布,平衡近似支持向量机

Proximal Support Vector Machines for Samples with Unbalanced Distribution

TAO Xiao-Yan[1,2] JI Hong-Bing[1] MA Zhi-Qiang[2]

(School of Electronic Engineering, Xidian University, Xi'an 710071)[1]

(The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077)[2]

Abstract For the problem of unbalanced data classification is not discussed in the standard Proximal Support Vector Machines, a new PSVM algorithm is presented, namely BPSVM. The different penalty factors are assigned to the positive and negative training sets according to the unbalanced population, and the penalty values are transformed into a diagonal matrix. Finally the decision functions for the linear and nonlinear PSVM are obtained. The experimental results show that the generalization of BPSVM could be better than PSVM, and comparable to SVM with higher efficiency.

Keywords Proximal SVM (PSVM), Unbalanced distribution, Balanced PSVM (BPSVM)

1 引言

支持向量机[1](Support Vector Machines, SVM)是基于统计学习理论中的结构风险最小原则提出来的,它利用再生核以及内积定理,有效地避免了经典学习中的过学习、维数灾难、局部极小等非常棘手的问题,在小样本条件下,仍然具有良好的泛化能力。SVM已成为模式识别、函数回归等领域一个最有力的处理方法[2,3]。

然而,标准SVM需要求解复杂的二次规划问题,训练时间较长。近年来,人们针对SVM提出了许多改进方法[4]。Platt[5]提出了SMO(Sequential Minimal Optimization)算法,其基本思想是循环迭代,将原来问题分解为一系列的子问题,按照某种策略,通过反复求解子问题,最终使结果收敛到原问题的最优解。SMO解决了占用内存的缺点,但这种方法是一种盲目的"爬山法",需要有好的启发式策略。为了克服这个难题,很多学者提出了多变量迭代更新方法,它不像SMO一次迭代只更新工作集的系数,而是更新所有训练集的系数,这种方法原理简单,容易实现,但更新率的值难以确定。最近,为了加速SVM的学习速度,一些学者提出了简化标准SVM优化问题的思路,出现了很多变异SVM算法。如Suykens[6]等人提出的LS-SVM(Least Square Support Vector Machines),Mangasarian[7]等人提出的广义支持向量机(Generalized Support Vector Machines)。这些方法应用都是直接求解非奇异的线性方程组,从而大幅度提高了学习速度。尤其是Fung与Mangasarian[8,9]提出的近似支持向量机(Proximal Support Vector Machines),更引起了人们的极大兴趣。但是,上述PSVM对正负样本误差的惩罚是一样的,然而实际情况下,训练样本的分布往往是不平衡的,取同样的误差惩罚就会影响分类结果[10]。针对标准的SVM和v-SVM的样本不平衡问题,H. G. Chew[11,12]等人提出了根据两类样本的数量,分别对其分配不同的惩罚因子,但优化问题的求解没有变。笔者采用文[11]中的思想定义惩罚因子,针对PSVM的特点,在优化问题中引入了一个新的对角阵,进而推导出了适用于非平衡分布数据分类的线性和非线性PSVM的决策函

*)基金项目:国防预研基金项目资助(No. 51407030103DZ0117)。陶晓燕 博士研究生,主要研究方向为机器学习,智能信息处理。

提高网络的构造性能。

PONN网络不同于传统网络的类划分方法,而由高维空间覆盖角度来分析和构造网络,也就克服了传统网络例如BP网络,难于训练、中间层神经元难于确定、知识表示缺陷等问题。当然,如何更好地构造PONN网络,还有待于利用待识别事物本身特征的空间分布特征作为先验知识,同时加入知识表示的各向异性特征,这一方面的问题值得下一步进行研究。

参 考 文 献

1 曹安照,田丽,陈俊,吕元峰. 神经网络及其研究展望. 自动化与仪器仪表, 2006, 1:1~3

2 Burges C J C. A tutorial on support vector machines for pattern recognition [J]. Data Mining and Knowledge Discvery, 1998, 2 (2):121~167

3 Blum A, Rivest R L. Training a 3-node neural network is NP-complete. Neural Networks, 1992,5:117~127

4 王守觉..仿生模式识别(拓扑模式识别)——一种模式识别新模型的理论与应用[J].电子学报,2002,30(10):1417~1420

5 Wang Showjue. Priority Ordered Neural Networks with Better Similarity to HumanKnowledge Representation [J]. Chinese Journal of Electronics,1999,8(1):1~4

6 Frey P W, Slate D J. Letter Recognition Using Holland-Style Adaptive Classifiers [J]. Machine Learning, 1991,6(2):161~182