

随机投影下的 Plane-Gaussian 人工神经网络

冯哲 杨绪兵 张福全

(南京林业大学信息科学技术学院, 南京, 210037)

摘要: 针对平面高斯神经(Plane-Gaussian, PG)网络采用 k-平面聚类算法得到网络参数,使得网络训练时间过长,且易陷入局部极小值的问题,借鉴极限学习机(Extreme learning machine, ELM)中网络参数随机选择的方式,提出了随机投影下的平面高斯神经网络(Plane-Gaussian network based on random projection, RandPG)。该网络采用随机投影的方式确定隐层激活函数的参数,然后利用 Moore-Penrose 广义逆求解输出层权值。理论上证明该网络具有全局逼近性。同时,对呈直线型和平面型的人工数据集以及 UCI 标准数据库中的分类数据集进行测试,结果表明, RandPG 网络提供了一种简便的参数学习方法,并且在继承了 PG 网络对呈子空间分布的数据分类具有优势的情况下,显著提高了网络的学习速度。

关键词: 人工神经网络; 随机投影; 平面; 高斯; 极限学习机

中图分类号: TP391 **文献标志码:** A

Plane-Gaussian Artificial Neural Network Based on Random Projection

Feng Zhe, Yang Xubing, Zhang Fuquan

(College of Information Science and Technology, Nanjing Forestry University, Nanjing, 210037, China)

Abstract: For the Plane-Gaussian (PG) artificial network, its network parameters are generated from k-plane clustering algorithm in training phase. Compared with random parameters of extreme learning machine (ELM), PG is a time-consumer and easy to trap into local optimal solution. To improve the performance of PG network, inspired by ELM in this paper, a new training method based on random projection for PG network, termed as RandPG, is proposed. Typically, for the three-layer network, the weight matrix between input and hidden layers is selected by random projection to speed training network, and the weight matrix between hidden and output layers is obtained by Moore-Penrose generalized inverse. It is proved that the network has global approximation theoretically. Meanwhile, the effectiveness of this network is tested on the line-distribute datasets, plane-distribute datasets and several UCI datasets. The results indicate that RandPG provides a simple and convenient way to train parameters of neural network, and it not only follows the advantage of PG network, which is more suitable for classifying subspace-distribute datasets, but also significantly accelerates its learning speed.

Key words: artificial neural network; random projection; plane; Gaussian; extreme learning machine

引言

近十几年来,针对神经网络的学术研究非常活跃,研究者们提出了上百种的神经网络模型,其中应用最广泛的神经网络模型有反向传播(Back propagation, BP)网络和径向基函数(Radial basis function, RBF)网络。但是 BP 网络存在计算量大、学习速度慢和易陷入局部极小值等问题;而 RBF 网络的精度和稳定性由隐含层核函数的中心与宽度决定,导致网络的结构难以确定且适应性较差。

针对以上问题,黄广斌等^[1]在单隐层前馈神经网络的基础上提出了一种高效的学习算法——极限学习机(Extreme learning machine, ELM)。与传统的神经网络不同,ELM 网络的输入权重和隐层偏置随机生成,无需进行反复调整,其输出权值用最小二乘估计法就能得到唯一的最优解。ELM 具有较快的学习速度和良好的泛化性能,且网络结构简单。越来越多的学者对其在特征学习、聚类、回归和分类问题上进行研究,现今 ELM 已经在模式识别^[2-3]、自动控制^[4-5]、预测^[6-7]和建模^[8-9]等领域得到了广泛的应用。但是 ELM 并不具备明确的几何意义。杨绪兵等^[10]从平面拟合的角度,采用“平面原型”^[11]代替 RBF 网络的“点原型”,设计出适合平面型分布数据的平面高斯(Plane-Gaussian, PG)神经网络模型。由于 PG 网络在训练过程中采用 k-平面聚类算法(k-plane clustering, kPC)完成样本的类别划分,这就不可避免地会造成网络训练时间过长,且易陷入局部极小值等问题。

为了克服这些缺点,本文借鉴 ELM 随机产生参数的思想,结合随机投影理论,对 PG 网络进行改进,提出了随机投影下的平面高斯神经网络(Plane-Gaussian network based on random projection, RandPG)。RandPG 网络中的分类方向随机选择,避免聚类,突破了陷入局部最优解的限制,网络训练时间大大降低。同时随机投影将高维数据映射到低维空间中,降低了运算的复杂性。RandPG 网络在保持对呈子空间分布数据分类更有效的优势的同时,学习速度得到了进一步提高。

1 平面高斯网络

PG 网络是以平面高斯为激活函数,即用超平面代替高斯函数中的数据中心点,通过最小化样本点到最近超平面距离之和来获得 k 个原型超平面,PG 函数定义为

$$\Phi^{\text{PG}}(\mathbf{x}) = \exp\left(-\frac{|\mathbf{w}^{\text{T}} \cdot \mathbf{x} - b|^2}{2\sigma^2}\right) \quad (1)$$

式中: \mathbf{w} ($\|\mathbf{w}\|=1$) 和 b 分别为超平面 $\mathbf{w}^{\text{T}} \cdot \mathbf{x} - b = 0$ 的法向量和阈值; $|\mathbf{w}^{\text{T}} \cdot \mathbf{x} - b|$ 表示样本点 \mathbf{x} 到超平面的距离; σ 为宽展度。

PG 函数的可视化如图 1 所示,其呈带状分布,兼具全局性和局部性。PG 网络中激活函数的参数 \mathbf{w} 和 b 可通过 kPC 算法^[12]获得,它是超平面为聚类原型,设法找到一组超平面 $P = \{p_1, p_2, \dots, p_k\}$,使得 n 个样本点到距其最近的超平面的平方距离之和最小,其目标函数定义为

$$J_{\text{kPC}} = \sum_{i=1}^k \sum_{j=1}^n \|\mathbf{w}_i^{\text{T}} \cdot \mathbf{x}_j - b_i\|^2 \quad (2)$$

式中 \mathbf{w}_i 和 b_i 分别表示超平面 p_i 的单位法向量和截距。

kPC 算法的执行过程包含反复将样本点分配给距其最近的超平面所对应的聚类和计算新的超平面的法向量和截距两步,直至样本归类不再变化为止。

kPC 算法能对未知样本集进行有效聚类,但是通过迭代方法求解,会导致网络训练速度过慢,且很容易产生局部极小值。

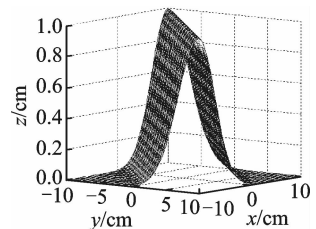


图 1 平面高斯函数的三维可视化

Fig. 1 3D visualization of PG function

2 随机投影原理

随机投影(Random projection, RP)技术^[13]是指将高维空间中的点投影到一个随机选取的低维子空间内,它可将原来的高维问题转化到低维空间来求解。下面给出随机投影的定义。

原始的 d 维数据矩阵 $\mathbf{X}_{d \times n}$, 在 $k \times d$ 维的随机矩阵 \mathbf{R} 的作用下, 被投影到一个 k 维的子空间中, 其中矩阵 \mathbf{R} 的每一行都单位化。投影问题可以描述为

$$\mathbf{X}_{k \times n}^{\text{RP}} = \mathbf{R}_{k \times d} \cdot \mathbf{X}_{d \times n} \quad (3)$$

式中 $\mathbf{X}_{k \times n}^{\text{RP}}$ 是降维后的 k 维数据矩阵。

随机投影可进行数据降维的理论依据主要是 Johnson-Lindenstrauss 定理^[14], 该定理表明: 一个高维欧氏空间可以被映射到一个 $O(\log \frac{M}{\epsilon})$ 维子空间上, 使得数据点之间的距离对于任意的 $0 < \epsilon < 1$ 能近似保留(即偏差不超过 $(1 \pm \epsilon)$ 的因素), 而且这个映射可以在多项式时间内找到。投影矩阵 \mathbf{R} 的选取至关重要。本文选取一个 $k \times d$ 维的随机矩阵 \mathbf{R} , 它的每一个元素 R_{ij} 都独立且满足 $0 \sim 1$ 之间的均匀分布。

最近一项研究表明, 人类能从极少量的数据中学习, 可能是因为随机映射算法^[15], 而神经网络正是根据人脑的生物结构所设计。因此, 在如今大数据时代, 随机映射算法必将成为机器学习领域内一项非常有用的技术。

3 RandPG 网络

RandPG 的网络结构如图 2 所示, 其包含一个输入层, 一个隐层和一个输出层。输入层中包含 d 个神经元, 表示样本的 d 维属性。与 RBF 网络一样, 输入层与隐层通过权值为 1 的权矩阵连接。隐层采用式(1)的 PG 函数作为激活函数, 通过随机投影的方法确定参数 $w_i, b_i (i = 1, 2, \dots, c)$ 后, 隐层和输出层的连接关系可用权值矩阵 \mathbf{U} 表示。因此, 输入输出层间的对应关系可描述为如下映射

$$g(\mathbf{x}) : \mathbf{R}^d \rightarrow \mathbf{R}^m \quad (4)$$

即

$$\mathbf{y} = g(\mathbf{x}) = \mathbf{U}^T \cdot \Phi = \mathbf{U}^T (\Phi_1^{\text{PG}}(\mathbf{x}), \Phi_2^{\text{PG}}(\mathbf{x}), \dots, \Phi_c^{\text{PG}}(\mathbf{x}))^T \quad (5)$$

式中

$$\Phi_i^{\text{PG}}(\mathbf{x}) = \Phi_i^{\text{PG}}(\mathbf{x}, w_i, b_i, \sigma_i) = \exp\left(-\frac{|\mathbf{w}_i^T \cdot \mathbf{x} - b_i|^2}{2\sigma_i^2}\right) \quad i = 1, 2, \dots, c \quad (6)$$

3.1 RandPG 网络的学习过程

RandPG 网络的训练过程分为两个阶段: 第一阶段采用随机投影的方法产生隐层中激活函数的参数 w_i 值, 由平面计算公式计算出 b_i 值; 第二阶段通过 Moore-Penrose 广义逆计算隐层与输出层的连接权值矩阵 \mathbf{U} 。具体的步骤描述如下。

(1) 计算激活函数的参数值。给定 N 个不同的训练样本 $\{(\mathbf{x}_i, t_i) \mid \mathbf{x}_i \in \mathbf{R}^d, t_i \in \mathbf{R}, 1 \leq i \leq N\}$, 根据随机投影原理, 将原始 d 维数据矩阵投影到与样本类别数相同的 c 维空间, 即产生投影矩阵 $w_j = (w_{1j}, w_{2j}, \dots, w_{dj}), 1 \leq j \leq c$, 其中每个 w_j 是独立的随机变量且都服从正态分布, 并将 w_j 按列进行归一化, 即 $\|w_j\| = 1$; 然后求出每类样本中所有点的各个属性的平均值, 即获得每类样本的中心点 $m_j (1 \leq j \leq c)$, 根据点到平面的距离公式 $|\mathbf{w}_j^T \cdot m_j - b_j| = 0$, 即可求出与 w_j 相对应的偏置 b_j , 即 $b_j = \mathbf{w}_j^T \cdot m_j$ 。宽度 σ_j 可以自行取值, 本文取值 0.5。

(2) 计算隐层输出矩阵。当隐层的激活函数确定后, 按式(1)定义, 可得到隐层输出矩阵 $\tilde{\Phi} = (\Phi_1,$

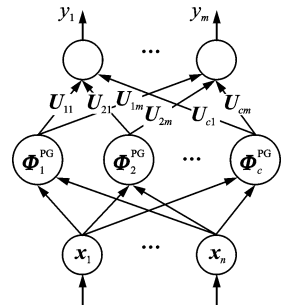


图 2 3 层 RandPG 网络结构图

Fig. 2 Three-layer RandPG neural network

Φ_2, \dots, Φ_N), 其中 $\Phi_j = (\Phi_1^{\text{PG}}(\mathbf{x}_j), \Phi_2^{\text{PG}}(\mathbf{x}_j), \dots, \Phi_j^{\text{PG}}(\mathbf{x}_j))^T, j=1, 2, \dots, N$ 。

(3) 采用 0-1 编码标记目标输出矩阵。样本类别标号用一个只有 0 和 1 组成的列向量来表示。例如某样本 \mathbf{x}_i 属于第 l 类, 则其第 l 个分量为 1, 其他元素都为 0, 则其 0-1 编码表示为 $\mathbf{y}_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ 。 N 个样本的目标输出矩阵表示为 $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ 。

(4) 连接隐层和输出层的连接权值矩阵。根据式(5), 利用 Moore-Penrose 广义逆求得 $\mathbf{U} = (\mathbf{Y} \cdot \tilde{\Phi}^+)^T$, 其中 $\tilde{\Phi}^+$ 表示矩阵 $\tilde{\Phi}$ 的广义逆矩阵。

按照上述步骤求得隐层激活函数的中心超平面的法向量 \mathbf{w}_i 、截距 b_i 和宽度 σ_i 及隐层与输出层的连接权值矩阵 \mathbf{U} 后, 就完成了 RandPG 网络的构建。

3.2 RandPG 网络的全局逼近性

RandPG 网络中第 i 个隐层神经元的激活函数形式为 $\Phi(\mathbf{x}) = \exp\left(-\frac{|\mathbf{w}_i^T \cdot \mathbf{x} - b_i|^2}{2\sigma_i^2}\right)$, 与其对应的激活区域呈带状, 如图 3 所示。

图 3 描述了该带状区域的中心是超平面, 在 RandPG 网络中, 与此超平面对应的 \mathbf{w}_i 和 b_i 由随机投影方法确定, σ_i 为半带宽。从几何解释图可以看出, PG 函数的激活区域表现出两种特性: 全局性和局部性, 即在超平面的法线方向上, 它是以 σ_i 为界的, 而在平行于超平面的方向上, 它是无界的。

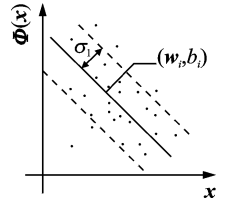


图 3 平面高斯函数带状激活区
Fig. 3 Band-shaped field of plane-Gaussian function

定理 1 (函数逼近定理) 设激活函数为任意非常数分段连续函数, 若通过调整隐层神经元的参数可以使单层前馈网络逼近任意目标函数 $t(\mathbf{x})$, 那么可以按照任意连续概率分布函数随机生成 $\{h_i(\mathbf{x})\}_{i=1}^N$, 并且可以找到合适的输出权值 β , 使 $\lim_{N \rightarrow \infty} \left\| \sum_{i=1}^N \beta_i h_i(\mathbf{x}) - t(\mathbf{x}) \right\| = 0$ 依概率收敛于 1。

下面从理论上证明 RandPG 网络的全局逼近性。给定一个有 N 个隐层节点的 RandPG 神经网络, 其隐层激活函数 $g: \mathbf{R} \rightarrow \mathbf{R}$ 在任意区间内无限可微, 对于 N 个不同的训练样本 $(\mathbf{x}_i, t_i), \mathbf{x}_i \in \mathbf{R}^n, t_i \in \mathbf{R}^m$, 随机选择输入层的权值 \mathbf{w}_i 和隐层偏置 $b_i, \mathbf{w}_i \in \mathbf{R}^n, b_i \in \mathbf{R}$, 根据连续概率分布, 隐层输出矩阵 \mathbf{H} 是可逆的且 $\|\mathbf{H}\mathbf{B} - \mathbf{T}\| = 0$ 。

证明: 隐层输出矩阵 \mathbf{H} 的第 i 列 $\mathbf{c}(b_i) = [g_i(\mathbf{x}_1), \dots, g_i(\mathbf{x}_N)]^T = [g(\mathbf{w}_i \cdot \mathbf{x}_1 + b_i), \dots, g(\mathbf{w}_i \cdot \mathbf{x}_N + b_i)]^T$, 在欧氏空间 \mathbf{R}^N 中 $b_i \in (a, b), (a, b)$ 为 \mathbf{R} 上的一段区间。可以证明向量 \mathbf{c} 不属于任何维度小于 N 的子空间。

由于 \mathbf{w}_i 是在连续概率分布上随机选择, 那么可以假设对于所有的 $k \neq k', \mathbf{w}_i \cdot \mathbf{x}_k \neq \mathbf{w}_i \cdot \mathbf{x}_{k'}$ 。假设 \mathbf{c} 属于 $N-1$ 维的子空间, 那么就存在一个向量 α 正交于这个子空间, 即有

$$(\alpha, \mathbf{c}(b_i) - \mathbf{c}(a)) = \alpha_1 \cdot g(b_i + d_1) + \alpha_2 \cdot g(b_i + d_2) + \dots + \alpha_N \cdot g(b_i + d_N) - z = 0 \quad (7)$$

式中: $d_k = \mathbf{w}_i \cdot \mathbf{x}_k, k=1, \dots, N$ 且对于 $b_i \in (a, b), z = \alpha \cdot \mathbf{c}(a)$ 。假设 $\alpha_N \neq 0$, 式(7)可以扩展成为

$$g(b_i + d_N) = - \sum_{p=1}^{N-1} \gamma_p g(b_i + d_p) + z/\alpha_N \quad (8)$$

式中 $\gamma_p = \alpha_p/\alpha_N, p=1, \dots, N-1$ 。RandPG 网络激活函数 $g(x)$ 中 $|\mathbf{w}^T \cdot \mathbf{x} - b|$ 是一个数值, 所以 $g(x)$ 在任意区间都无限可微, 因此有

$$g^{(l)}(b_i + d_N) = - \sum_{p=1}^{N-1} \gamma_p g^{(l)}(b_i + d_p) \quad l=1, 2, \dots, N, N+1, \dots \quad (9)$$

式中 $g^{(l)}$ 为激活函数 g 关于 b_i 的第 l 阶导数。但是, 只有 $N-1$ 个自由系数 $(\gamma_1, \dots, \gamma_{N-1})$ 对应于推导出的超过 $N-1$ 个的线性方程, 这是相背的。因此向量 \mathbf{c} 不属于任何维度小于 N 的子空间。

所以从任意区间 (a, b) 上随机选择 N 个隐层节点的偏置 b_1, \dots, b_N 使得相应的向量 $\mathbf{c}(b_1), \mathbf{c}$

$(b_2), \dots, c(b_N)$ 扩展到 \mathbf{R}^N 是可能的, 这意味着对于任意的 $w_i \in \mathbf{R}^n, b_i \in \mathbf{R}$, 根据连续概率分布, \mathbf{H} 矩阵满秩。

4 实验结果

为验证 RandPG 网络的分类性能, 本文分别在呈直线型分布和呈平面型分布的人工数据集和 8 个 UCI 数据集上进行实验, 并与 PG 网络和 ELM 在分类精度和分类时间上进行比较。实验均是从整个样本数据集中随机抽取一半作为训练样本, 剩下的一半作为测试样本。每个数据集分别进行 50 次分类实验。对于实验结果的评价标准, 由于 RandPG 网络和 PG 网络的投影方向具有几何意义, 代表的是拟合平面, 所以选取 50 次分类实验中训练精度和测试精度的最大值作为最优投影方向, 而 ELM 算法没有明确的几何意义, 所以选取 50 次分类实验中训练精度和测试精度的平均值作为最终结果, 训练时间和测试时间都取 50 次实验的平均值。为了保证 3 种算法的公平性, PG 网络和 ELM 的隐层节点数都与类别个数相同。

4.1 人工数据集

人工数据集的实验主要是为了检验 RandPG 网络对呈子空间分布的数据的分类是否具有优越性。对此, 实验设计了二维空间中的多类直线型分布数据和多维空间中的平面型分布数据。

4.1.1 直线型数据集

在二维空间中分别设计了 2~6 类呈直线型分布的数据, 每个类别含 100 个样本点, 其中 2 类数据集抽样于线段 $y=2x+1$ 和 $y=-x-1, x \in [-5, 5]$; 3 类数据集分别抽样于线段 $y=3x, y=8-x$ 和 $y=(x-5)/2, x \in [-2, 8]$; 4 类数据集分别抽样于 4 条线段 $y=\frac{1}{2}x+4, y=x-5, y=-2x+5$ 和 $y=3x-2, x \in [-2, 8]$; 5 类数据集分别抽样于 5 条线段 $y=x-9, y=-\frac{1}{2}x, y=-3x-7, y=-x-14$ 和 $y=x-18, x \in [-2, 8]$; 6 类数据集分别抽样于 6 条线段 $y=-2x-2, y=3-x, y=(3x+2)/4, y=2x+14, y=5x-18$ 和 $y=-\frac{2}{3}x-15, x \in [-2, 8]$ 。每个数据集都在 y 上附加均值为 0, 标准差为 0.5 的标准正态分布的噪声扰动。各个数据集的分布图如图 4 所示。

分别用 RandPG, PG 和 ELM 3 种方法训练上述 5 个二维空间中呈直线分布的人工数据集, 比较 3 种神经网络随着类别数的增加对直线型数据的分类性能, 训练精度和测试精度分别如图 5 和图 6 所示, 训练时间和测试时间分别如图 7 和图 8 所示。

从图 5~8 可以看出, 当类别数较少时, PG 网络的分类精度较高, 随着类别数的增加, RandPG 网络在分类精度方面展现出了优势, 而 ELM 对直线型分布的数据分类精度都不高。每种方法的训练精度和测试精度的趋势走向基本一致, 因此通过训练过程得到的参数对网络的泛化能力起着至关重要的作用。从训练时间的角度看, RandPG 网络和 ELM 的训练速度都很快, 而 PG 网络的训练时间大概是它们的 10 倍; 这和每种网络参数产生的原理相吻合, PG 网络产生参数时需要逐步迭代, 势必会耗费更多的时间, 而其他两种网络都是随机产生参数, 所以在训练时间上比较有优势。

4.1.2 平面型数据集

分别在三维、四维和五维空间设计呈平面状分布的数据, 每个空间都是包含 3 类平面。三维空间中的第一类样本抽样于平面 $x-y-z=0, [y, z] \in (-4, 5)$, 在 x 轴上附加范围在 $(-0.5, 0.5)$ 的随机噪声扰动; 第二类平面抽样于平面 $2x+y+5z-2=0, [x, z] \in (-4, 5)$, 在 y 轴上附加范围在 $(-0.5, 0.5)$

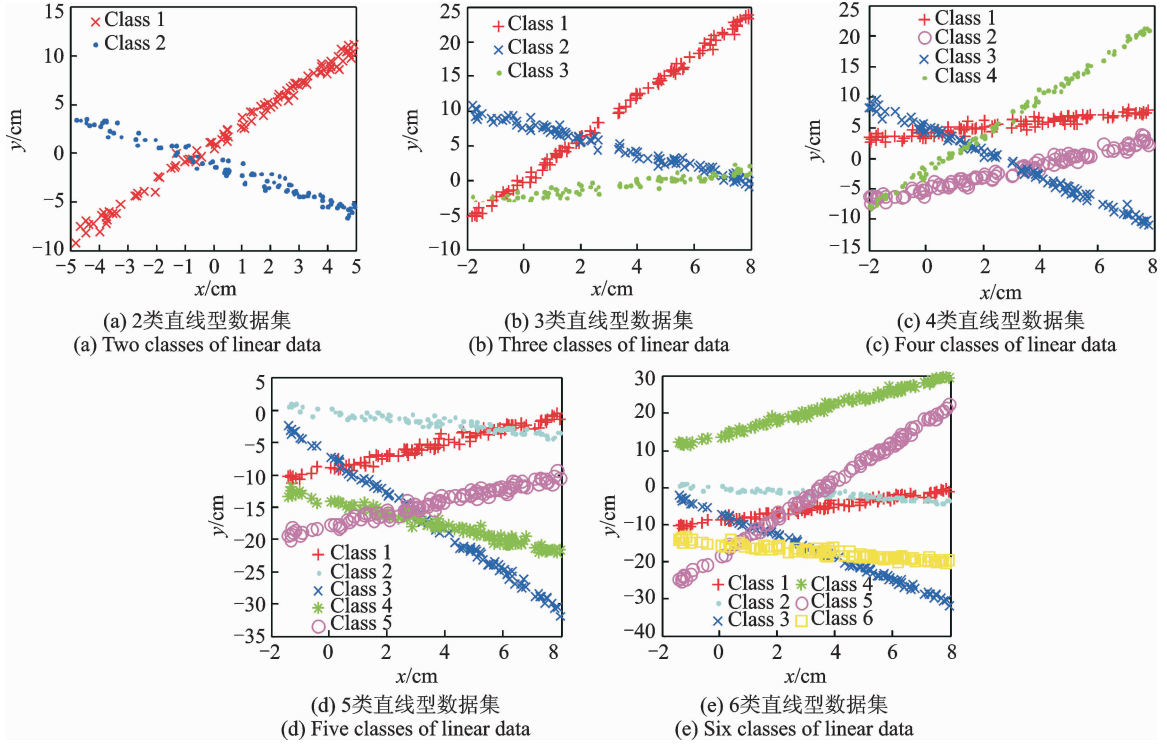


图 4 2~6 类直线型数据分布图

Fig. 4 Different classes of linear data distribution diagram

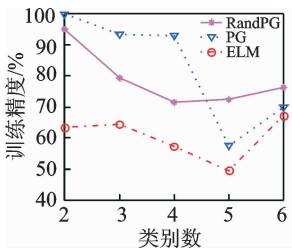


图 5 直线型分布数据的训练精度

Fig. 5 Training accuracy of linear distribution data

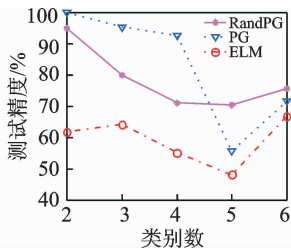


图 6 直线型分布数据的测试精度

Fig. 6 Testing accuracy of linear distribution data

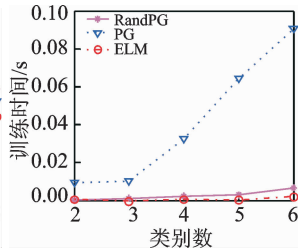


图 7 直线型分布数据的训练时间

Fig. 7 Training time of linear distribution data

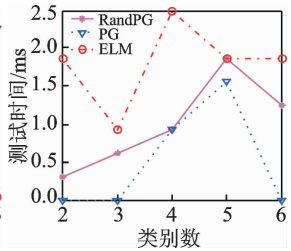


图 8 直线型分布数据的测试时间

Fig. 8 Testing time of linear distribution data

的随机噪声扰动;第三类平面抽样于 $3x - 2y + z - 3 = 0, [x, y] \in (-4, 5)$, 并在 z 轴上附加范围在 $(-0.5, 0.5)$ 的随机噪声, 每类包含 100 个点。数据分布如图 9 所示。

四维空间中的第一类样本抽样于平面 $a + b - 2c - 4d - 1/2 = 0, [b, c, d] \in (-4, 5)$, 在 a 轴上附加范围在 $(-0.5, 0.5)$ 的随机噪声扰动;第二类平面抽样于平面 $a - 2b + c + 2/3d + 3 = 0, [a, b, d] \in (-4, 5)$, 在 c 轴上附加范围在 $(-0.5, 0.5)$ 的随机噪声扰动;第三类平面抽样于 $-2a - 3b + 5c + d - 1 = 0, [a, b, c] \in (-4, 5)$, 并在 d

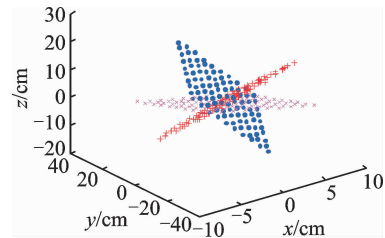


图 9 三维平面型数据分布图

Fig. 9 3D plane data distribution

轴上附加范围在 $(-0.5, 0.5)$ 的随机噪声。每类样本包含 1 000 个点。

五维空间中的样本点分别抽样于平面 $a + 2b - 3c + 4d + 2e + 2 = 0$, $3a - b + 5c - 2d + 3e + 4 = 0$ 和 $-a + 3b - 2c + d + 4e + 3 = 0$, a, b, c, d 轴的范围是均值为 0, 方差为 1 的正态随机分布, e 轴的值通过平面公式计算得到。每类样本包含 1 000 个点。

对三维、四维和五维空间中平面型数据的实验结果如图 10~12 所示, 其中图 10 和图 11 分别表示训练精度和测试精度, 图 12 和图 13 分别描绘了训练时间和测试时间。

实验结果表明, PG 网络在平面型数据集上的分类具有较大的优势, 不管在哪一维空间中, 分类精度都达到了 90% 以上。但是随着空间维数的增加, PG 网络的训练时间急剧增加, 是 RandPG 网络的百倍以上, 所以对于高维数据集, RandPG 网络在分类时间上具有较大的优势。与 ELM 相比, RandPG 网络对平面型分布数据的分类精度略高, 而在训练时间上, 两者相差无几。

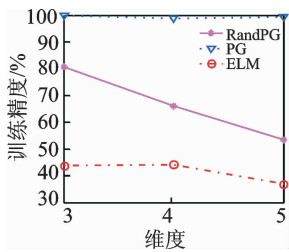


图 10 平面型分布数据的训练精度

Fig.10 Training accuracy of plane data

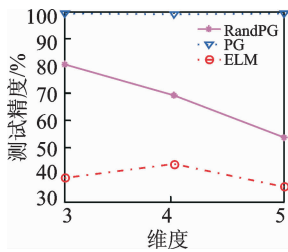


图 11 平面型分布数据的测试精度

Fig.11 Testing accuracy of plane data

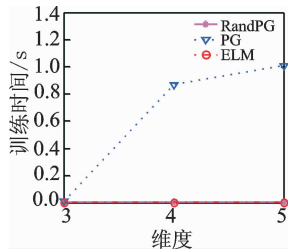


图 12 平面型分布数据的训练时间

Fig.12 Training time of plane data

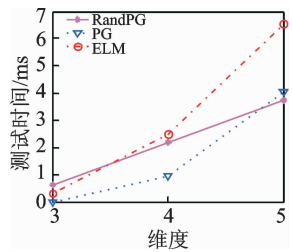


图 13 平面型分布数据的测试时间

Fig.13 Testing time of plane data

4.2 UCI 数据集

为了检验 RandPG 网络对真实数据集的分类能力, 本文选取 glass, ionosphere, iris, lenses, new_thyroid, soybean_small, dermatology 和 ecoli 8 个 UCI 数据集 (<http://archive.ics.uci.edu/ml>) 进行实验, 数据集的基本信息如表 1 所示。

表 1 UCI 数据集基本信息

Tab.1 Main information about UCI data sets

数据集	维数	样本数	类别信息	数据集	维数	样本数	类别信息
glass	9	214	2 [51,163]	new_thyroid	5	215	3 [150,35,30]
ionosphere	34	351	2 [126,225]	soybean_small	35	47	4 [10,10,10,17]
iris	4	150	3 [50,50,50]	dermatology	33	366	6 [112,61,72,49,52,20]
lenses	4	24	3 [4,5,15]	ecoli	6	332	6 [143,77,52,35,20,5]

在 8 个 UCI 数据集上的实验结果如表 2 所示。从表 2 可以很明显地看出 RandPG 网络对其中大部分数据集的分类精度都是最高的。PG 网络在 ionosphere, new_thyroid, dermatology 和 ecoli 等数据集上的分类精度虽然与 RandPG 网络很接近, 但是需花费数十倍甚至百倍的训练时间来完成每一次实验。这说明采用随机投影的方式, 经过多次实验不仅能寻找到最优的投影方向, 而且能节省算法一步步迭代所耗费的时间, RandPG 网络在分类精度和分类时间上都具有较大的优势。在 3 种方法的隐层节点数都与样本类别数相同的条件下, ELM 的分类精度明显处于劣势。这是因为 ELM 的网络结构决定了必须提供足够多的隐层节点个数才能保障其分类精度, 而在实际情况下不可能提供无限多个隐层节点数, 这样会增加网络的复杂性。在分类时间上, 由于 ELM 算法的隐层参数也是随机产生, 所以与 RandPG

有着相同数量级的训练时间。

表 2 UCI 数据集上的实验结果

Tab. 2 Experimental results of UCI data sets

数据集	方法	训练精度/%	测试精度/%	训练时间/s	测试时间/s
glass	RandPG	94.33	95.37	0.001 872	0.000 936
	PG	87.73	87.96	0.012 168	0.000 000
	ELM	76.86	76.24	0.000 312	0.000 624
ionosphere	RandPG	81.71	80.11	0.001 872	0.000 000
	PG	78.28	81.25	0.063 024	0.000 936
	ELM	65.54	64.79	0.000 624	0.001 248
iris	RandPG	96.00	96.00	0.000 624	0.000 000
	PG	88.00	84.00	0.005 616	0.000 000
	ELM	77.94	77.12	0.000 624	0.000 000
lenses	RandPG	90.90	76.92	0.003 432	0.000 000
	PG	90.90	92.30	0.006 864	0.000 624
	ELM	75.09	62.15	0.000 624	0.000 000
new_thyroid	RandPG	81.30	80.55	0.000 936	0.000 000
	PG	82.24	80.55	0.009 672	0.000 000
	ELM	69.25	68.35	0.001 248	0.001 560
soybean_small	RandPG	82.60	75.00	0.001 872	0.000 000
	PG	95.65	95.83	0.016 536	0.000 000
	ELM	69.73	62.58	0.000 000	0.002 496
dermatology	RandPG	75.82	79.89	0.006 240	0.000 000
	PG	75.27	77.17	0.215 281	0.000 936
	ELM	70.18	67.41	0.002 184	0.002 808
ecoli	RandPG	84.75	82.73	0.004 992	0.000 936
	PG	84.14	82.14	0.032 760	0.000 000
	ELM	83.18	81.78	0.001 248	0.002 184

5 结束语

针对 PG 网络通过聚类方法确定隐层激活函数参数的缺点,本文借鉴 ELM 随机产生网络参数的思想,提出了 RandPG 网络。该网络运用随机投影的方法确定网络的结构和隐层的权值,一定程度上避免了网络训练易陷入局部极小值的问题,使得分类结果更稳定。通过对直线型、平面型数据集及 UCI 标准数据库中部分分类数据集的测试,结果表明,RandPG 神经网络不仅继承了 PG 网络适用于平面型数据分类的特性,而且其学习速度明显有所提高。较之 ELM,该网络具有清晰的几何意义,且隐节点个数与数据集类别数相同,无需人为干预设定。将所提出的 RandPG 神经网络应用于复杂的呈子空间分布的数据分类问题,可以大大提高分类的效率和准确率。在今后的研究工作中,可以考虑将 RandPG 网络应用于回归问题,也可以设计多隐层的 RandPG 网络。

参考文献:

- [1] Huang Guangbin, Zhu Qinyu, Siew C K. Extreme learning machine: Theory and application[J]. Neurocomputing, 2006, 12(1): 489-501.
- [2] 卢涛,杨威,万永静.基于图像超分辨率学习机的极低分辨率人脸识别[J].计算机应用,2016,36(2):580-585.

- Lu Tao, Yang Wei, Wan Yongjing. Very low resolution face recognition via super-resolution based on extreme learning machine[J]. *Journal of Computer Applications*, 2016, 36(2): 580-585.
- [3] An Le, Yang Songfan, Bhanu B. Efficient smile detection by extreme learning machine[J]. *Neurocomputing*, 2015, 149: 354-363.
- [4] Rong Haijun, Wei Jintao, Bai Jianming, et al. Adaptive neural control for a class of MIMO nonlinear systems with extreme learning machine[J]. *Neurocomputing*, 2015, 149: 405-414.
- [5] Gao Xianghui, Wong K I, Wong P K, et al. Adaptive control of rapidly time-varying discrete-time system using initial-training-free online extreme learning machine[J]. *Neurocomputing*, 2016, 194: 117-125.
- [6] 王保义, 赵硕, 张少敏. 基于云计算和极限学习机的分布式电力负荷预测算法[J]. *电网技术*, 2014, 38(2): 526-531.
Wang Baoyi, Zhao Shuo, Zhang Shaomin. A distributed load forecasting algorithm based on cloud computing and extreme learning machine[J]. *Power System Technology*, 2014, 38(2): 526-531.
- [7] Mulia I E, Asano T, Nagayama A. Real-time forecasting of near-field tsunami waveforms at coastal areas using a regularized extreme learning machine[J]. *Coastal Engineering*, 2016, 109: 1-8.
- [8] 张英堂, 马超, 李志宁, 等. 基于快速留一交叉验证的核极限学习机在线建模[J]. *上海交通大学学报*, 2014, 48(5): 641-646.
Zhang Yingtang, Ma Chao, Li Zhining, et al. Online modeling of kernel extreme learning machine based on fast leave-one-out cross-validation[J]. *Journal of Shanghai Jiaotong University*, 2014, 48(5): 641-646.
- [9] Liu Zhen, Li Hanxiong. Extreme learning machine based spatiotemporal modeling of lithium-ion battery thermal dynamics [J]. *Journal of Power Sources*, 2015, 277: 228-238.
- [10] Yang Xubing, Chen Songcan, Chen Bin. Plane-Gaussian artificial neural network[J]. *Neural Computing and Applications*, 2012, 21(2): 305-317.
- [11] 王颖. 基于超平面原型的聚类算法及相应扩展神经网络的研究[D]. 南京: 南京航空航天大学, 2006.
Wang Ying. Research on hyperplane prototype clustering algorithm and its generalization in neural networks[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2006.
- [12] Bradley P S, Mangasarian O L. k-Plane clustering[J]. *Journal of Global Optimization*, 2000, 16(1): 23-32.
- [13] Fern X Z, Brodley C E. Random projection for high dimensional data clustering: A cluster ensemble approach[C]// *International Conference on Machine Learning(ICML)*. Washington D C: [s. n], 2003: 186-193.
- [14] Johnson W B, Lindenstrauss J. Extensions of Lipschitz mappings into a Hilbert space[J]. *Contemporary Mathematics*, 1984, 26(1): 189-206.
- [15] Arriaga R I, Rutter D, Cakmak M, et al. Visual categorization with random projection[J]. *Neural Computation*, 2015, 27(10): 2132-2143.

作者简介:



冯哲(1992-),女,硕士研究生,研究方向:模式识别、神经计算, E-mail: 535716987@qq.com。



杨绪兵(1973-),男,副教授,研究方向:模式识别、神经计算。



张福全(1977-),男,副教授,研究方向:无线传感器网络,网络通信和林业物联网。

