权重随机正交化的极速非线性判别分析网络

谢群辉[1],田 青[2]
(1. 南京航空航天大学 计算机科学与技术学院,江苏 南京 211106;
2. 南京信息工程大学 计算机与软件学院,江苏 南京 210044)

摘 要:极速学习机(Extreme Learning Machine,ELM)以其训练速度快、易实现、泛化性能好等优点受到了广泛关注。然而在数据维度较高的场景,数据中往往蕴含着较多冗余信息,而经典 ELM 尚未能很好地应对这个问题。此外,经典 ELM 也未能对标记数据的判别信息有效地加以融合利用。针对传统 ELM 方法的不足之处,提出一种权重随机正交的判别分析网络(O-ENDA)。在 O-ENDA 中,一方面对 ELM 输入层权重施加正交约束,这就降低了输入特征的冗余信息以减低过拟合的风险(尤其在小样本场景下);另一方面将隐层特征与判别分析相融合进行联合学习,实现数据判别信息在 ELM 中的融合利用。实验结果表明,提出方法在保持数据判别特征的同时能够去除其冗余信息、提高模型的泛化能力并能获得更高的分类精度。

关键词:极速学习机;线性判别分析;神经网络;降维;分类
中图分类号:TP389.1 **文献标识码**:A **文章编号**:1673-629X(2018)01-0023-05
doi:10.3969/j.issn.1673-629X.2018.01.005

Nonlinear Discriminant Analysis Networks with Random and Orthogonalized Input Weights

XIE Qun-hui[1],TIAN Qing[2]
(1. School of Computer Science and Technology,Nanjing University of Aeronautics and Astronautics,
Nanjing 211106,China;
2. School of Computer and Software,Nanjing University of Information Science and Technology,
Nanjing 210044,China)

Abstract:Extreme Learning Machines (ELM) has attracted increasing attention due to its fast training speed,simplicity and good generalization. However,in applications with high-dimensional features,a large amount of redundant information may exist in the data,which has not been concerned in classical ELM. Moreover,the discriminant information behind data has also not been incorporated in the ELM learning. To overcome the drawbacks of classical ELM,a weight-orthogonal discriminant analysis network (O-ENDA) is put forward. In O-ENDA,the input weights of ELM are restricted to be orthogonal,removing the redundant features to alleviate the risk of over-fitting (especially in the scenario of small samples);simultaneously the transformed hidden features are combined with discriminant analysis to improve the discriminant ability of O-ENDA. Experiment demonstrates that the proposed approach not only can remove redundant input information while preserving the necessary feature information,but also entirely yields preferable classification accuracy than classical ELM.

Key words:extreme learning machine;linear discriminant analysis;neural networks;dimension reduction;classification

1 概 述

极速学习机(Extreme Learning Machines,ELM)[1]以其训练速度快,易实现,泛化性能好等优点受到广泛关注。相关研究证明,ELM 具有很好的万能逼近能力(Universal Approximation Capability)[2]。相比 SVM 等机器学习分类方法,ELM 具有更好的分类泛化性和学习快速性[3],并在特征学习、分类、回归和聚类等方面获得了一系列拓展[4-6]。例如,极速非线性判别分析方法(Extreme Nonlinear Discriminant Analysis,ENDA)就是在 ELM 基础上提出的,通过随机初始化输入

收稿日期:2016-12-31 **修回日期**:2017-05-02 **网络出版时间**:2017-09-27
基金项目:国家自然科学基金资助项目(61472186);中国博士后科学基金特别资助项目(20133218110032);南京信息工程大学人才启动基金
作者简介:谢群辉(1984-),男,硕士研究生,研究方向为机器学习。

权重取代了传统的前馈神经网络迭代训练,提高了计算效率;利用 LDA 进行特征提取并能获得全局最优,避免了神经网络局部最小的问题;通过对线性不可分数据进行非线性化特征选择,在高维空间中进行 LDA 降维和特征判别分析,并且获得可视化效果,非常有利于后续的分类。

非线性化方法主要包含神经网络方法和核方法两大类,其中神经网络非线性化方法又包含深度学习和浅层网络(ELM)两大类。文献[6-7]表明,随机初始化正交权重,可以改善其泛化性能。权重的正交化方法更适用于高维的图像聚类或分类。在处理小样本、高维度的数据时,权重正交化的设计可以去除特征以外的噪声,提高模型的计算效率。通过初始化正交权重,避免传统深度学习的反向迭代计算权重,达到优化非线性组合的目的。文献[8-10]表明,正交性在深度网络学习中至关重要,正交权重层通常被设计成滤波器组,实现更完备特征的提取,达到非常好的无监督特征分类学习效果,给高维大数据学习正交化提供理论依据。在 MATLAB 环境下实现卷积神经网络(Convolutional Neural Networks)的深度集成化[11],其中滤波器模块利用权重正交化约束的设计。通过权重的随机正交化可以过滤冗余信息,以达到快速训练深度神经网络权重的目的。这种权重正交层被称为施蒂费尔(Stiefel Layer),在很多 BP 网络中被广泛应用,它的作用是降维和特征提取。文献[7]提出的传统的深度学习,不仅学习速度缓慢,而且在计算资源上花费巨大。为解决这一问题,通过极速学习机自动编码(ELM-AE)代替传统的 BP 梯度下降算法,大大提高了计算速度,节省了计算成本。ELM-AE 正是通过随机正交初始化权重、偏置以及输出权重达到子空间约束的目的。利用多层正交化 ELM 网络组成多层神经网络,在计算速度上要比传统的深度学习快很多。考虑到深度学习计算的时效性问题,核方法显然比深度的神经网络方法要快,而有关核方法的权重正交设计方法[12-13]相继被提出,这足以说明,在学习理论中正交概念是至关重要的。

在机器学习中,由于多层网络训练困难,核方法又受到核函数隐节点线性与样本数的限制,当样本数很大时,核方法计算花费大。而只有一层隐含层的浅层模型备受欢迎,作为机器学习的重要发展方向-极速化的浅层学习,ELM 正是这种浅层学习代表。在此基础上,文中提出了正交约束设计特征选择方法(O-ENDA)。受 ELM-AE 的启发,将原有浅层神经网络 ENDA 改造成 O-ENDA,不仅能过滤图像的冗余信息,而且与前面提到的深度学习对比充分发挥了浅层神经网络计算速度快的优势,在保证分类性能的基础上,O-ENDA 比深度学习和核方法更容易实现极速化。在保证数据结构化多样化特性的前提下,对输入权重正交化强制约束,能够提取结构特征效果,正交后输入权重映射在保持多样性的基础上同样可以提高计算效率。考虑到隐层的节点数与数据维度空间问题,从而实现降维(或者升维)作用,在保证数据原有多样化特性的情况下,降低了数据的冗余信息,提取数据特征。

2 正交特征映射

2.1 极速学习机

极速学习机是由黄广斌提出的快速求解单隐层神经网络的快速学习算法。与传统的 Back-Propagation(BP)算法不同的是,利用权重的随机初始化设置输入权重和偏差,代替传统的梯度下降的权重学习。假设网络层输入为 $\boldsymbol{x}$,目标输出为 $\boldsymbol{T}$,估计输出为 $\boldsymbol{y}$。

随机初始化隐层节点,给定一个训练集 $\{(\boldsymbol{x}_i,\boldsymbol{t}_i)\mid \boldsymbol{x}_i\in\mathbb{R}^n,\boldsymbol{t}_i\in\mathbb{R}^m,i=1,2,\cdots,N\}$,$\boldsymbol{x}_i$ 为特征输入向量,$\boldsymbol{t}_i$ 为对应的输出目标向量,L 为隐节点个数。ELM 代价函数最小的目的是最小化训练误差,最小化输出权重:

$$\min_{\boldsymbol{\beta}_i} L_{\mathrm{ELM}}=\frac{1}{2}\|\boldsymbol{\beta}\|^2+\frac{C}{2}\sum_{j=1}^{N}\|\boldsymbol{H}\boldsymbol{\beta}-\boldsymbol{T}\|^2 \tag{1}$$

其中,$\boldsymbol{H}$ 为隐层输出矩阵。

$$\boldsymbol{H}=\begin{bmatrix}\boldsymbol{h}(x_1)\\ \boldsymbol{h}(x_2)\\ \vdots\\ \boldsymbol{h}(x_N)\end{bmatrix}=\begin{bmatrix}h_1(x_1) & \cdots & h_L(x_1)\\ h_1(x_2) & \cdots & h_L(x_2)\\ \vdots & \vdots & \vdots\\ h_1(x_N) & \cdots & h_L(x_N)\end{bmatrix}\in\mathbb{R}^{N\times L} \tag{2}$$

$\boldsymbol{T}$ 为训练数据目标矩阵。

$$\boldsymbol{T}=\begin{bmatrix}\boldsymbol{t}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{t}_1^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}t_{11}\cdots t_{1m}\\ \vdots\ \vdots\ \vdots\\ t_{N1}\cdots t_{Nm}\end{bmatrix} \tag{3}$$

ELM 训练算法流程如下:

步骤 1:初始化权重 $\boldsymbol{w}_i^{(1)}$ 和偏置 b_i,$i=1,2,\cdots,L$,分别服从(-1,1)和(0,1)上的随机均匀分布。

步骤 2:计算隐层输出矩阵 $\boldsymbol{H}$。

步骤 3:获得输出权重向量。

$$\boldsymbol{\beta}^{\dagger}=\boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{4}$$

当 $\boldsymbol{T}=[\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_N]^{\mathrm{T}}\in\mathbb{R}^{N\times m}$,$\boldsymbol{H}^{\dagger}$ 为 $\boldsymbol{H}$ 的 Moore-Penrose 广义逆矩阵。

考虑到 ELM 优化方法中更好的泛化性能,解得:

$$\boldsymbol{\beta}^{\dagger}=\left(\frac{\boldsymbol{I}}{C}+\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{T} \tag{5}$$

对应的 ELM 输出函数表示为:

$$f(\boldsymbol{x}_j)=h(x)\beta=h(x)\left(\frac{\boldsymbol{I}}{C}+\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{T} \tag{6}$$

2.2　权重随机初始化

回顾 ENDA 的训练过程,其网络结构为输入层、隐层和输出层三层网络,共分为两个步骤。第一步,随机生成输入层的连接权重和偏置,对输入进行随机特征映射;第二步,计算 LDA 层投影权重,将隐层输出作为 LDA 输入数据,同时输出层可用于可视化,如图 1 所示。由于第一层为随机生成权重,所以只需计算第二层的投影权重,整个模型计算简单、有效。

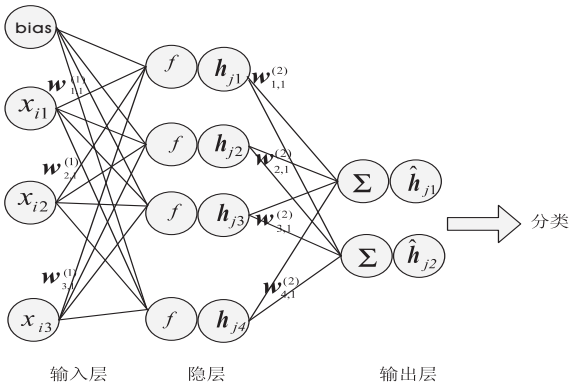

图 1　O-ENDA 网络结构

与 ELM 步骤类似,只需计算第二步使用 LDA 对 $\boldsymbol{H}$ 进行降维。对 $\boldsymbol{H}$ 经过式(4)投影后,得到新样本 $\hat{\boldsymbol{H}}=[\hat{\boldsymbol{h}}_1^{\mathrm{T}},\cdots,\hat{\boldsymbol{h}}_N^{\mathrm{T}}]\in\mathbb{R}^{N\times P}$。设:

$$\boldsymbol{W}=[\boldsymbol{w}_1^{(2)},\boldsymbol{w}_2^{(2)},\cdots,\boldsymbol{w}_P^{(2)}] \tag{7}$$

$$\hat{\boldsymbol{H}}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{H} \tag{8}$$

其中,$\boldsymbol{w}_i^{(2)}$ 为 LDA 的投影权重。

样本总类内散度矩阵 $\boldsymbol{S}_{\mathrm{b}}$ 和类间散度矩阵 $\boldsymbol{S}_{\mathrm{w}}$ 分别定义为:

$$\boldsymbol{S}_{\mathrm{b}}=\sum_{i=1}^{c}n_i(\boldsymbol{\mu}_i-\boldsymbol{\mu})(\boldsymbol{\mu}_i-\boldsymbol{\mu})^{\mathrm{T}} \tag{9}$$

$$\boldsymbol{S}_{\mathrm{w}}=\sum_{i=1}^{c}\sum_{\hat{\boldsymbol{h}}\in\boldsymbol{H}_i}(\hat{\boldsymbol{h}}-\boldsymbol{\mu}_i)(\hat{\boldsymbol{h}}-\boldsymbol{\mu}_i)^{\mathrm{T}} \tag{10}$$

其中,$\hat{\boldsymbol{H}}_i$ 表示 $\hat{\boldsymbol{H}}$ 中属于第 i 类的样本;$\boldsymbol{\mu}_i=\frac{1}{n_i}\sum_{\hat{\boldsymbol{h}}\in\boldsymbol{H}_i}\hat{\boldsymbol{h}}$;$\boldsymbol{\mu}=\frac{1}{N}\sum_{i=1}^{c}n_i\boldsymbol{\mu}_i$,$n_i$ 为第 i 类样本数。构建的优化目标函数如下:

$$\max_{\boldsymbol{W}}\boldsymbol{J}(\boldsymbol{W})=\frac{|\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{b}}\boldsymbol{W}|}{|\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{w}}\boldsymbol{W}|} \tag{11}$$

然后对其进行求解,最大化 $\boldsymbol{J}(\boldsymbol{W})$ 等价于求解如下广义特征问题:

$$\boldsymbol{S}_{\mathrm{b}}\boldsymbol{W}=\lambda\boldsymbol{S}_{\mathrm{w}}\boldsymbol{W} \tag{12}$$

通过求解式(8)可计算出投影矩阵 $\boldsymbol{W}$,它由 $\boldsymbol{S}_{\mathrm{w}}^{-1}\boldsymbol{S}_{\mathrm{b}}$ 的 $c-1$ 个最大广义特征值所对应的特征向量组成,此 $\boldsymbol{W}$ 即为最佳的判别投影矩阵。由 Kasun 等提出的正交 ELM 自编码(ML-ELM)在性能上取得了优势而且快速,具有很好的泛化性能。受此启发,对极速非线性判别分析网络随机映射进行正交化。

2.3　正交特征选择

正交特征选择通过优化算法选择出有利于分类的特征子集[6],为避免“维数灾难”,处理高维数据,提出权重正交试验设计,从而选出数据中有代表的特征向量,在保留完备和均匀特征信息的条件下,减少了计算复杂度。假设样本空间 $\boldsymbol{x}_i=[x_1,x_2,\cdots,x_N]$,$x_j\in\mathbb{R}^n$,对应的标签 $\boldsymbol{T}=[t_{j1},t_{j2},\cdots,t_{jm}]\in\mathbb{R}^m$,将样本投影权重矩阵 $\boldsymbol{B}$ 到隐层,从而得到特征子空间 $\boldsymbol{H}$。

首先讨论隐层节点参数 L 构建不同的或者相等的维度空间,L 与样本维度 n 的关系:当 $n=L$ 时,属于平等维度表示;当 $n>L$ 时,属于压缩表示(降维);当 $n<L$ 时,属于升维表示(稀疏)。假设样本特征维度 $n>L$,这里需要对输入层进行降维,随机初始化正交权重进行改造。

$$\begin{cases}\boldsymbol{H}=g(\boldsymbol{B}\boldsymbol{x}_i+b_i)\\ \text{s.t. }\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}\\ \boldsymbol{B}=\boldsymbol{A}+\boldsymbol{E}\end{cases} \tag{13}$$

其中,$\boldsymbol{E}$ 为高斯扰动;$g(\cdot)$ 为 Sigmoid 激励函数。

正交化后必须保持子空间的基不变,等效优化效果。$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}$ 的约束条件是 $\boldsymbol{A}\in\{\boldsymbol{R}\boldsymbol{W}\}$,$\boldsymbol{R}\in\{\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}=\boldsymbol{I}_p\}$ 等效优化效果,$\boldsymbol{A}$ 是正交的,通过计算经验误差度量学习效果。O-ENDA 模型的经验损失如下:

$$\begin{aligned}\boldsymbol{E}&\triangleq\frac{1}{N}\sum_{i=1}^{N}(\|\boldsymbol{y}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i\|^2)=\\&\frac{1}{N}\sum_{i=1}^{N}[(\boldsymbol{y}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i)^{\mathrm{T}}(\boldsymbol{y}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i)]=\\&\frac{1}{N}\sum_{i=1}^{N}\{[\boldsymbol{y}_i-(\boldsymbol{R}\boldsymbol{W})^{\mathrm{T}}\boldsymbol{x}_i]^{\mathrm{T}}[\boldsymbol{y}_i-(\boldsymbol{R}\boldsymbol{W})^{\mathrm{T}}\boldsymbol{x}_i]\}\end{aligned} \tag{14}$$

O-ENDA 算法步骤如下:

步骤 1:对数据进行预处理;

步骤 2:计算 O-ENDA 第一部分中初始化 Gram-Schmidt 正交化权重向量 $\boldsymbol{w}_i^{(1)}$ 和随机初始化偏置 b_i,$i=1,2,\cdots,L$;

步骤 3:计算 O-ENDA 的隐层输出 h_i(通过式(13));

步骤 4:对隐层输出 h_i 进行判别分析,通过式(8)计算投影权重 $\boldsymbol{W}$,得到 O-ENDA 的输出 $\hat{\boldsymbol{h}}$;

步骤 5:对输出层结果进行分类。

3　实　验

仿真环境如下:MATLAB R2015a,Intel(R) Core™ i5-3470 CPU @3.20 GHz,16.0 GB 内存,64 位 Win10 专业版操作系统。采用 UCI 机器学习数据库 CNAE-9

数据、手写数字 MNIST 数据和 CIFAR10 数据分别进行实验。实验中完整的数据集被划分为 10 部分,训练和测试数据集使用了 10 重交叉验证方法。对输出层采用 KNN 算法进行分类,比较原 ENDA 和正交化改进后的 O-ENDA 的分类性能,三种数据如表 1 所示。分别为文本、手写数字、图片的数据,冗余信息较多,维度逐渐加大,对于了解正交化改进后的效果有一定代表性。

表 1 数据集

数据集	Instances	Features	Classes
CNAE-9	1 080	857	9
MNIST	70 000	784	10
CIFAR10	50 000	3 072	10

3.1 UCI 数据

CNAE-9 数据集包含 1 080 个文档的自由文本,共分为 9 个类。从原始文本进行预处理获得当前数据集。每个文档被表示为一个向量,每个词的权重是它在文档中出现的频率。其中,高斯扰动参数设置均值 μ 为 0,方差 σ^2 为 1。

调整隐节点参数获得重复 20 次实验平均值和方差,在 CNAE-9 数据集上的表现如图 2 所示。实验结果表明,在维度较大的情况下,ENDA 和 O-ENDA 的分类效果都很明显,O-ENDA 分类的性能更稳定。

从 O-ENDA 和 ENDA 的误差率来看,O-ENDA 分类效果更好。讨论了隐节点参数 L 与数据原有维度 n 的关系,当 $L > n$ 时,都正交化需要调整权重矩阵,与 $L < n$ 时的权重正好相反。这个数据集因为较简单,$L < n$ 就能达到较好的效果。这里讨论 L 从 50 到 250 的情况,当 L 持续增大时,出现了过拟合。从 CNAE-9 的特征维度到隐层实际上是进行了降维,正交化后数据的多样性得以保留,数据冗余噪声减少,性能更具有独立性,分类的效果也更稳定。

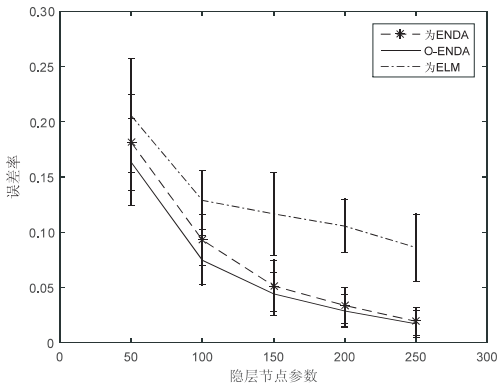

图 2 CNAE-9 数据集误差率对比曲线

3.2 MNIST 数据

该数据为手写数字 MNIST 字体库,包含 70 000 个样本。要对 0 至 9 手写数字图像进行识别,每一个手写数字样本是一个 28 * 28 像素的图像,因此对于每一个样本,其输入信息就是每一个像素对应的灰度,总共有 784(28 * 28)个像素,也就是数据中包含 784 个特征,实验前需对转换成普通的图像格式进行预处理。

手写数字识别大多来源于邮政编码以及银行业务自动识别,由于字体变化大,识别要求高,在多数情况下通常采用多网络的深度学习来提高识别率,主要利用神经网络非线性化学习能力和快速并行来提高识别率。目的在于验证浅层的 O-ENDA 可以通过正交过滤器消除冗余信息来解决此类问题。其中,高斯扰动参数设置均值 μ 为 0,方差 σ^2 为 0.1。

从图 3 可以看出,ENDA 具有很好的分类精度,而且 O-ENDA 的误差率平均值比 ENDA 还要小,这充分表明,在 MNIST 数据特征维度到隐层维度,实际上是进行了非线性化升维再进行判别分析。O-EDNA 去除了更多冗余信息,更加易于分类,保留特征具有线性独立的特点,因为正交的随机权重分布更均匀,提取特征更加完备,无论是从分类精度还是标准差上,O-ENDA 都要优于原来 ENDA。

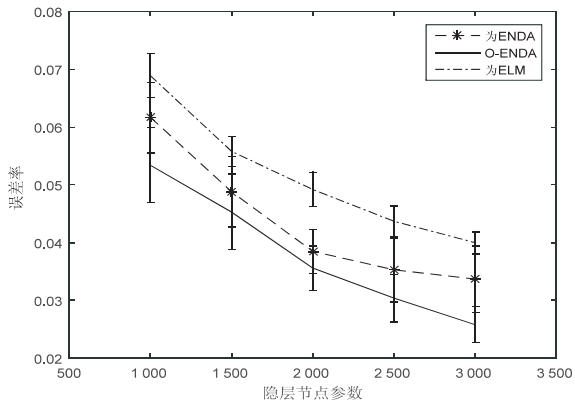

图 3 MNIST 数据集误差率对比曲线

3.3 CIFAR10 数据

CIFAR10 数据显示的每一行存储 32×32 的彩色图像,共 50 000 个数据集,第 1 024 项包含红色通道值,1 024绿色,最后 1 024 蓝色。标签范围为 0 ~ 9,10 个类别每个类图片 5 000 张。随机采集 10 000 个样本为 10 000×3 072 的矩阵进行实验。

比较 MNIST 数据和 UCI 数据,CIFAR10 维度更大,权重正交化效果更明显。从误差率比较 ENDA 和 O-ENDA,结果如图 4 所示。其中,高斯扰动参数设置均值 μ 为 0,方差 σ^2 为 0.1。

从图 4(重复 20 次实验)可以看出,当 L = 3 500 时,O-ENDA 权重正交化的误差性能更加稳定。当隐节点参数在 2 000 ~ 4 000 范围变化时,与数据维度 3 072相比,分别进行了降维、等维和升维三种情况的表示。如图 5 所示,可以得到 CIFAR10 数据集 20 次误差率平均值和节点参数关系图,从图中可以看到,正交化 O-ENDA 使得物体更加线性独立和易于分类。

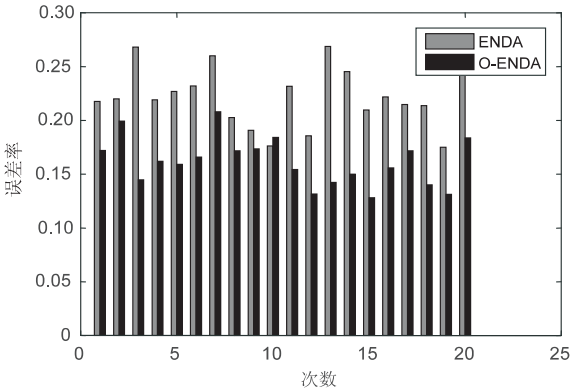

图 4　隐节点参数 3 500 时的误差率直方图

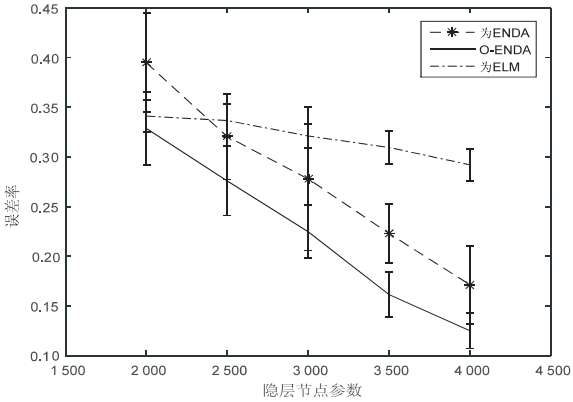

图 5　CIFAR10 数据集误差率对比曲线

3.4　实验结果分析

从三个数据集上的实验结果表明,对较高维度的数据进行分类时,O-ENDA 比 ENDA 分类效果更好,特别是在 CIFAR10 数据集作用更加明显。由于图像数据冗余信息越多,权重正交化取得分类效果就越明显,O-ENDA 算法分类性能越好,且隐层更能代表原样本数据多样性特征,充分验证了正交作为神经网络一种重要的特征提取方法的有效性,从而也开启了浅层神经网络对高维度正交提取完备特征,能有效防止“维数灾难”。

4　结束语

在极速非线性化判别分析网络基础上提出正交约束。权重 Gram-Schmidt 正交化方法作为神经网络一种重要的优化方法,与传统神经网络方法和核方法相比,具有速度和可伸缩性两方面的优势。权重正交的对象为高维度小样本数据集,通过正交化局部保持投影的度量学习分析,正交的随机权重分布使得特征之间均匀更加线性独立,一方面降低了数据的冗余信息,利于后续分类;另一方面提取更为完备的特征,整体上提高了模型泛化性能。实验表明正交优化能提高算法的性能。

集成学习能够对模型比较独立的样本进行训练,然后把结果整合起来进行整体的投票决策。对于随机映射每个不同的基分类器集成[14],这是一种非常有效的方法。计算机的并行化和批量扩展能力更加适合集成学习的未来发展,这也是下一步研究的方向。

参考文献:

[1] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: a new learning scheme of feedforward neural networks [C]//International joint conference on neural networks. [s. l.]: IEEE, 2004: 985-990.

[2] HUANG G B, CHEN L, SIEW C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes [J]. IEEE Transactions on Neural Networks, 2006, 17(4): 879-892.

[3] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 2012, 42(2): 513-529.

[4] HUANG G B, BAI Z, KASUN L L C, et al. Local receptive fields based extreme learning machine [J]. IEEE Computational Intelligence Magazine, 2015, 10(2): 18-29.

[5] WIDROW B, GREENBLATT A, KIM Y, et al. The no-prop algorithm: a new learning algorithm for multilayer neural networks [J]. Neural Networks, 2013, 37(1): 182-188.

[6] ONETO L, BISIO F, CAMBRIA E, et al. Statistical learning theory and ELM for big social data analysis [J]. IEEE Computational Intelligence Magazine, 2016, 11(3): 45-55.

[7] KASUNL L C, ZHOU H, HUANG G B, et al. Representational learning with ELMs for big data [J]. IEEE Intelligent Systems, 2013, 28(6): 31-34.

[8] TURK M, PENTLAND A. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86.

[9] HYVARINEN A. Fast and robust fixed-point algorithms for independent component analysis [J]. IEEE Transactions on Neural Networks, 1999, 10(3): 626-634.

[10] CAI D, HE X, HAN J, et al. Orthogonal laplacianfaces for face recognition [J]. IEEE Transactions on Image Processing, 2006, 15(11): 3608-3614.

[11] VEDALDI A, LENC K. Matconvnet: convolutional neural networks for matlab [C]//Proceedings of the 23rd ACM international conference on multimedia. [s. l.]: ACM, 2015: 689-692.

[12] 黄金杰,常英丽. 基于支持向量机和正交设计的特征选择方法[J]. 计算机工程与应用,2008,44(17):135-137.

[13] 金　一,阮秋琦. 基于核的正交局部保持投影的人脸识别[J]. 电子与信息学报,2009,31(2):283-287.

[14] YIN X C, HUANG K, YANG C, et al. Convex ensemble learning with sparsity and diversity [J]. Information Fusion, 2014, 20: 49-59.