极限学习机综述

陆思源，陆志海，王水花，张煜东
（南京师范大学 计算机科学与技术学院，江苏 南京 210023）

摘要：极限学习机是一种单隐层前向网络的训练算法，主要特点是训练速度极快，而且可以达到很高的泛化性能。回顾了极限学习机的发展历程，分析了极限学习机的数学模型，详细介绍了极限学习机的各种改进算法，并列举了极限学习机在识别、预测和医学诊断领域的应用。最后总结预测了极限学习机的改进方向。

关键词：极限学习机；机器学习；人工神经网络；综述
中图分类号：TP181　　**文献标识码**：A　　**文章编号**：1000 - 8829(2018)10 - 0003 - 07
doi：10.19708/j.ckjs.2018.10.001

Review of Extreme Learning Machine

LU Si-yuan, LU Zhi-hai, WANG Shui-hua, ZHANG Yu-dong
(School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China)

Abstract: Extreme learning machine (ELM) is a novel training algorithm of single-hidden layer feedforward network (SLFN). The training of ELM is extremely fast while obtaining good generalization ability. The development of ELM is reviewed, the mathematical model of ELM is analyzed, the different improvements of ELM are introduced, and the applications of ELM in pattern recognition, forecasting and medical diagnosis are listed. Finally, several ways for improving ELM are summarized.

Key words: extreme learning machine(ELM); machine learning; artificial neural network; review

近年来，机器学习成为研究的热点主题。机器学习中最著名的传统学习算法主要有：支持向量机(Support Vector Machine，SVM)、决策树(Decision Tree，DT)、K 近邻(K-Nearest Neighbors，KNN)和反向传播神经网络(Back Propagation Neural Network，BPNN)。SVM 的思想是寻找最大分类间隔，把分类问题转化为凸二次规划问题解决。SVM 适合解决小样本分类问题。DT 算法的核心是通过信息熵等度量方法对数据集进行分析，来构建树状决策结构，每个内部节点表示一个属性上的测试，每个分支代表一个测试输出，每个叶节点代表一种类别。常用的训练算法为 ID3 和 C4.5。KNN 是一种典型的非参数分类方法，它把数据集分为训练集和测试集；对于每个测试集的样本，KNN 先找到距离该样本最近的 K 个训练样本，然后把这 K 个样本标签中占比最多的标签作为该样本的标签。BPNN 是一种前向神经网络，在训练时，反向传播误差，利用梯度下降法迭代更新网络参数，直到收敛。

极限学习机(Extreme Learning Machine，ELM)是由新加坡南洋理工大学的 Huang 等人[1]提出的一种单隐层前向神经网络(Single-Hidden Layer Feedforward Network，SLFN)的训练算法。不同于传统的训练算法(如 BP 算法等)，ELM 算法对输入层的权值和偏置进行随机赋值，然后用求 Moore-Penrose 广义逆矩阵的方法直接解出隐层到输出层的权值。ELM 的优势有：需要手动设置的参数只有隐含层结点个数，算法执行过程中不需要人工调整参数；避免了传统训练算法反复迭代的过程，快速收敛，极大地减少了训练时间；所得解是唯一最优解，保证了网络的泛化性能。目前 ELM 已经广泛应用到各种回归和分类问题中，也出现了许多 ELM 的改进算法[2-5]来应对具体的问题。

收稿日期：2017 - 10 - 05
基金项目：国家自然科学基金项目(61602250)；江苏省自然科学基金(BK20150983)；国家重点研发计划(2017YFB1103200/02)
作者简介：陆思源(1991—)，男，硕士，主要研究方向为模式识别；陆志海(1969—)，男，硕士，高级工程师，主要研究方向为图像处理；王水花(1985—)，女，博士，讲师，主要研究方向为医学图像处理；张煜东(1985—)，男，博士，教授，博士生导师，主要研究方向为人工智能与医学图像处理。

1　理论基础

1.1　SLFN 结构

一个典型的 SLFN 是由输入层、隐含层和输出层构成,如图 1 所示。其中输入层和输出层的结点数是由具体问题确定的,网络中的待定参数有:隐含层结点数、输入层到隐含层权值、隐含层到输出层的权值和隐含层的偏置。

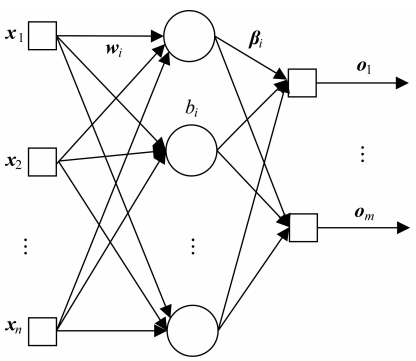


图 1　SLFN 的结构图

1.2　问题描述

SLFN 的一般学习模式可以描述为:对于任意的 N 个不同的样本$(\boldsymbol{x}_i,\boldsymbol{t}_i)$,$\boldsymbol{x}_i=[x_{i1},x_{i2},\cdots,x_{in}]^{\mathrm{T}}\in\mathbf{R}^n$,$\boldsymbol{t}_i=[t_{i1},t_{i2},\cdots,t_{im}]^{\mathrm{T}}\in\mathbf{R}^m$,$g(x)$为激活函数,标准的有$\hat{N}$个隐含结点的 SLFNs 可以表示为

$$\sum_{i=1}^{\hat{N}}\boldsymbol{\beta}_i g_i(\boldsymbol{x}_j)=\sum_{i=1}^{\hat{N}}\boldsymbol{\beta}_i g_i(\boldsymbol{w}_i\cdot\boldsymbol{x}_j+b_i)=\boldsymbol{o}_j,\ j=1,\cdots,N \tag{1}$$

式中,$\boldsymbol{w}_i=[w_{i1},w_{i2},\cdots,w_{in}]^{\mathrm{T}}$ 为连接第 i 个隐含结点和输入层的权值向量;$\boldsymbol{\beta}_i=[\beta_{i1},\beta_{i2},\cdots,\beta_{im}]^{\mathrm{T}}$ 为连接第 i 个隐含结点和输出层的权值向量;b_i 为第 i 个隐含结点的偏置;$\boldsymbol{w}_i\cdot\boldsymbol{x}_j$ 为 $\boldsymbol{w}_i$ 和 $\boldsymbol{x}_j$ 的内积。

这样的一个 SLFN 可以无限逼近这 N 个样本:$\sum_{j=1}^{N}\|\boldsymbol{o}_j-t_j\|=0$,也就是说,存在$\boldsymbol{\beta}_i$、$\boldsymbol{w}_i$、$b_i$ 使得:

$$\sum_{i=1}^{\hat{N}}\boldsymbol{\beta}_i g_i(\boldsymbol{w}_i\cdot\boldsymbol{x}_j+b_i)=\boldsymbol{t}_j,\ j=1,\cdots,N \tag{2}$$

这$\hat{N}$个等式可以写成

$$\boldsymbol{H}\boldsymbol{\beta}=\boldsymbol{T} \tag{3}$$

其中,

$$\boldsymbol{H}(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{\hat{N}},b_1,\cdots,b_{\hat{N}},\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)=\begin{bmatrix}g(\boldsymbol{w}_1\cdot\boldsymbol{x}_1+b_1)&\cdots&g(\boldsymbol{w}_{\hat{N}}\cdot\boldsymbol{x}_1+b_{\hat{N}})\\ \vdots&\cdots&\vdots\\ g(\boldsymbol{w}_1\cdot\boldsymbol{x}_N+b_1)&\cdots&g(\boldsymbol{w}_{\hat{N}}\cdot\boldsymbol{x}_N+b_{\hat{N}})\end{bmatrix}_{N\times\hat{N}} \tag{4}$$

$$\boldsymbol{\beta}=\begin{bmatrix}\boldsymbol{\beta}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{\beta}_{\hat{N}}^{\mathrm{T}}\end{bmatrix}_{\hat{N}\times m},\quad \boldsymbol{T}=\begin{bmatrix}\boldsymbol{t}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{t}_N^{\mathrm{T}}\end{bmatrix}_{N\times m} \tag{5}$$

式中,$\boldsymbol{H}$ 为神经网络隐含层的输出矩阵,$\boldsymbol{H}$ 的第 i 列为神经网络隐含层的第 i 个结点的输出。

1.3　理论证明

接着,Guang-Bin Huang[6] 严格证明了如下的结论。

① 如果激活函数 $g(x)$ 是无限可微的,那么可以随机地选择输入层到隐含层的权值 $\boldsymbol{w}_i$ 和隐含层的偏置 b_i,SLFN 就可以被当成一个线性系统,只需要通过解析求出隐含层到输出层的权值 $\boldsymbol{\beta}$ 就可以确定整个网络。$\boldsymbol{H}\boldsymbol{\beta}=\boldsymbol{T}$,$\boldsymbol{H}$ 是一个 $N\times\hat{N}$的矩阵,如果 $N=\hat{N}$,那么 $\boldsymbol{H}$ 是可逆的;如果 $N\neq\hat{N}$,那么求 $\boldsymbol{H}$ 的 Moore-Penrose 广义逆矩阵。这样得到的网络不仅具有最小的训练误差,而且权值的范数也最小。

根据 Bartlett 定律,神经网络的训练误差越小,权值的范数越小,那么它的泛化性能就越好。

② 对于给定的 SLFN,有 N 个隐含结点,激活函数 $g:R\to R$ 是无限可微的。对于任意的 N 个不同的样本$(\boldsymbol{x}_i,\boldsymbol{t}_i)$,$\boldsymbol{x}_i=[x_{i1},x_{i2},\cdots,x_{in}]^{\mathrm{T}}\in\mathbf{R}^n$,$\boldsymbol{t}_i=[t_{i1},t_{i2},\cdots,t_{im}]^{\mathrm{T}}\in\mathbf{R}^m$,任意随机的 $\boldsymbol{w}_i$ 和 b_i,不论它们服从什么样的连续分布,以下公式成立的概率为 100%。

$$\|\boldsymbol{H}\boldsymbol{\beta}-\boldsymbol{T}\|=0 \tag{6}$$

③ 给定一个任意小的正数 $\varepsilon>0$,无限可微的激活函数 $g:R\to R$,存在$\hat{N}\leqslant N$,使得对于任意的 N 个不同的样本$(\boldsymbol{x}_i,\boldsymbol{t}_i)$,$\boldsymbol{x}_i=[x_{i1},x_{i2},\cdots,x_{in}]^{\mathrm{T}}\in\mathbf{R}^n$,$\boldsymbol{t}_i=[t_{i1},t_{i2},\cdots,t_{im}]^{\mathrm{T}}\in\mathbf{R}^m$,任意随机的 $\boldsymbol{w}_i$ 和 b_i,不论它们服从什么样的连续分布,以下公式成立的概率为 100%。

$$\|\boldsymbol{H}\boldsymbol{\beta}-\boldsymbol{T}\|<\varepsilon \tag{7}$$

1.4　ELM 基本算法

基于以上的 3 个结论,Guang-Bin Huang[6] 提出了 ELM 基本算法:给定一个训练集 $S=\{(\boldsymbol{x}_i,\boldsymbol{t}_i)\mid\boldsymbol{x}_i\in\mathbf{R}^n,\boldsymbol{t}_i\in\mathbf{R}^m,i=1,\cdots,N\}$,激活函数 $g(x)$ 和隐含层结点个数$\hat{N}$。算法实现步骤如下:

① 对输入层到隐含层的权值 $\boldsymbol{w}_i$ 和隐含层的偏置 b_i 随机赋值,$i=1,\cdots,\hat{N}$;

② 计算隐含层的输出矩阵 $\boldsymbol{H}$;

③ 计算隐含层到输出层的权值 $\boldsymbol{\beta}$,$\boldsymbol{\beta}=\boldsymbol{H}^{+}\boldsymbol{T}$,其中 $\boldsymbol{H}^{+}$ 是矩阵 $\boldsymbol{H}$ 的 Moore-Penrose 广义逆矩阵。

通过 ELM 所得的解为唯一最优解,保证了网络的泛化性能。

2　改进算法

但是 ELM 也存在着一些问题,例如:由于隐含层参数是随机生成的,会导致使用不同初始化参数训练出的 ELM 泛化性能有差异,影响 ELM 的稳定性和鲁棒性,所以有学者对 ELM 进行分析并提出各种改进算法。

Guang-Bin Huang[6]提出了在线增量极限学习机(Online Sequential Extreme Learning Machine, OS-ELM),可以学习不断增长数据集。Guang-Bin 等人[7]提出了凸增量极限学习机(Convex Incremental Extreme Learning Machine,CI-ELM),可以解决增量模型中新增结点的训练问题。Huang 等人[2]分析比较了 ELM 和 SVM,得出以下结论:① SVM 的最大间隔特性与前向神经网络的权重范数最小化理论是一致的;② 从标准的优化算法角度来看,SVM 和 ELM 是等价的,但是由于 ELM 具有特殊的可分离特性,所以 ELM 的优化限制条件更少;③ 经过理论分析和模拟实验,ELM 比传统的 SVM 具有更好的泛化性能。由于 ELM 的输入层权值和偏置是随机生成的,这样可能导致隐含层的输出矩阵不是列满秩矩阵,会降低 ELM 的性能。针对这一问题,Wang 等人[8]提出了有效极限学习机(Effective Extreme Learning Machine,EELM),在计算输出层权值之前,调整输入层的权值和偏置,使得隐含层的输出矩阵满足列满秩条件,改进后的 EELM 算法可以减少训练时间,提高网络的分类准确率和网络的鲁棒性。Cao 等人[3]提出基于投票的极限学习机(Voting based Extreme Learning Machine,V-ELM)来避免随机生成的隐含层权值和偏置对分类结果造成的不稳定性。V-ELM 训练多个独立的具有相同结构和激活函数的 ELM,然后利用投票法来整合各 ELM 的结果。为了解决非均衡分布数据的分类问题,Zong 等人[4]提出了加权极限学习机(Weighted Extreme Learning Machine, Weighted ELM),该算法可以直接用于多分类问题,并且可以推广到代价敏感学习。Rong 等人[9]提出并应用模块化的极限学习机(Modular Extreme Learning Machine,M-ELM)来识别图像中的飞行器。Bai 等人[10]把 ELM 应用到对象通用类识别问题中,提出了一种基于局部感知域的极限学习机(Local Receptive Fields Based Extreme Learning Machine, ELM-LRF)。Liu 等人[11]提出一种通用的学习框架——多核极限学习机(Multiple Kernel Extreme Learning Machine,MK-ELM)来解决 ELM 核函数的选择和优化。Wang 等人[12]尝试给出了 ELM 泛化能力的振荡范围,并分析了 ELM 对隐含层结点个数不敏感的原因;实验结果表明,当隐含层结点数接近无穷时,ELM 的核函数与激活函数对参数是不敏感的,它的泛化性能与原始的 ELM 相当,可以避免过拟合。Deng 等人[13]为解决高维数据分类问题对 ELM 进行了改进,把奇异值分解的隐层结点嵌入极限学习机,提出了基于快速奇异值分解隐层结点的极限学习机(Fast Singular Value Decomposition-Hidden-Nodes Based Extreme Learning Machine, FSVD-H-ELM)。Xu 等人[14]深入研究了误差最小化极限学习机(Error Minimized Extreme Learning Machine, EM-ELM),并对其进行改进,提出了增量正则化极限学习机(Incremental Regularized Extreme Learning Machine, IR-ELM)和它的增强版(Enhancement of IR-ELM,EIR-ELM)。每当加入一个新的隐含层结点,IR-ELM 可以快速地递归求解输出权值。实验证明,IR-ELM 比 EM-ELM 的泛化性能更优。Wang 等人[15]使用灰狼优化算法(Grey Wolf Optimization,GWO)训练核极限学习机,寻找最优的 ELM 参数,提出了灰狼优化的核极限学习机(GWO-KELM)。Wang 等人[16]运用核融合方法来降低选择不同核函数对 ELM 分类性能的影响,并且使用削减核降低计算开销,在此基础上提出了混合迁移学习和削减核的极限学习机(Transfer Learning Mixed and Reduced Kernel Extreme Learning Machine,TransM-RKELM)。为了解决流数据的学习问题,Xu 等人[17]基于在线学习的方法训练 ELM,在训练过程中动态地添加隐含层结点,提出了动态极限学习机(Dynamic Extreme Learning Machine,DELM)。Zeng 等人[18]采用一种新颖的延时转换的粒子群优化算法(Switching Delayed Particle Swarm Optimization, SDPSO)来训练 ELM,把 ELM 的输入权值和隐含层偏置作为粒子的参数,用延时的局部最优解和全局最优解来更新粒子的位置和移动速度。

表 1 对上述改进算法进行了分析和总结,从中可以看出 ELM 的改进算法主要有以下 3 个方面:ELM 与在线学习结合、ELM 隐含层结点结构的改造和 ELM 的网络参数优化。

3 应用

3.1 识别领域

Yang 等人[19]提出基于面部图像的性别识别系统,采用局部三元模式(LTP)提取图像特征。LTP 是局部二元模式的一种推广,对面部光照差异具有更好的识别力并且对噪声有一定的容忍度。他们使用 ELM 作为分类器来识别图像中人物的性别。在实验中,使用公开数据集测试系统的性能。ELM 达到 87.13%的识别准确率,并且运行总时间仅为 1.87 s,性能明显优于 BP 网络和 SVM。

Weimin Huang[20]提出一种基于视频信息的人类动作识别算法。首先从视频中找到感兴趣的时间空间信息,再提取出局部形状和运动信息作为动作的描述特征。然后,提出一种新的最小类别变换的极限学习机(Minimum Class Variance Extreme Learning Machine,MCVELM)作为分类器来识别动作。MCVELM 在 4 个公开数据集上都取得了较先进的识别准确率。

表 1　ELM 改进算法的分析与总结

文献	改进算法	结果
文献[6]	提出 OS-ELM,可以根据数据集的增长动态添加隐含层结点	OS-ELM 在回归任务中的训练时间比其他顺序算法短,且精度较高
文献[7]	提出 CI-ELM,可以解决增量模型中新增结点的训练问题	在回归实验中,CI-ELM 的泛化性能总体上优于传统的增量极限学习机
文献[2]	将 ELM 算法应用到支持向量网络的训练	证明了 ELM 算法优于传统的 SVM
文献[8]	提出 EELM,在计算输出层权值之前,调整输入层的权值和偏置,使得隐含层的输出矩阵满足列满秩条件,提高网络的分类准确率和网络的鲁棒性	总体而言,EELM 在数据集较小的情况下,分类和回归的性能比传统 ELM 好
文献[3]	提出 V-ELM,通过训练多个 ELM 来降低随机性对网络泛化性能的影响,用投票法生成最终结果	在实验测试的 19 个数据集中 V-ELM 的性能均好于 ELM,并且在其中的 7 个数据集上有明显的提高
文献[4]	提出 Weighted ELM,可以解决非均衡分布数据的分类问题	在各种分布比例的非均衡数据集测试中,Weighted ELM 的泛化性能比传统 ELM 高
文献[9]	提出 M-ELM,把训练集划分成多个子集,分别用于训练不同的 ELM,并用聚类选择的方法整合它们的输出结果	在飞行器识别的实验中,M-ELM 的准确率高于单个 ELM、多模块的 SVM、C4.5 算法、K 近邻算法和朴素贝叶斯算法
文献[10]	提出 ELM-LRF,可以更好地检测识别局部结构	ELM-LRF 在 ETH-80 数据集上的准确率达到先进水平,在 NORB 和 COIL 上创造了新的记录
文献[11]	提出 MK-ELM 来解决 ELM 核函数的选择和优化	MK-ELM 在减少计算量的同时,可以达到甚至超过现有的最先进水平
文献[12]	尝试给出 ELM 泛化能力的振荡范围,并分析了 ELM 对隐含层结点个数不敏感的原因	当 ELM 的隐含层结点个数接近无穷时,它的学习能力与原始 ELM 相当,而且可以有效避免过拟合
文献[13]	提出 FSVD-H-ELM,引入快速奇异值分解隐含层结点	通过在 12 个高维数据集上的实验,FSVD-H-ELM 的准确率已经达到或超过现有的最先进算法
文献[14]	提出 IR-ELM 和 EIR-ELM,快速递归计算新增结点的权值	在 5 个基准数据集的实验中,EIR-ELM 都可以用更少的隐含层结点取得比 IR-ELM 更好的分类准确率
文献[15]	提出 GWO-KELM,利用群体计算优化 KELM 参数	GWO-KELM 在破产预测中达到 84.86% 的准确率,高于基于粒子群优化、基因算法和网格搜索的 KELM
文献[16]	提出 TransM-RKELM,运用核融合方法来降低选择不同核函数对 ELM 分类性能的影响,并且使用削减核降低计算开销	TransM-RKELM 提高了对于跨地域行为的识别准确率
文献[17]	提出 DELM,处理流数据	DELM 可以根据数据流自适应地调整隐含层结点参数,准确率已经达到最先进水平
文献[18]	提出 SDPSO-ELM,利用延时转换的粒子群优化算法训练网络	在负载预测实验中,SDPSO-ELM 的泛化性能高于传统的径向基函数神经网络

Chen 等人[21]把 KELM 运用到高光谱图像的识别分类问题中。分别使用 Gabor 算子和多假设预测(MH)的方法从高光谱图像中提取空间特征和光谱特征。然后分别用这两种特征训练 KELM。结果显示,基于 KELM 的算法准确率不仅高于传统的按像素的分类方法,而且也优于 SVM 算法。

Rong 等人[9]提出一种基于 M-ELM 的飞行器识别算法。首先从图像中提取 Hu 矩、Zernike 矩和小波矩,再通过多特征选择的算法得到最优特征集。然后把训练集划分成若干子集,分别用不同的子集训练不同的 ELM,最后采用聚类选择的方法来整合所得的各 ELM 的分类结果。实验结果显示,M-ELM 的分类性能优于多模块的 SVM、C4.5、K 近邻和朴素贝叶斯分类器。

对象通用类识别问题一直是一个难点,因为类内可变性很大。Bai 等人[10]用 ELM-LRF 作为分类算法,尝试解决这一问题。在 ELM 中加入了卷积和池化方法。池化层的结点权值为随机赋值,并且采用平方或开方的池化方法。最后,应用最小二乘算法计算输出权值。实验证明,ELM-LRF 可以直接应用于不同的数据集。该算法在 NORB、ETH-80 和 COIL 数据集上都取得了很好的分类准确率。

Cheng 等人[22]把 ELM 和深度学习结合起来,提出一种去噪深度极限学习机的自编码器(DDELM-AE)来解决手写数字识别问题。在实验中,用 USPS 数据集的 500 个样本进行训练,200 个样本进行测试。DDELM-AE 可以达到 96.1% 的识别准确率,高于 K-SVD 的 93.10% 和 SAE 的 95.25%。

普通的 ELM 输出多为标量,而在实际应用中往往是复杂的输出结果。Maliha 等人[23]对 ELM 进行改进,提出了具有结构化输出空间的极限学习机(ELM for Structured Output Spaces),用于图像中的对象识别定位问题,并利用 PASCAL 和 VOC2006 数据集对这种算法进行评估。结果表明,这种方法可以准确地识别对象的位置范围,识别效果比结构化的 SVM 好。

Ding 等人[24]在 KELM 中加入卷积和降采样,提出了卷积核极限学习机(CKELM),用来解决手写数字识别问题。卷积和降采样可以有效地从图像中提取特征信息,KELM 用于分类识别。在实验中,利用 MNIST 和 USPS 数据集对 CKELM 进行性能评估。CKELM 在

MNIST 上的准确率为 96.80%，在 USPS 上的准确率为 96.21%，均高于 KELM。

Wang 等人[16]使用惯性传感器收集数据，然后利用 TransM-RKELM 识别其中的人类行为。使用核融合的方法来降低选择不同核函数对识别性能的影响，核融合还可以降低计算复杂度。在线学习阶段，利用高置信度的识别结果对 TransM-RKELM 进行训练，使它能够适应不同的地点位置。实验结果表明，TransM-RKELM 可以自适应传感器位置，可以达到很高的识别准确率。

3.2 预测领域

Song 等人[25]提出一种基于 ELM 的多元电力消费预测模型。结合电力参数和环境参数作为特征，选择 ELM 作为预测算法。提出一种 DDMS-PSO 算法来进行所有离散型参数的优化，包括特征选择和 ELM 的隐含层结点个数。基于现实数据的预测实验显示，提出的算法比 SVR 的预测更准确。

Deng 等人[13]用改进后的 FSVD-H-ELM 解决高维数据分类问题。为了从高维数据中有效地提取出内在特征，把奇异值分解嵌入到 ELM 的隐含层结点。由于大型数据的奇异值分解的计算复杂度过高，引入一种分治逼近方案来平衡计算量。在训练时，没有直接使用整个数据集，而是先从原始数据中随机选取若干子集，然后利用这些子集生成隐含层的奇异值分解结点。实验结果表明，FSVD-H-ELM 具有较好的泛化性能和计算效率。

Wang 等人[15]提出 GWO-KELM，并应用到企业破产预测问题上。GWO 是一种群体优化算法，它模拟自然界中狼的等级制和狩猎行为来寻找解空间中的最优解。实验中利用 GWO 来训练 ELM 的隐含层结点参数，并与基于粒子群优化、基因算法和网格搜索的 KELM 进行对比。GWO-KELM 取得了 86.84% 的准确率，比其他 3 种优化算法的结果都要好。

Xu 等人[17]运用 DELM 对流数据进行学习分析。为了使分类器能够适应连续数据流中的概念漂移，采用了在线学习的机制来训练 ELM，并设计了一个双隐含层的结构来提高 ELM 的性能。在训练过程中，DELM 可以动态地调整网络的隐含层结点数量，来适应数据流的改变。经过实验发现，DELM 可以提高对数据流的分类准确率，并且可以很快地适应新的概念。

Zeng 等人[18]采用 SDPSO-ELM 来进行短期电力负载预测，实验中，对比了 SDPSO-ELM 和径向基函数神经网络（Radial Basis Function Neural Network，RBFNN）的性能，结果显示所提出的 SDPSO-ELM 比 RBFNN 的预测结果更接近实际情况。

3.3 医学诊断领域

Wang 等人[26]基于乳房 X 射线成像，运用 ELM 检测乳腺疾病。首先利用图像纹理分析从乳房 X 射线成像中提取的特征，再使用启发式搜索确定特征的数量和 ELM 隐含层结点的数量。ELM 在实验中的平均诊断准确率达到了 91%。

Ramalho 等人[27]利用 KELM 基于核磁共振成像进行病脑诊断，首先对脑部核磁共振图像进行小波变换，然后从所得的小波子带中提取熵，作为图像特征，最后训练 KELM 进行自动分类诊断。实验结果显示，该诊断方法可以达到 97.48% 的灵敏度、94.44% 的特异度和 97.04% 的准确率。随后，Arunadevi 等[28]采用蝙蝠算法（Bat Algorithm，BA）训练 ELM，并应用于病脑诊断，将诊断准确率提高到了 98.33%。

Zhu 等人[29]使用 Jaya 算法训练 ELM，并用于病脑检测中。为了解决脑核磁共振成像数据集样本分布不均衡的问题，首先对数量较少的正常脑样本进行重采样。然后，对图像进行小波包变换，再提取 Tsallis 熵作为样本的特征向量，最后使用 Jaya 算法训练 ELM 作为分类器。对比实验显示，Jaya-ELM 的泛化性能比 GA-ELM、PSO-ELM 和 BA-ELM 高。

表 2 对 ELM 的应用进行了梳理，列出了 ELM 的应用领域以及实验结果。可以发现，相比于传统的 SVM 和 BP 网络等分类算法，基于 ELM 的算法往往具有明显的优势，不仅运行时间极短，而且泛化性能较好。专家学者针对具体问题，对 ELM 做出各式各样的改进，来进一步提高它的分类精度和工作效率。

4 结束语

本文回顾了 ELM 的发展历程，分析了 ELM 的数学模型，详细介绍了 ELM 的各种改进算法，并列举了 ELM 在识别、预测和医学诊断领域的应用。可以发现，相比于传统的 SLFN 训练算法，ELM 的最大特点是算法没有迭代过程，大幅减少了网络的训练时间，同时有效地避免局部最优解，求得唯一全局最优解，保证了网络的泛化性能。

ELM 的改进思路主要有以下 3 种：

① 降低随机性对网络性能的影响，通过群体优化算法（如 PSO、GA、GWO 等）寻找最优的参数提高网络的鲁棒性，但此类方法一般需要参数迭代更新，所以计算复杂度较高；

② 将 ELM 与在线学习和增量模型结合，提升网络的可塑性，此类方法往往更针对具体的数据；

③ 将 ELM 与深度学习结合，加入自编码器、卷积、降采样等深度学习的方法[30-32]，可以实现自动提取特征，提高分类准确率。

表 2　ELM 的应用

应用领域	作者	结果
识别	Yang 等人[19]	在性别识别实验中 ELM 的识别准确率达到 87.13%，运行时间仅为 1.87s，不仅准确率高于 SVM 和 BP 网络，而且所需时间极短
	Huang 等人[20]	MCVELM 在动作识别实验中达到了先进水平，并且对于视角的变换具有很好的鲁棒性
	Chen 等人[21]	KELM 在高光谱图像的识别分类中的准确率高于传统按像素的分类算法和基于 SVM 的算法
	Rong 等人[9]	在飞行器识别的实验中，M-ELM 的准确率高于单个 ELM、多模块的 SVM、C4.5 算法、K 近邻算法和朴素贝叶斯算法
	Cheng 等人[22]	通过 USPS 的手写数字识别实验，发现 DDELM-AE 的识别准确率比 K-SVD 和 SAE 高，但它的运行时间略短于 K-SVD，比 SAE 长
	Maliha 等人[23]	结构化的 ELM 可以达到 ELM 的泛化性能，并且在对象识别定位问题中可以取得比结构化 SVM 更准确的精度
	Ding 等人[24]	CKELM 可以自动地提取图像中的特征，并且在 MNIST 和 USPS 两个数据集上的识别准确率均高于传统的 ELM 和 KELM，但是运行时间介于 KELM 和 ELM 之间
	Bai 等人[10]	ELM-LRF 在 ETH-80 数据集上的准确率达到先进水平，在 NORB 和 COIL 上创造了新的记录
	Wang 等人[16]	TransM-RKELM 提高了对于跨地域行为的识别准确率
预测	Song 等人[25]	采用 DDMS-PSO 优化的特征和 ELM 模型可以找到最优参数，实验中的预测准确率比 SVR 更高
	Deng 等人[13]	FSVD-H-ELM 在 12 个高维数据集上的准确率已经达到或超过现有的最先进算法
	Wang 等人[15]	GWO-KELM 在企业破产预测问题上取得了 86.84% 的准确率，比基于粒子群优化、基因算法和网格搜索的 KELM 的准确率都高
	Xu 等人[17]	DELM 可以提高对数据流的分类准确率，并且可以很快地适应新的概念
	Zeng 等人[18]	在短期电力负载预测实验中，SDPSO-ELM 比 RBFNN 的预测结果更接近实际情况
医学诊断	Wang 等人[26]	在乳腺疾病的诊断中，ELM 达到了 90% 的灵敏度、98% 的特异度和 91% 的平均准确率，性能优于 SVM、KNN 等算法
	Ramalho 等人[27]	KELM 可以达到 97.48% 的灵敏度、94.44% 的特异度和 97.04% 的准确率
	Arunadevi 等人[28]	基于 BA-ELM 的诊断方法可以达到 98.33% 的诊断准确率
	Zhu 等人[29]	结合小波包变换、Tsallis 熵和 Jaya-ELM 的病脑检测系统可以达到 99.64% 的灵敏度、99.14% 的特异度和 99.57% 的准确率

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]// Proceedings of 2004 IEEE International Joint Conference on Neural Networks. 2004: 985 – 990.

[2] Huang G B, Ding X J, Zhou H M. Optimization method based extreme learning machine for classification[J]. Neurocomputing, 2010, 74(1/2/3): 155 – 163.

[3] Cao J W, Lin Z P, Huang G B, et al. Voting based extreme learning machine[J]. Information Sciences, 2012, 185(1): 66 – 77.

[4] Zong W W, Huang G B, Chen Y Q. Weighted extreme learning machine for imbalance learning[J]. Neurocomputing, 2013, 101: 229 – 242.

[5] Geng Z Q, Dong J G, Chen J, et al. A new self-organizing extreme learning machine soft sensor model and its applications in complicated chemical processes[J]. Engineering Applications of Artificial Intelligence, 2017, 62: 38 – 50.

[6] Huang G B, Liang N Y, Rong H J, et al. On-line sequential extreme learning machine[C]//Proceedings of the IASTED International Conference on Computational Intelligence. 2005: 232 – 237.

[7] Huang G B, Chen L. Convex incremental extreme learning machine[J]. Neurocomputing, 2007, 70(16/17/18): 3056 – 3062.

[8] Wang Y G, Cao F L, Yuan Y B. A study on effectiveness of extreme learning machine[J]. Neurocomputing, 2011, 74(16): 2483 – 2490.

[9] Rong H J, Jia Y X, Zhao G S. Aircraft recognition using modular extreme learning machine[J]. Neurocomputing, 2014, 128(5): 166 – 174.

[10] Bai Z, Kasun L L C, Huang G B. Generic object recognition with local receptive fields based extreme learning machine[J]. Procedia Computer Science, 2015, 53: 391 – 399.

[11] Liu X W, Wang L, Huang G B, et al. Multiple kernel extreme learning machine[J]. Neurocomputing, 2015, 149: 253 – 264.

[12] Wang D, Wang P, Ji Y. An oscillation bound of the generalization performance of extreme learning machine and corresponding analysis[J]. Neurocomputing, 2015, 51: 883 – 890.

[13] Deng W Y, Bai Z, Huang G B, et al. A fast SVD-hidden-nodes based extreme learning machine for large-scale data analytics[J]. Neural Networks, 2016, 77: 14 – 28.

[14] Xu Z X, Yao M, Wu Z W, et al. Incremental regularized extreme learning machine and it's enhancement[J]. Neurocomputing, 2016, 174: 134 – 142.

[15] Wang M J, Chen H L, Li H Z, et al. Grey wolf optimization evolving kernel extreme learning machine: application to bankruptcy prediction[J]. Engineering Applications of Artificial Intelligence, 2017, 63: 54 – 68.

[16] Wang Z L, Wu D H, Gravina R, et al. Kernel fusion based extreme learning machine for cross-location activity recognition[J]. Information Fusion, 2017, 37: 1 – 9.

[17] Xu S L, Wang J H. Dynamic extreme learning machine for

data stream classification[J]. Neurocomputing,2017,238:433 – 449.

[18] Zeng N Y,Zhang H,Liu W B,et al. A switching delayed PSO optimized extreme learning machine for short-term load forecasting[J]. Neurocomputing,2017,240:175 – 182.

[19] Yang J C,Jiao Y B,Xiong N X,et al. Fast face gender recognition by using local ternary pattern and extreme learning machine[J]. KSII Transactions on Internet and Information Systems,2013,7(7):1705 – 1720.

[20] Huang W M,Li N,Lin Z P,et al. Liver tumor detection and segmentation using kernel-based extreme learning machine[C]//2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2013:3662 – 3665.

[21] Chen C,Li W,Su H J,et al. Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine[J]. Remote Sensing,2014,6(6):5795 – 5814.

[22] Cheng X Y,Liu H P,Xu X Y,et al. Denoising deep extreme learning machine for sparse representation[J]. Memetic Computing,2016,9(3):199 – 212.

[23] Maliha A,Yusof R,Shapiai M I. Extreme learning machine for structured output spaces[J]. Neural Computing and Applications,2018,30(4):1251 – 1264.

[24] Ding S F,Guo L L,Hou Y L. Extreme learning machine with kernel model based on deep learning[J]. Neural Computing and Applications,2017,28(8):1975 – 1984.

[25] Song H,Qin A K,Salim F D. Multivariate electricity consumption prediction with extreme learning machine[C]// International Joint Conference on Neural Networks. 2016.

[26] Wang Z Q,Yu G,Kang Y,et al. Breast tumor detection in digital mammography based on extreme learning machine[J]. Neurocomputing,2014,128:175 – 184.

[27] Ramalho G L B,Reboucas Filho P P,Cortez P C. Lung disease detection using feature extraction and extreme learning machine[J]. Revista Brasileira de Engenharia Biomédica,2014,30(3):207 – 214.

[28] Arunadevi B,Deepa S N. Brain tumor tissue categorization in 3D magnetic resonance images using improved PSO for extreme learning machine[J]. Progress in Electromagnetics Research B,2013,49:31 – 54.

[29] Zhu W Z,Huang W M,Lin Z P,et al. Data and feature mixed ensemble based extreme learning machine for medical object detection and segmentation[J]. Multimedia Tools and Applications,2015,75(5):2815 – 2837.

[30] Sun J H,Xiao Z W,Xie Y X. Automatic multi-fault recognition in TFDS based on convolutional neural network[J]. Neurocomputing,2017,222:127 – 136.

[31] Bai S. Growing random forest on deep convolutional neural networks for scene categorization[J]. Expert Systems with Applications,2017,71:279 – 287.

[32] Ma J L,Wu F,Zhu J,et al. A pre-trained convolutional neural network based method for thyroid nodule diagnosis[J]. Ultrasonics,2017,73:221 – 230.

□

征 稿 启 事

《测控技术》以国防及工业领域中的测试测量与控制技术为核心,涉及多个学科领域,如计算机技术、电子技术、自动控制技术、传感器及仪表技术、网络与通信技术、自动测试技术和虚拟与仿真技术等。本刊开辟的栏目和内容主要有:大家论坛、综述、模式识别(先进算法)与人工智能、机器人技术与应用、网络技术与应用、试验与测试、智能感知与仪器仪表、数据采集与处理、计算机与控制系统、理论专栏、虚拟现实技术、飞行器控制技术、兵器科学测控技术与运用、科技动态等。

欢迎业界专家、教授、中青年学者及工程技术人员投稿,其论文应反映当前国内外测控技术及其应用方面的先进理论、技术成果或发展趋势;具有实际应用背景,并有相应的实验结果;未在国内外期刊上公开发表过的有创新性的内容。具体要求如下:

★ 文题醒目切题,简明扼要,字数不超过 20 字为宜,中英文对照。

★ 给出中英文摘要、关键词。摘要应包括研究目的、方法和过程、结果和结论,准确反映论文要点,在 200 ~ 300 字之间。关键词 4 ~ 6 个,应反映论文的本质、特征,且便于检索。

★ 综述性文章字数要求在 6000 字以内,其他技术论文在 5000 字以内,典型图表一般不超过 8 幅,图表中的英文须译成中文。

★ 给出必要的参考文献,并在文内用上角标标出。新理论、观点应给出出处。文献数量不少于 8 条。

★ 请勿一稿多投。超过 3 个月未收到审稿结果者可投他刊,录用稿编辑部有删改权。录用稿件将酌情收取少量版面费,稿费从优。

★ 来稿请注明所投栏目,若课题受到资助,请注明资助项目名称类别及编号。

★ 投稿一律采用网上在线投稿,不接收邮箱投稿。详情请登录我刊网站。

《测控技术》网站:www. mct. com. cn

通信地址: 北京市亦庄经济技术开发区经海二路 29 号院 9 号楼 7 层《测控技术》杂志社 **邮编:** 101111

联系电话: 010 – 65667497,65665486,65665345