

DOI:10.19651/j.cnki.emt.2210187

含类信息的极限学习机自动编码器特征学习方法^{*}程 蓉 白艳萍 胡红萍 谭秀辉 续 婷
(中北大学理学院 太原 030051)

摘要: 极限学习机自动编码器(ELM-AE)将极限学习机(ELM)技术与自动编码器(AE)结合,可以无监督学习数据特征且克服了参数迭代调整的昂贵时间消耗。然而,以最小化重构误差为目标的 ELM-AE 并不能有效利用分类问题中的数据类别信息,导致特征的类别可分性较差。针对此现象,本文提出一种面向数据分类的含类信息极限学习机自编码(CELM-AE)特征学习方法,该方法将投影特征向量的类间离散度与类内相似度限制到 ELM-AE 的目标函数中,且可通过解析算法求得更具类别分辨力的最优数据表示。对 6 种 UCI 数据集分别使用基于 CELM-AE、ELM-AE 和 AE 的特征表示进行分类实验,结果表明,CELM-AE 得到的数据特征在两种分类器(ELM/KNN)下的分类精度与稳定性表现均优于 ELM-AE 与 AE,且时间代价很小,说明了 CELM-AE 在提取可分性数据特征表示方面的优势。

关键词: 极限学习机;自动编码器;特征学习;数据分类

中图分类号: TP3 **文献标识码:** A **国家标准学科分类代码:** 520.20

Feature learning method of extreme learning machine auto-encoder
with category informationCheng Rong Bai Yanping Hu Hongping Tan Xiuhui Xu Ting
(School of Science, North University of China, Taiyuan 030051, China)

Abstract: Extreme learning machine auto-encoder (ELM-AE) combines extreme learning machine (ELM) technology with auto-encoder (AE), which can learn data features unsupervised and overcome the expensive time consumption of parameter iterative adjustment. However, ELM-AE, which aims to minimize reconstruction errors, cannot effectively use the data category information in classification problems, resulting in features with poor category separability. In view of this phenomenon, this paper proposes a data classification-oriented feature learning method of extreme learning machine auto-encoder with category information (CELM-AE), which limits the inter class dispersion and intra class similarity of the projected feature vector to the objective function of ELM-AE, and can obtain the optimal data representation with more class resolution through analytical algorithm. The classification experiments of 6 UCI data sets are carried out using the feature representation based on CELM-AE, ELM-AE and AE respectively. The results show that the classification accuracy and stability of the data features obtained by CELM-AE under the two classifiers (ELM/KNN) are better than ELM-AE and AE, and the time cost is very small, which shows the advantages of CELM-AE in extracting the separable feature representation of data.

Keywords: extreme learning machine; auto-encoder; feature learning; data classification

0 引 言

机器学习算法的泛化能力取决于数据集,因此合理设计数据集的特征来表示数据的显著结构就显得十分重要。特征学习或表征学习^[1]是一种能够从训练数据中自动提取分类或其他特定任务所需的有效表征的技术。在各种特征

表示学习方法中,自动编码器(auto-encoder, AE)^[2]是一种前馈神经网络(包括输入层、隐藏层和输出层),它以无监督的方式学习数据表示。为了构建更深层次的网络结构,将多个 AE 进行堆叠的网络设计被提出,称为堆栈式自编码(stacked auto-encoder, SAE)^[3]。AE 或 SAE 的学习训练采用的是误差反向传播方式,当面向分类任务时,只需在自

收稿日期:2022-06-01

^{*} 基金项目:山西省基础研究计划资助项目(202103021223189, 202103021224195, 202103021224212, 20210302123019)、国家自然科学基金(61774137)、山西省回国留学人员科研项目(2020-104, 2021-108)资助

编码结构的顶层添加一个输出层,然后采用基于梯度下降的策略,对整个系统的参数(包括在 SAE 中构建新数据表示的参数)进行调整,从而最小化分类误差。换句话说,分类错误被反向传播到 SAE,以指导各层 AE 构建更具代表性和区分性的特征,从而实现更好的分类。但由于基于误差方向传播的训练方法非常耗时,且容易陷入局部极值等,AE 的训练方式限制了其在实际应用中的效果。

Huang 等^[4]于 2006 年提出了基于单层前馈网络(single-layer feedforward network, SLFN)的极限学习机(extreme learning machine, ELM)技术,该方法的关键原理是从输入层到隐藏层随机选择权重,然后通过最小化损失函数解析地确定 SLFN 的输出权重,因此 ELM 比传统 SLFN 的训练算法在回归与分类时更为简单高效。自那时以来,大量理论和实验证据证明了 ELM 的多重优势,如快速训练、强大的泛化能力和通用逼近能力^[5]。到目前为止,大量的 ELM 已成功应用于许多领域,例如模式分类^[6-8]、图像重建^[9]、图像分割^[10-11]、预测^[12-13]、聚类^[14-15]等。2013 年, Kasun 等^[16]首先将极限学习机技术用于数据自编码,提出了 ELM-AE 模型,该方法借助 ELM 算法的优势,能够通过最小化输入训练数据的重建误差,有效地学习数据表示,并将其推广为多层结构(multilayer extreme learning machine, ML-ELM),在 MNIST 数据集上取得了优于其它算法的特征表示分类效果,进而引起了相关学术界的关注。之后几年,ELM-AE 被应用于声信号分类^[17]、缺失数据插补^[18]、数据流分类^[19]、光伏功率预测^[20]、故障检测^[21]等多种领域,同时 ELM-AE 的多种变形与改进方法也被相继提出,如 Wong 等^[22]受到核学习的启发,提出了核极限学习机自编码器(kernel-based extreme learning machine auto-encoder, KELM-AE),消除了 ELM-AE 隐层节点数量的影响。Sun 等^[23]考虑到数据流形结构,将流形正则化加入 ELM-AE 的目标函数中,提出了图正则化极限学习机自动编码器(generalized extreme learning machine auto-encoder, GELM-AE),使其具有更强的能力以提取适合聚类的特征。陈晓云等^[24]将子空间结构学习与 ELM-AE 结合,提出可以保持数据子空间结构的极限学习机自动编码器,在数据聚类方面取得良好效果。Liu 等^[25]提出了类约束极限学习机(class-constrained extreme learning machine, C2ELM),对不同类别数据分别训练 ELM-AE,将所有解码权重组合为投影变换矩阵,并验证了该方法的有效性。Chang 等^[26]提出了一种简单、有效、快速的深层结构 ELMAENet,使用 ELM-AE 来获得卷积层滤波器,不再需要参数调整,但仍具有良好的图像分类性能。

基本 ELM-AE 与 ML-ELM 在特征提取阶段采用的都是完全无监督的训练方式,即通过最小化重构误差获得数据的特征表示。由于其沿袭了 ELM 的学习方式,各编码层的权重不随训练而调整,只根据最小二乘法求得解码权重,使其能够保持快速学习的优势,但这同时也使 ELM-

AE 不能像传统 AE 一样利用增加有监督的分类输出层去反向调整自动编码阶段的权重,从而获得更具有代表性及可分性的数据特征。为了保持 ELM-AE 对特征进行快速学习的特点,同时又获得更具类别代表性的数据特征而提高数据分类精度,本文提出了一种将数据类别信息加入到 ELM-AE 中的数据表示方法,即含类信息的极限学习机自动编码器(extreme learning machine with category information, CELM-AE),该方法将投影数据的类间离散度与类内相似度限制到 ELM-AE 的目标函数中去,并通过解析算法获得最优数据表示,保持了运算快速优势的同时获得了相较于原算法更具类别代表性的特征表示。

1 相关工作

本节将简要介绍 ELM 与 ELM-AE,并基于此引出下节经改进的含类别信息的极限学习机自动编码器。

1.1 极限学习机(ELM)

ELM 由 Huang 等^[27]于 2006 年提出,是一种基于 SLFN 的高效学习算法。ELM 理论表明,尽管 SLFN 中的隐藏节点起着关键作用,但它们不需要调整,只要这些神经元是非线性分段连续且随机生成的,它们就具有普适逼近能力^[4-5,27-28]。这自然克服了许多传统学习算法(如 BP)遇到的学习速度慢,容易得到局部最优解等相关的学习瓶颈和障碍。

如图 1 所示,ELM 网络结构的网络输出如下:

$$f_L(x) = \sum_{i=1}^L \beta_i^T h_i(x) = \mathbf{h}(x)\boldsymbol{\beta} \quad (1)$$

其中, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_L)^T$ 是隐层节点与输出节点间的权重矩阵, $\mathbf{h}(x) = (h_1(x), h_2(x), \dots, h_L(x))$ 是隐层输出向量。设激活函数为 $g(x)$, 则 $h_i(x) = g(\mathbf{a}_i \cdot \mathbf{x} + b_i)$, 其中的权重向量 \mathbf{a}_i 与偏置 b_i 通过随机方式生成并在训练中保持不变。从而在 ELM 网络中,输入向量 \mathbf{x} 被投影为 L 维 ELM 随机特征 $\mathbf{h}(x)$, 当给定 n 个独立的训练样本 $\{(\mathbf{x}_i, \mathbf{t}_i) \mid i = 1, 2, \dots, n\}$ 时,考虑到算法的泛化性能与鲁棒性,ELM 解决的是如下优化问题:

$$\text{Minimize: } \|\boldsymbol{\beta}\|^2 + \frac{C}{2} \|\mathbf{T} - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}\|^2 \quad (2)$$

其中,

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}(x_1) \\ \mathbf{h}(x_2) \\ \vdots \\ \mathbf{h}(x_n) \end{pmatrix} = \begin{pmatrix} g(\mathbf{a}_1 \cdot \mathbf{x}_1 + b_1) & g(\mathbf{a}_2 \cdot \mathbf{x}_1 + b_2) & \cdots & g(\mathbf{a}_L \cdot \mathbf{x}_1 + b_L) \\ g(\mathbf{a}_1 \cdot \mathbf{x}_2 + b_1) & g(\mathbf{a}_2 \cdot \mathbf{x}_2 + b_2) & \cdots & g(\mathbf{a}_L \cdot \mathbf{x}_2 + b_L) \\ \vdots & \vdots & \ddots & \vdots \\ g(\mathbf{a}_1 \cdot \mathbf{x}_n + b_1) & g(\mathbf{a}_2 \cdot \mathbf{x}_n + b_2) & \cdots & g(\mathbf{a}_L \cdot \mathbf{x}_n + b_L) \end{pmatrix}$$

代表 ELM 网络隐层神经元的输出矩阵, $\mathbf{T} = (\mathbf{t}_1^T, \mathbf{t}_2^T, \dots, \mathbf{t}_n^T)^T$ 为目标输出矩阵。

由目标函数(2)可得最优输出权值的解析解为:

$$\boldsymbol{\beta} = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T} \quad (3)$$

由此可知,对比传统前向网络的训练方式,ELM 具有更加快速的学习速度,且可证明其具有通用逼近与分类能力。

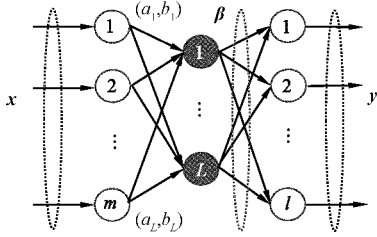


图 1 ELM 结构

1.2 极限学习机自动编码器(ELM-AE)

ELM-AE 具有严格轴对称结构,如图 2 所示,不同于捆绑权重自动编码器(tied weight auto-encoder, TAE),ELM-AE 并未采用捆绑权重,且由于只有解码阶段的权重需要计算,其计算开销远小于 TAE。而与 ELM 相比,ELM-AE 采用的是将网络输出数据 \mathbf{T} 设置为输入数据 \mathbf{X} 的无监督学习方式,其目标是最小化输入 \mathbf{X} 的重建误差,从而获得原数据的特征表示,而输出层权值的计算则与 ELM 相似。Kasun 等^[29]曾指出 AE 的隐层需要保持输入数据的信息,因此 ELM-AE 采用正交随机参数来保持输入数据的欧几里得信息,其隐层节点输出具体计算如下:

$$\begin{cases} \mathbf{h}(\mathbf{x}) = g(\mathbf{x}\mathbf{A} + \mathbf{b}) = (g(\mathbf{a}_1 \cdot \mathbf{x} + b_1), \dots, g(\mathbf{a}_L \cdot \mathbf{x} + b_L)) \\ \mathbf{A}^T \mathbf{A} = \mathbf{I}, \mathbf{b}\mathbf{b}^T = 1, \quad L < m \\ \mathbf{A}\mathbf{A}^T = \mathbf{I}, \mathbf{b}\mathbf{b}^T = 1, \quad L > m \end{cases} \quad (4)$$

其中, $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_L)$ 为输入节点与隐层节点间的正交随机权值矩阵, $\mathbf{b} = (b_1, b_2, \dots, b_L)$ 为隐层节点的偏置向量。从而,ELM-AE 的优化目标函数为:

$$\text{Minimize: } \|\boldsymbol{\beta}\|^2 + \frac{C}{2} \|\mathbf{X} - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}\|^2 \quad (5)$$

对 $\boldsymbol{\beta}$ 求导,并令其为 0,可得最优输出权值解矩阵 $\boldsymbol{\beta}$ 为:

$$\boldsymbol{\beta} = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{X} \quad (6)$$

输入数据 \mathbf{X} 通过 ELM-AE 的解码权值 $\boldsymbol{\beta}$ 得到新的特征表示:

$$\mathbf{X}_{proj} = \mathbf{X}\boldsymbol{\beta}^T \quad (7)$$

2 含类信息的极限学习机自动编码器(CELM-AE)

从 ELM-AE 的设计原理可以看出,基于极限学习方式的自动编码器具有学习速度快,能够计算解析最优解的优点。然而,由于 ELM-AE 的特征表示只考虑对输入数据 \mathbf{X} 的重建,而与数据本身的类别标签信息无关,所以仅基于目标函数式(5)提取的数据特征并不能保证在数据分类任务中获得令人满意的效果。为了提取更具代表性和区分性的

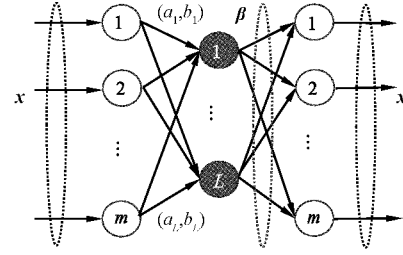


图 2 ELM-AE 结构

数据表示,面向分类任务,目标函数中应该增加已有数据的类别信息,从而帮助提高对未知数据的分类准确率。具体来说,CELM-AE 就是将带标签样本投影向量的类间离散度与类内相似度同时限制到学习任务的目标函数中去,以期获得更有利于分类任务的特征表示投影变换矩阵。

2.1 模型提出

设原样本空间的均值(中心)向量为:

$$\mathbf{m}_k = \frac{1}{N} \sum_{\mathbf{x} \in X_k} \mathbf{x}, \quad (k = 1, \dots, K) \quad (8)$$

则投影后的样本均值为:

$$\begin{aligned} \tilde{\mathbf{m}}_k &= \frac{1}{N} \sum_{\mathbf{x}_{proj} \in X_k} \mathbf{x}_{proj} = \frac{1}{N} \sum_{\mathbf{x} \in X_k} \mathbf{x}\boldsymbol{\beta}^T = \mathbf{m}_k \boldsymbol{\beta}^T, \\ (k &= 1, \dots, K) \end{aligned} \quad (9)$$

CELM-AE 的目标函数具体表达式如下:

$$\begin{aligned} \text{Minimize: } F &= \|\boldsymbol{\beta}\|^2 + \frac{C}{2} \|\mathbf{X} - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}\|^2 - \\ &\frac{\lambda_1}{2} \sum_{k \neq l} \|\tilde{\mathbf{m}}_k - \tilde{\mathbf{m}}_l\|^2 + \frac{\lambda_2}{2} \sum_{k=1}^K \sum_{i=1}^{N_k} \|\mathbf{x}_{ki}\boldsymbol{\beta}^T - \tilde{\mathbf{m}}_k\|^2 \end{aligned} \quad (10)$$

其中,第 1 项用于控制模型复杂度,第 2 项为模型的误差损失,第 3 项为各类投影数据的均值向量间距离和,反映投影数据总类间离散程度,第 4 项为各类投影数据与均值向量的总距离和,反映投影数据总类内相似程度。 K 表示类别数, N_k 表示第 k 类中的数据样本个数, \mathbf{x}_{ki} 表示第 k 类数据中的第 i 个样本, C, λ_1, λ_2 为正化因子。

2.2 模型求解

为了得到使目标函数达到最小的权值矩阵 $\boldsymbol{\beta}$,将目标函数 F 对 $\boldsymbol{\beta}$ 求导数,并令其为 0,即可得最优解。

令 $f_1 = \|\boldsymbol{\beta}\|^2 + \frac{C}{2} \|\mathbf{X} - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}\|^2$,则 f_1 可等价地写为:

$$f_1 = \text{tr}(\boldsymbol{\beta}^T \boldsymbol{\beta}) + \frac{C}{2} \text{tr}[(\mathbf{X}^T - \boldsymbol{\beta}^T \mathbf{H}(\mathbf{X})^T)(\mathbf{X} - \mathbf{H}(\mathbf{X})\boldsymbol{\beta})] \quad (11)$$

f_1 对 $\boldsymbol{\beta}$ 求导,得:

$$\frac{df_1}{d\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{C}\mathbf{H}(\mathbf{X})^T \mathbf{H}(\mathbf{X})\boldsymbol{\beta} - \mathbf{C}\mathbf{H}(\mathbf{X})^T \mathbf{X} \quad (12)$$

令 $f_2 = \frac{\lambda_1}{2} \sum_{k \neq l} \|\tilde{\mathbf{m}}_k - \tilde{\mathbf{m}}_l\|^2 = \frac{\lambda_1}{2} \sum_{k \neq l} \|(\mathbf{m}_k - \mathbf{m}_l)\boldsymbol{\beta}^T\|^2$,则 f_2 可等价地写为:

$$f_2 = \frac{\lambda_1}{2} \sum_{k \neq l} \text{tr}(\boldsymbol{\beta}(\mathbf{m}_k - \mathbf{m}_l)^T(\mathbf{m}_k - \mathbf{m}_l)\boldsymbol{\beta}^T) = \frac{\lambda_1}{2} \sum_{k \neq l} \text{tr}(\boldsymbol{\beta} \mathbf{S}_{kl} \boldsymbol{\beta}^T) \quad (13)$$

f_2 对 $\boldsymbol{\beta}$ 求导,得:

$$\frac{df_2}{d\boldsymbol{\beta}} = \frac{\lambda_1}{2} \sum_{k \neq l} \boldsymbol{\beta}(\mathbf{S}_{kl} + \mathbf{S}_{kl}^T) = \lambda_1 \boldsymbol{\beta} \sum_{k \neq l} \mathbf{S}_{kl} \quad (14)$$

其中,

$$\sum_{k \neq l} \mathbf{S}_{kl} = \sum_{k \neq l} (\mathbf{m}_k - \mathbf{m}_l)^T(\mathbf{m}_k - \mathbf{m}_l), (k, l = 1, \dots, K) \quad (15)$$

为原数据总类间散度矩阵,刻画原数据各类的类间离散程度。

$$\text{令 } f_3 = \frac{\lambda_2}{2} \sum_{k=1}^K \sum_{i=1}^{N_k} \| \mathbf{x}_{ki} \boldsymbol{\beta}^T - \tilde{\mathbf{m}}_k \|^2 =$$

$$\frac{\lambda_2}{2} \sum_{k=1}^K \sum_{i=1}^{N_k} \| \mathbf{x}_{ki} \boldsymbol{\beta}^T - \mathbf{m}_k \|^2,$$

则 f_3 可等价地写为:

$$f_3 = \frac{\lambda_2}{2} \sum_{k=1}^K \sum_{i=1}^{N_k} \text{tr}(\boldsymbol{\beta}(\mathbf{x}_{ki} - \mathbf{m}_k)^T(\mathbf{x}_{ki} - \mathbf{m}_k)\boldsymbol{\beta}^T) =$$

$$\frac{\lambda_2}{2} \sum_{k=1}^K \sum_{i=1}^{N_k} \text{tr}(\boldsymbol{\beta} \tilde{\mathbf{S}}_{ki} \boldsymbol{\beta}^T) \quad (16)$$

f_3 对 $\boldsymbol{\beta}$ 求导,得:

$$\frac{df_3}{d\boldsymbol{\beta}} = \frac{\lambda_2}{2} \boldsymbol{\beta} \sum_{k=1}^K \sum_{i=1}^{N_k} (\tilde{\mathbf{S}}_{ki} + \tilde{\mathbf{S}}_{ki}^T) = \lambda_2 \boldsymbol{\beta} \sum_{k=1}^K \sum_{i=1}^{N_k} \tilde{\mathbf{S}}_{ki} \quad (17)$$

其中,

$$\sum_{k=1}^K \sum_{i=1}^{N_k} \tilde{\mathbf{S}}_{ki} = \sum_{k=1}^K \sum_{i=1}^{N_k} (\mathbf{x}_{ki} - \mathbf{m}_k)^T(\mathbf{x}_{ki} - \mathbf{m}_k) \quad (k = 1, \dots, K, i = 1, \dots, N_k) \quad (18)$$

为原数据总类内散度矩阵,刻画原数据各类的类内离散程度。

综合式(12)、(14)和(17),可得:

$$\frac{dF}{d\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{CH}(\mathbf{X})^T \mathbf{H}(\mathbf{X}) \boldsymbol{\beta} - \mathbf{CH}(\mathbf{X})^T \mathbf{X} -$$

$$\lambda_1 \boldsymbol{\beta} \sum_{k \neq l} \mathbf{S}_{kl} + \lambda_2 \boldsymbol{\beta} \sum_{k=1}^K \sum_{i=1}^{N_k} \tilde{\mathbf{S}}_{ki} \quad (19)$$

令 $\frac{dF}{d\boldsymbol{\beta}} = 0$, 得:

$$(\mathbf{I} + \mathbf{CH}(\mathbf{X})^T \mathbf{H}(\mathbf{X})) \boldsymbol{\beta} + \boldsymbol{\beta} (\lambda_2 \sum_{k=1}^K \sum_{i=1}^{N_k} \tilde{\mathbf{S}}_{ki} - \lambda_1 \sum_{k \neq l} \mathbf{S}_{kl}) =$$

$$\mathbf{CH}(\mathbf{X})^T \mathbf{X} \quad (20)$$

式(20)为西尔维斯特(Sylvester)矩阵方程,借助矩阵 Kronecker 积的朴素求解法会增加方程阶数从而提高计算复杂度,在数据特征维数较高时并不适用。可采用经典 Bartels-Stewart 方法求解,将式(20)中的系数矩阵 $\mathbf{I} +$

$\mathbf{CH}(\mathbf{X})^T \mathbf{H}(\mathbf{X})$ 与 $\lambda_2 \sum_{k=1}^K \sum_{i=1}^{N_k} \tilde{\mathbf{S}}_{ki} - \lambda_1 \sum_{k \neq l} \mathbf{S}_{kl}$ 进行 Schur 分解,即:

$$\begin{cases} \mathbf{I} + \mathbf{CH}(\mathbf{X})^T \mathbf{H}(\mathbf{X}) = \mathbf{UR}_1\mathbf{U}^T \\ \lambda_2 \sum_{k=1}^K \sum_{i=1}^{N_k} \tilde{\mathbf{S}}_{ki} - \lambda_1 \sum_{k \neq l} \mathbf{S}_{kl} = \mathbf{VR}_2\mathbf{V}^T \end{cases} \quad (21)$$

则式(20)可变化为三角形矩阵方程 $\mathbf{R}_1\mathbf{B} + \mathbf{BR}_2 = \mathbf{D}$,

其中 $\mathbf{B} = \mathbf{U}^T \boldsymbol{\beta} \mathbf{V}$, $\mathbf{D} = \mathbf{U}^T (\mathbf{CH}(\mathbf{X})^T \mathbf{X}) \mathbf{V}$ 。

记 $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m)$, $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m)$, 其中 $\mathbf{b}_k, \mathbf{d}_k \in R^L$, $\mathbf{R}_2 = (r_{ij}^{(2)})$, $i, j = 1, \dots, m$ 。则可得 \mathbf{B} 的列递推公式:

$$(\mathbf{R}_1 + r_{kk}^{(2)} \mathbf{I}) \mathbf{b}_k = \mathbf{d}_k - \sum_{i=1}^{k-1} \mathbf{b}_i r_{ik}^{(2)}, (k = 1, \dots, m) \quad (22)$$

通过向前代换依次求解,可得矩阵 \mathbf{B} 的所有列向量 \mathbf{b}_k , 再通过:

$$\boldsymbol{\beta} = \mathbf{UBV}^T \quad (23)$$

即可求出式(20)的解矩阵 $\boldsymbol{\beta}$, 从而得到目标函数 F 的最优解。

基于以上讨论,CELM-AE 方法的总体框架如图 3 所示,其具体算法步骤总结如算法 1 所示。

算法 1:CELM-AE 特征学习算法

输入:原始数据矩阵 $\mathbf{X}_{n \times m}$, 投影后的特征表示维数 L , 正则化因子 $C > 0, \lambda_1 > 0, \lambda_2 > 0$

输出:含类别信息的数据特征表示矩阵 \mathbf{X}_{proj}

步骤 1)根据式(4)随机初始化正交权值矩阵 \mathbf{A} 与偏置向量 \mathbf{b} ;

步骤 2)计算 CELM-AE 的隐层输出矩阵 $\mathbf{H}(\mathbf{X})_{n \times L}$;

步骤 3)结合样本类标签信息,根据式(15)和(18)计算样本总类间散度矩阵 $\sum_{k \neq l} \mathbf{S}_{kl}$ 和总类内散度矩阵 $\sum_{k=1}^K \sum_{i=1}^{N_k} \tilde{\mathbf{S}}_{ki}$;

步骤 4)确定西尔维斯特矩阵方程式(20),由 Bartels-Stewart 方法求出解矩阵 $\boldsymbol{\beta}$ (式(23)),则 $\mathbf{X}_{n \times m}$ 的含样本类别信息的数据特征表示矩阵 $\mathbf{X}_{proj} = \mathbf{X}\boldsymbol{\beta}^T$ 。

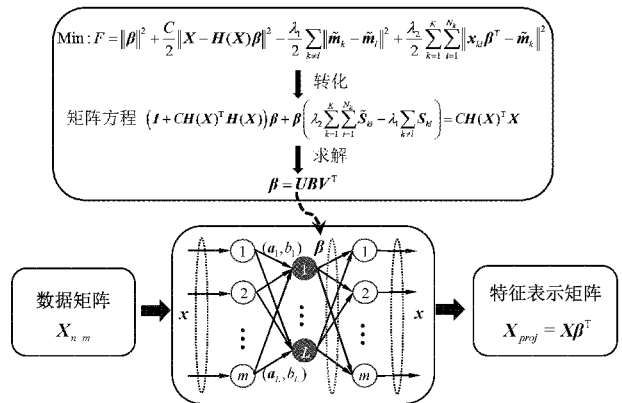


图 3 CELM-AE 总体框架

3 实验结果与分析

本节将对 CELM-AE 关于数据表示的区分能力进行检

验,并采用两种分类网络分别对此进行实验验证,并与 ELM-AE 及 AE 方法进行对比分析。

3.1 数据说明与实验设置

实验选取了 6 个 UCI 公开数据集 (Iris, Wine, Image Segmentation, Vehicle Silhouettes, Breast Cancer Wisconsin(Diagnostic), Spambase), 具体的数据集说明如表 1 所示。

表 1 数据集说明

数据集	维数	样本数	类别数
Iris	4	150	3
Wine	13	178	3
Segmentation	19	2 310	7
Vehicle	18	846	4
Breast	30	569	2
Spambase	57	4 601	2

所有实验全部随机选取数据集的 70% 作为训练集 \mathbf{X}_{train} , 30% 作为测试集 \mathbf{X}_{test} , 并为减少因数据选取随机性产生的差异, 采用 100 次独立实验的平均分类表现作为该数据集上的最终分类评估依据。编码器激活函数采用 sigmoid 函数, 输出向量采用 One-Hot 二进制编码方式。目标函数 F 中的正则化因子在 3.2 节的实验中全部设置为 1 ($C = 1, \lambda_1 = 1, \lambda_2 = 1$), 也即在目标函数中同时考虑投影特征向量的类间离散度与类内相似度两个损失项, 对其中一种损失项的消融实验将在 3.3 节中进行并给出相应分析。

3.2 分类实验与结果分析

1) 基于 ELM 分类器的对比实验

首先采用 ELM 网络分类器对 CELM-AE 的数据表示区分能力进行检验。当设置了数据集在两种编码器下的投影向量维数 L 后, 为了避免 ELM 隐层节点个数在训练分类器时对分类结果的进一步影响, 实验直接将 CELM-AE

的投影向量 $\mathbf{X}_{proj} = \mathbf{X}\boldsymbol{\beta}^T$ 作为 ELM 的隐层输入向量, 如图 4 所示。故 CELM-AE+ELM 的数据分类实验步骤为: 首先, 根据训练集 \mathbf{X}_{train} 与分类目标 \mathbf{T}_{train} 中包含的类别信息, 利用 CELM-AE 算法得到解码权值 $\boldsymbol{\beta}$; 其次, 将自编码的解码权值 $\boldsymbol{\beta}^T$ 作为 ELM 网络的输入权值, 且不设置随机偏差向量, 并使用式 (24) 确定 ELM 分类器的输出层权值 $\boldsymbol{\beta}_1$ (实验中将 C_1 设置为 10^8); 最后, 将测试集 \mathbf{X}_{test} 输入 ELM 网络, 得到相应分类结果并与目标 \mathbf{T}_{test} 对比得出分类准确率。

$$\boldsymbol{\beta}_1 = \left(\frac{\mathbf{I}}{C_1} + \mathbf{G}(\mathbf{X}_{proj})^T \mathbf{G}(\mathbf{X}_{proj}) \right)^{-1} \mathbf{G}(\mathbf{X}_{proj})^T \mathbf{T}_{train} \quad (24)$$

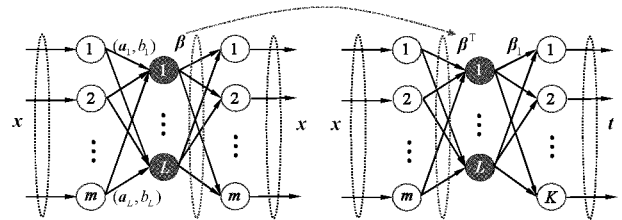


图 4 CELM-AE+ELM 结构图

表 2 给出了在某个特定投影特征维数 L 时, 3 种自编码算法 (CSELM-AE, ELM-AE, AE) 下的 ELM 分类精度 (100 次独立实验的平均正确率), 多次实验分类正确率的标准差, 及每次实验的平均耗时统计对比。从表 2 中可以看出, 在相同的特定投影特征维数 L 前提下, 基于 CSELM-AE 与 ELM-AE 特征学习算法的分类实验耗时远小于基于 AE 算法的分类实验, 这是由于 AE 算法在特征学习时使用的是基于梯度下降的反向传播学习方式, 而 CSELM-AE 与 ELM-AE 则使用解析方法优化, 可以更高效地学习到有效特征表示。此外, 通过比较平均分类正确率与标准差指标可以看出, 基于 CSELM-AE 的分类正确率在 6 种不同数据集上均高于基于 ELM-AE 与 AE 的分类表现, 且实验结果更为稳定, 说明加入数据类别信息的 ELM 自编码对数据分类效果起到了明显的积极作用。

表 2 3 种编码算法在 ELM 分类器下的分类表现

数据集	CELM-AE+ELM			ELM-AE+ELM			AE+ELM			投影维数
	正确率	标准差	运行时间/s	正确率	标准差	运行时间/s	正确率	标准差	运行时间/s	
Iris	0.953 6	0.032 7	0.165 3	0.898 9	0.035 1	0.152 6	0.921 1	0.043 4	3.224 7	$L = 20$
Wine	0.980 9	0.019 6	0.163 2	0.972 8	0.020 3	0.152 2	0.971 5	0.023 3	3.608 3	$L = 20$
Segmentation	0.928 7	0.009 9	0.181 1	0.925 3	0.010 1	0.156 5	0.924 8	0.009 4	10.117	$L = 60$
Vehicle	0.805 4	0.022 5	0.173 2	0.784 4	0.023 0	0.151 8	0.738 0	0.026 3	7.198 3	$L = 80$
Breast	0.969 3	0.013 0	0.171 4	0.965 3	0.014 6	0.155 5	0.959 5	0.015 9	6.654 5	$L = 80$
Spambase	0.915 3	0.007 3	0.269 0	0.908 8	0.007 7	0.183 6	0.871 1	0.013 9	35.312	$L = 120$

另外, 图 5 给出了两种高效率自编码算法对 6 种数据集在设置不同投影特征维数 L 时的平均分类正确率 (100 次实验) 对比曲线, 说明对维数变化的特征表达 \mathbf{X}_{proj} , CELM-AE 的数据区分能力均高于 ELM-AE。为了更直

观地比较 CELM-AE、ELM-AE 与 AE 在数据区分能力方面的差别, 图 6 绘出了 Iris 数据集中相同的测试数据利用 3 种不同编码算法投影到 $L = 2$ 维空间后的图像, 并标注了相应算法的特征学习时间, 可以看出 CELM-AE 的学习

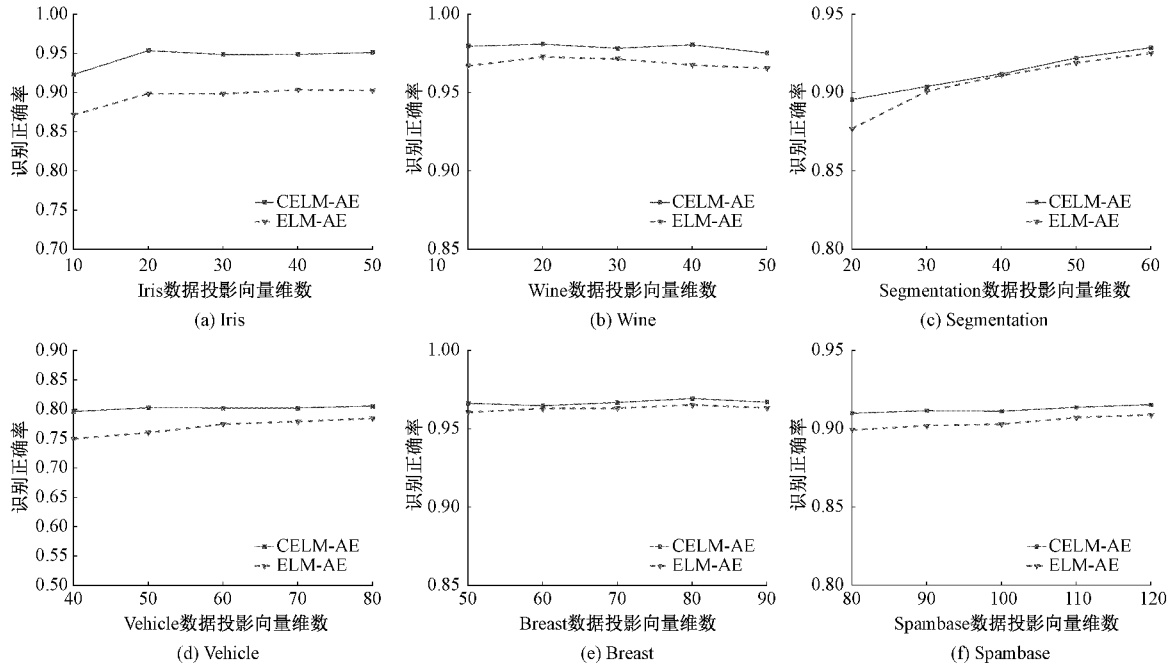


图 5 不同投影向量维数时的 ELM 分类比较

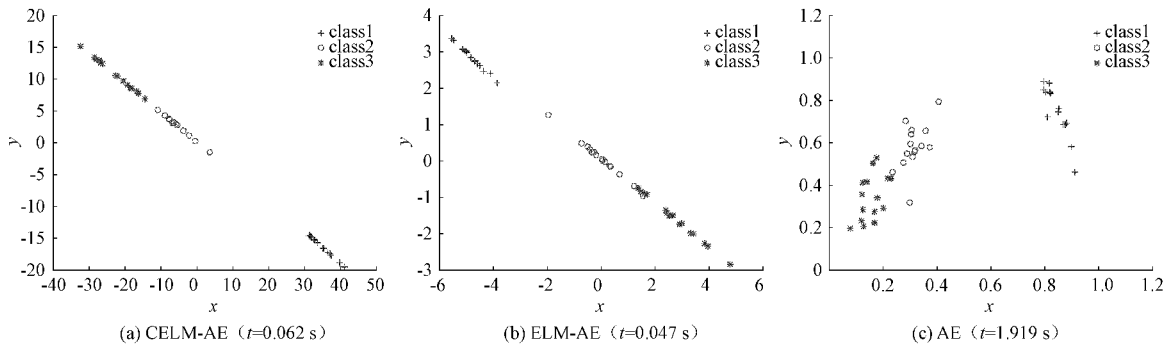


图 6 可视化的二维投影数据与特征学习时间对比 (t 为特征学习时间)

时间消耗虽略高于 ELM-AE,但仍然十分快速,且相较于 ELM-AE 和 AE,基于 CELM-AE 的投影测试数据具有更加明显的可分性。

2) 基于 KNN 分类器的对比实验

为了进一步验证 CELM-AE 相比于 ELM-AE 及 AE 在分类特征表示方面的优势,我们又选择了 K 最近邻分类算法(k-nearest-neighbours, KNN)^[30]对 6 类数据集在特定维数下的投影特征进行分类,即首先在给定投影维数 L 后,根据训练集 X_{train} 与分类目标 T_{train} 中包含的类别信息,利用 CELM-AE 算法得到解码权值 β ,并得出 X_{train} 的数据表示 $X_{proj}^{tr} = X_{train}\beta^T$;其次,计算测试集 X_{test} 的数据表示 $X_{proj}^{te} = X_{test}\beta^T$;最后,根据训练数据特征 X_{proj}^{tr} 与相应类别标签 T_{train} ,通过 KNN 算法确定测试数据在特征表示 X_{proj}^{te} 下的分类结果,并与目标 T_{test} 对比得出分类正确率。

表 3 给出了特定数据表示维数下 6 种数据集在 100 次独立分类实验中的平均分类正确率,多次实验分类正确率

的标准差,及每次实验的平均耗时统计。通过平均分类正确率与标准差两项指标的对比,可以看出,经过 100 次独立实验,相比 ELM-AE 与 AE 算法,CELM-AE 确定的数据特征向量通过 KNN 算法获得了更高的分类正确率,且算法稳定性更高。图 7 同样给出了两种高效自编码方法在设置不同投影数据维数 L 时,KNN 算法的平均分类正确率结果对比,可以看出对 6 种不同数据集,CELM-AE+KNN 也都优于 ELM-AE+KNN 的分类表现,进一步说明了 CELM-AE 在面向分类问题方面,具有可确定更具代表性与可分性特征的优势。

3.3 损失项消融实验与结果分析

本节将对 CELM-AE 的目标函数中增加的两个新损失项(即反映投影数据总类间离散程度的损失函数 f_2 与反映投影数据总类内相似程度的损失函数 f_3)对分类效果的影响进行进一步讨论。

从 CELM-AE 的目标函数式(10)中可以看出,正则化

表 3 3 种编码算法在 KNN 分类器下的分类表现

数据集	CELM-AE+KNN			ELM-AE+KNN			AE+KNN			投影 维数
	正确率	标准差	运行时间/s	正确率	标准差	运行时间/s	正确率	标准差	运行时间/s	
Iris	0.967 8	0.022 2	0.565 8	0.946 2	0.026 8	0.548 6	0.935 3	0.034 0	2.948 6	$L = 10$
Wine	0.981 7	0.018 9	0.580 5	0.954 4	0.025 2	0.572 5	0.952 6	0.026 1	4.502 8	$L = 40$
Segmentation	0.968 6	0.005 7	0.603 0	0.963 1	0.007 0	0.589 6	0.959 8	0.007 5	9.092 0	$L = 50$
Vehicle	0.721 8	0.023 5	0.602 8	0.634 8	0.026 6	0.578 0	0.580 0	0.025 4	6.806 9	$L = 80$
Breast	0.964 7	0.011 9	0.590 5	0.954 3	0.014 4	0.579 3	0.954 3	0.015 2	6.506 7	$L = 60$
Spambase	0.908 3	0.006 5	0.763 3	0.894 3	0.007 6	0.688 6	0.803 6	0.012 8	35.858	$L = 120$

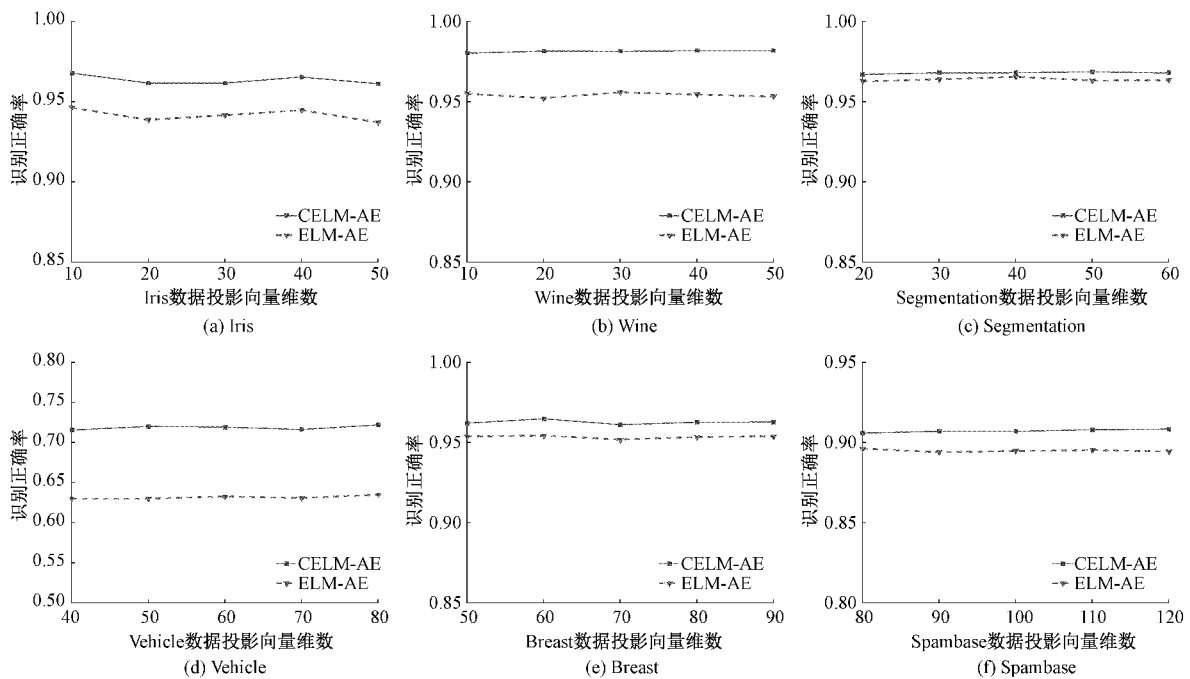


图 7 不同投影向量维数时的 KNN 分类比较

参数 λ_1, λ_2 的取值分别控制了损失项 $(-f_2)$ 与 f_3 的权重。在 3.2 节的分类实验中,将其分别设置为 $\lambda_1 = 1, \lambda_2 = 1$, 说明实验同时考虑了两类损失的影响。本节中我们将结合 ELM 分类器,分别讨论 $\lambda_1 = 1, \lambda_2 = 0$ 与 $\lambda_1 = 0, \lambda_2 = 1$ 这两种参数设置对分类效果的影响,其中 $\lambda_1 = 1, \lambda_2 = 0$ 时说明在目标函数中仅考虑投影数据总类间离散程度的损失,而 $\lambda_1 = 0, \lambda_2 = 1$ 时则说明在目标函数中仅考虑投影数据总类内相似程度的损失。

结合 ELM 分类器,表 4 给出了 6 种数据集通过 3 种参数设置的 CELM-AE 所得投影特征的分类结果对比,统计结果仍为 100 次独立实验的平均正确率与标准差。从实验结果可以看出,若原目标函数中仅增加损失函数

$(-f_2)$ ($\lambda_1 = 1, \lambda_2 = 0$),除 Wine 数据集外,其它数据集的特征分类效果均优于 ELM-AE 所得特征的分类效果;若原目标函数中仅增加损失函数 f_3 ($\lambda_1 = 0, \lambda_2 = 1$),则所有数据集的分类效果均得到有效提升,说明在目标函数中分别增加含类信息的损失函数 $(-f_2)$ 与 f_3 可在一定程度上提升投影特征的可分性。对比所有结果可以看出,在 CELM-AE 的目标函数中同时考虑两个损失项 ($\lambda_1 = 1, \lambda_2 = 1$) 的分类正确率最高,说明在 $(-f_2)$ 与 f_3 的协同限制下,得到具有更高类别分辨力特征的可能性更大。可在今后的研究中进一步考虑正则化参数 λ_1 与 λ_2 的其它设置组合,使 CELM-AE 在分类任务中获得更加优良的表现。

表 4 CELM-AE 损失项对分类效果影响的对比

数据集	CELM-AE+ELM						ELM-AE+ELM		投影维数
	$\lambda_1 = 1, \lambda_2 = 1$		$\lambda_1 = 1, \lambda_2 = 0$		$\lambda_1 = 0, \lambda_2 = 1$		正确率	标准差	
	正确率	标准差	正确率	标准差	正确率	标准差			
Iris	0.953 6	0.032 7	0.936 2	0.033 4	0.901 8	0.034 9	0.898 9	0.035 1	$L = 10$
Wine	0.980 9	0.019 6	0.892 8	0.092 3	0.979 3	0.018 8	0.972 8	0.020 3	$L = 40$
Segmentation	0.928 7	0.009 9	0.926 8	0.014 8	0.926 6	0.008 9	0.925 3	0.010 1	$L = 50$
Vehicle	0.805 4	0.022 5	0.789 1	0.023 0	0.802 0	0.023 1	0.784 4	0.023 0	$L = 80$
Breast	0.969 3	0.013 0	0.967 1	0.012 8	0.967 4	0.013 9	0.965 3	0.014 6	$L = 60$
Spambase	0.915 3	0.007 3	0.908 8	0.006 9	0.913 6	0.007 4	0.908 8	0.007 7	$L = 120$

4 结 论

本文在分析了 ELM-AE 优缺点的基础上,提出了一种含类别信息的极限学习机自动编码特征学习方法:CELM-AE。通过利用原有数据的类别信息,在 ELM-AE 的目标函数中引入投影数据类间离散度与类内相似度的限制,弥补了原算法不能有效利用数据类别信息这一劣势,使其获得更具类别可分性的有效特征表示。此外,提出的 CELM-AE 仍可通过解析方式获得特征投影系数矩阵,避免了传统 AE 对参数进行迭代带来的昂贵时间消耗与不稳定性,其投影特征提取时间虽略高于 ELM-AE,但仍十分高效,且算法稳定性也高于 ELM-AE。通过对 UCI 数据集进行的分类实验,证明了本文提出的 CELM-AE 特征表示算法分别结合两种不同分类器(ELM,KNN)时,在数据分类正确率与稳定性方面均获得了相比 ELM-AE 与 AE 更加优良的分类表现,从而为面向分类任务的数据特征表示提供了可靠有效的新方法。今后我们还将继续研究基于 CELM-AE 的深度堆栈式模型及模型中相关参数的最优设置等问题。

参考文献

- [1] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: A review and new perspectives [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828.
- [2] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [3] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion [J]. Journal of Machine Learning Research, 2010, 11: 3371-3408.
- [4] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1-3): 489-501.
- [5] 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 (Cybernetics), 2011, 42(2): 513-529.
- [6] 魏迪,刘德山,闫德勤,等.应用于人脸图像识别的邻域保持极限学习机 [J]. 计算机工程与应用, 2019, 55(11):187-191.
- [7] TIAN Y M, WANG X T, GENG Y L, et al. Inertial sensor-based human activity recognition via ensemble extreme learning machines optimized by quantum-behaved particle swarm [J]. Journal of Intelligent & Fuzzy Systems, 2020, 38(2): 1443-1453.
- [8] LAL P V, SRILAKSHMI U, VENKATESWARLU D. Face recognition system using extreme learning machines [C]. 2021 5th International Conference on Information Systems and Computer Networks (ISCON), IEEE, 2021: 1-5, DOI: 10.1109/ISCON52037.2021.9702467.
- [9] 张立峰,戴力.基于极限学习机求解正问题的 ECT 图像重建 [J]. 仪器仪表学报, 2021, 42(10):64-71.
- [10] KAYA E, SERT E. A new 3D segmentation approach using extreme learning machine algorithm and morphological operations [J]. Computers & Electrical Engineering, 2020, 84, DOI: 10.1016/j.compeleceng.2020.106638.
- [11] SONG Y, HE B, ZHAO Y, et al. Segmentation of sidescan sonar imagery using markov random fields and extreme learning machine [J]. IEEE Journal of Oceanic Engineering, 2019, 44(2):502-513.
- [12] LIU Z F, LI L L, TSENG M L, et al. Prediction short-term photovoltaic power using improved chicken swarm optimizer-Extreme learning machine model [J]. Journal of Cleaner Production, 2020, 248, DOI: 10.1016/j.jclepro.2019.119272.
- [13] TIAN, Z D, REN Y, WANG G. Short-term wind speed prediction based on improved PSO algorithm

- optimized EM-ELM [J]. Energy Sources Part A-Recovery Utilization and Environmental Effects, 2019, 41(1): 26-46.
- [14] HOMAYOUNI H, MANSOORI E G. Manifold regularization ensemble clustering with many objectives using unsupervised extreme learning machines[J]. Intelligent Data Analysis, 2021, 25(4): 847-862.
- [15] CHEN J C, ZENG Y J, LI Y, et al. Unsupervised feature selection based extreme learning machine for clustering[J]. Neurocomputing, 2020, 386:198-207.
- [16] KASUN L 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.
- [17] ZHANG J, YIN J, ZHANG Q, et al. Robust sound event classification with bilinear multi-column ELM-AE and two-stage ensemble learning [J]. EURASIP Journal on Audio, Speech, and Music Processing, 2017, 2017:11, DOI: 10.1186/s13636-017-0109-1.
- [18] LU C, MEI Y. An imputation method for missing data based on an extreme learning machine auto-encoder[J]. IEEE Access, 2018, 6:52930-52935.
- [19] AYDOGDU O, EKINCI M. A new approach for data stream classification: Unsupervised feature representational online sequential extreme learning machine [J]. Multimedia Tools and Applications, 2020, 79:27205-27227.
- [20] 靳果,朱清智,闫奇. 基于 PCA 和 ML-ELM-AE 的短期光伏功率预测[J]. 控制工程, 2021, 28(9):1787-1796.
- [21] KALPANA V, MAHESWAR R, NANDAKUMAR E. Multiple parametric fault diagnosis using computational intelligence techniques in linear filter circuit[J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11:5533-5545.
- [22] WONG C M, VONG C M, WONG P K, et al. Kernel-based multilayer extreme learning machines for representation learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 29(3): 757-762.
- [23] SUN K, ZHANG J, ZHANG C, et al. Generalized extreme learning machine autoencoder and a new deep neural network[J]. Neurocomputing, 2017, 230: 374-381.
- [24] 陈晓云,陈媛. 子空间结构保持的多层极限学习机自编码器[J]. 自动化学报, 2022, 48(4):1091-1104.
- [25] LIU X, MIAO J, QING L, et al. Class-constrained extreme learning machine[C]. Proceedings of ELM-2015 Volume 1. Springer, Cham, 2016: 521-530, DOI: 10.1007/978-3-319-28397-5_41.
- [26] CHANG P, ZHANG J, WANG J, et al. ELMAENet: A simple, effective and fast deep architecture for image classification [J]. Neural Processing Letters, 2020, 51(1): 129-146.
- [27] 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.
- [28] HUANG G B. An insight into extreme learning machines: Random neurons, random features and kernels[J]. Cognitive Computation, 2014, 6(3): 376-390.
- [29] KASUN L L C, YANG Y, HUANG G B, et al. Dimension reduction with extreme learning machine[J]. IEEE Transactions on Image Processing, 2016, 25(8): 3906-3918.
- [30] GUO G, WANG H, BELL D, et al. KNN model-based approach in classification [C]. OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", Springer, Berlin, Heidelberg, 2003: 986-996, DOI:10.1007/978-3-540-39964-3_62.

作者简介

程蓉,工学博士,讲师,主要研究方向为信号与信息处理、模式识别、机器学习等。

E-mail: chengro@nuc.edu.cn

白艳萍,工学博士,教授,主要研究方向为人工智能,信号处理与辨识,优化计算等。

E-mail: baiyp666@163.com

胡红萍,工学博士,教授,主要研究方向为优化算法、机器学习、图像处理等。

E-mail: hhp92@163.com