

控制与决策

Control and Decision

基于极限学习机的无监督领域适应分类器

王雪松, 赵季娟, 程玉虎, 许德智

引用本文:

王雪松, 赵季娟, 程玉虎, 等. 基于极限学习机的无监督领域适应分类器[J]. 控制与决策, 2020, 35(4): 861–869.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2018.1082>

您可能感兴趣的其他文章

Articles you may be interested in

求解区间数分布式流水线调度的混合离散果蝇优化算法

A hybrid discrete fruit fly optimization algorithm for distributed permutation flowshop scheduling with interval data
控制与决策. 2020, 35(4): 930–936 <https://doi.org/10.13195/j.kzyjc.2018.1274>

脉冲神经网络:模型、学习算法与应用

Spiking neural networks: Model, learning algorithms and applications
控制与决策. 2018, 33(5): 923–937 <https://doi.org/10.13195/j.kzyjc.2017.1444>

基于ELM–PNN算法的第24周太阳黑子预测预报

Prediction and forecast of sunspot numbers in 24 cycle based on ELM–PNN algorithms
控制与决策. 2017, 32(4): 642–646 <https://doi.org/10.13195/j.kzyjc.2016.0093>

具有广义正则化与遗忘机制的在线贯序超限学习机

Online sequential extreme learning machine with generalized regularization and forgetting mechanism
控制与决策. 2017, 32(2): 247–254 <https://doi.org/10.13195/j.kzyjc.2015.1385>

基于相似度学习的多源迁移算法

Similarity–learning based multi–source transfer learning algorithm
控制与决策. 2017, 32(11): 1941–1948 <https://doi.org/10.13195/j.kzyjc.2016.1279>

基于自适应快速决策树的不确定数据流概念漂移分类算法

Classifying algorithm for concept drifting of uncertain data streams based on adapting fast decision tree algorithm
控制与决策. 2016, 31(9): 1609–1614 <https://doi.org/10.13195/j.kzyjc.2015.1110>

改进的灵敏度剪枝极限学习机

Improved sensitivity–analysis based pruning extreme learning machine
控制与决策. 2016(2): 249–255 <https://doi.org/10.13195/j.kzyjc.2014.1825>

自适应分组混沌云模型蛙跳算法求解连续空间优化问题

Adaptive grouping chaotic cloud model shuffled frog leaping algorithm for continuous space optimization problems
控制与决策. 2015, 30(5): 923–928 <https://doi.org/10.13195/j.kzyjc.2014.0387>

基于极限学习机的无监督领域适应分类器

王雪松, 赵季娟, 程玉虎[†]

(中国矿业大学 信息与控制工程学院, 江苏 徐州 221116)

摘要: 在构建基于极限学习机的无监督自适应分类器时, 隐含层的参数通常都是随机选取的, 而随机选取的参数不具备领域适应能力. 为了增强跨领域极限学习机的知识迁移能力, 提出一种新的基于极限学习机的无监督领域适应分类器学习方法, 该方法主要利用自编码极限学习机对源域和目标域数据进行重构学习, 从而可以获得具有领域不变特性的隐含层参数. 进一步, 结合联合概率分布匹配和流形正则的思想, 对输出层权重进行自适应调整. 所提出算法能对极限学习机的两层参数均赋予领域适应能力, 在字符数据集和对象识别数据集上的实验结果表明其具有较高的跨领域分类精度.

关键词: 领域适应; 极限学习机; 无监督; 分类器

中图分类号: TP391.4

文献标志码: A

Unsupervised domain adaption classifier via ELM

WANG Xue-song, ZHAO Ji-juan, CHENG Yu-hu[†]

(School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: When constructing an unsupervised adaptive classifier based on extreme learning machines (ELMs), the hidden layer parameters are usually randomly selected, which do not have domain adaption capability. In order to enhance the knowledge transfer ability of the cross-domain ELM, a new unsupervised domain adaptive classifier learning method via ELMs is proposed. The method mainly uses the extreme learning machine autoencoder to reconstruct the data of both source and target domains, which can obtain the hidden layer parameters with domain invariant features. Furthermore, by using the ideas of joint distribution adaption and manifold regularization, the output layer weights of ELMs can be adaptively adjusted. The proposed classifier can provide the domain adaption capability to both hidden layer and output layer parameters of ELMs. Experiments on the digital and object recognition datasets show that the proposed classifier has higher cross-domain classification accuracy.

Keywords: domain adaption; extreme learning machine; unsupervised; classifier

0 引言

作为机器学习领域中的一重要分支, 神经网络模型被广泛研究并应用于图像分类、目标检测、语音识别等各个场景. 常见的神经网络模型如自动编码器、卷积神经网络等都是利用反向传播算法, 通过逐层求导的方法学习网络参数. 近年来, 黄广斌教授提出的极限学习机 (ELM) 为神经网络开辟了一个崭新的视角^[1-2]. 作为一种单隐层神经网络模型, 极限学习机的网络参数由隐含层权重和输出层权重组成, 其隐含层权重是随机选取的, 而输出层权重可通过矩阵的广义逆求解出来. 虽然隐含层参数是随机选取的, 但是文献 [3] 通过理论知识证明了 ELM 具有强大的函数逼近能力. 大量研究表明, ELM 及其相关优化模型能有效解决如半监督^[4]、数据的缺失与失衡^[5-6]、过

程神经网络^[7]等许多具体的模式识别任务.

作为迁移学习方法中的一类, 领域适应算法^[8]可以使用源域数据训练一个分类模型, 该模型能有效作用于目标域样本, 其中源域和目标域的数据不同但具有相关性. 对于领域适应问题来说, 最常见的两类解决思想为: 1) 基于特征层面的跨领域知识迁移^[9-11], 主要做法是利用特征变换来减小领域间的概率分布差异; 2) 基于分类器层面的跨领域知识迁移^[12-13], 主要方法是将分布适配与模型正则相结合来设计自适应分类器.

随着 ELM 算法的不断发展进步, 许多基于 ELM 的优化方法可以有效解决领域适应问题. 根据不同的领域适应思想, 这些方法可分为两类: 1) 在 ELM 特征空间中的分布适应; 2) ELM 自适应分类器设计. 文

收稿日期: 2018-08-26; 修回日期: 2018-10-29.

基金项目: 国家自然科学基金项目 (61772532).

[†]通讯作者. E-mail: chengyuhu@163.com.

献[14]首次提出基于极限学习机的跨领域共享子空间学习方法,该方法通过最小化领域间的数据均值分布差异,构造出源域和目标域具有相似特征分布的子空间.进一步,文献[15]利用联合概率分布匹配^[9]的思想对源域和目标域数据进行特征重构,使得重构后的数据具有相似的特征分布.不同于传统的特征分布匹配方法,文献[16]提出领域适应特征增广ELM(AELM). AELM的主要思想为:首先,利用ELM的自编码重构学习来获得包含源域数据非线性结构特征的模型;然后,利用该模型对目标域数据进行异域特征的重构和增广来实现领域间的适配.但是AELM在训练过程中需要根据每一类学习一个特定的自编码ELM (ELM-AE)模型,当领域中存在较多类时,整个训练过程将会非常耗时.上述基于ELM特征空间分布适应的方法有一个共同特点:在训练阶段仅对领域间的特征分布进行适配,然后用变换后的特征训练一个普通的分类器,均未在分类器层面实现领域间的自适应.

由Zhang等^[17]首次提出的极限学习机跨领域学习机制是通过构造自适应ELM分类器实现的,主要思想是借助少量目标域样本的标签信息对源域样本训练得到的分类器进行微调,从而达到领域适应的目的.基于相同的学习机制,文献[17]提出两种学习框架,分别为源领域适应ELM(DAELM-S)和目标领域适应ELM(DAELM-T),这两者之间的区别在于DAELM-T在采用少量目标域标签信息微调的同时,还充分利用了目标域无标签样本的信息.同样是设计跨领域分类器的思想,文献[18]通过将目标域的ELM分类器参数迁移至源域的参数空间中使得两个领域的分类器参数分布尽可能相近,从而构造出自适应ELM分类器.然而,上述两种基于构造跨领域分类器的方法均要求目标域存在部分带标签的样本,不能实现无监督领域适应.文献[19]将概率分布匹配方法融入ELM输出层权重的学习中,可以实现在目标域无监督情况下的跨领域分类器学习.利用该方法获得的ELM输出层权重既能实现分类的功能,同时还具有较好的知识迁移能力.基于这种学习框架,根据不同类型的概率分布匹配,文献[19]提出两种算法:基于边缘概率分布的交叉领域ELM(CDELMM)、基于条件概率分布的交叉领域ELM(CDELMC).更进一步,文献[20]将联合概率分布匹配正则项加入求解分类器参数的目标函数中,有效提升了ELM分类器参数的领域适应性能.

通过上述分析可以发现:已有的基于构建自适

应ELM分类器的方法都是仅针对输出层权重进行跨领域适配,忽略了ELM隐含层权重的自适应学习.为此,本文提出一种基于极限学习机的无监督领域适应分类器(UDAC-ELM):首先,利用自编码极限学习机模型的重构学习获得具有领域适应能力的ELM隐含层权重,经过该隐含层权重映射后得到的特征具有较好的领域不变特性;然后,将联合概率分布匹配正则项和流形正则项嵌入求解输出层参数的目标函数中,使得学习到的输出层权重获得较好的迁移性能;最后,将该ELM跨领域分类器用于目标域样本,可直接获得其预测标签.

1 极限学习机

极限学习机是一种快速高效的单隐层神经网络模型,给定 n 个训练样本 $\mathbf{X} = (x_1, x_2, \dots, x_n) \in R^{d_1 \times n}$,其标签为 $\mathbf{Y} = (y_1, y_2, \dots, y_n) \in R^{n \times d_2}$, d_1 和 d_2 分别表示输入数据和输出数据的维度.极限学习机的隐含层权重 $\mathbf{W} = w_{ij} \in R^{d_1 \times L}$ 是随机选取的^[3],其中 L 表示隐含层神经元的个数.计算隐含层与计算传统的前向传播网络相同,有 $\mathbf{H} = g(\mathbf{X}, \mathbf{W})$,其中 $\mathbf{H} \in R^{n \times L}$, $g(\cdot)$ 为激活函数.

极限学习机的学习目标是通过最小化预测误差损失函数之和来求解输出权重 β ^[1],目标函数为

$$\min_{\beta \in R^{L \times d_2}} \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \|\mathbf{Y} - \mathbf{H}\beta\|^2. \quad (1)$$

其中:第1项是防止参数过拟合的正则项,第2项中的 C 为预测误差项的惩罚系数.求解上述目标函数可视为最小二乘优化问题,令目标函数对于 β 的梯度为零,可得

$$\nabla_{\beta} L_{\text{ELM}} = \beta + C\mathbf{H}^T(\mathbf{Y} - \mathbf{H}\beta) = 0. \quad (2)$$

根据Moore-Penrose广义逆矩阵,可以求出 β 的最优解.当训练样本数量 n 和隐含层神经元个数 L 大小不同时,求解 β 存在如下两种情况^[2]:

$$\beta^* = \begin{cases} \left(\mathbf{H}^T\mathbf{H} + \frac{\mathbf{I}_L}{C}\right)^{-1} \mathbf{H}^T\mathbf{Y}, & n \geq L; \\ \mathbf{H}^T\left(\mathbf{H}^T\mathbf{H} + \frac{\mathbf{I}_n}{C}\right)^{-1} \mathbf{Y}, & n < L. \end{cases} \quad (3)$$

其中 \mathbf{I}_n 和 \mathbf{I}_L 分别表示 n 维和 L 维的单位矩阵.

2 基于极限学习机的无监督领域适应分类器

2.1 系统结构框图

给定源域和目标域的输入样本 $\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_t] \in R^{d_1 \times n}$,源域样本的标签为 $\mathbf{Y}_s \in R^{n_s \times 1}$,目标域中的样本没有标签,其中 d_1 表示输入数据的维数, $n = n_s + n_t$ 表示源域和目标域的样本个数之和.如图1

所示, UDAC-ELM分类器参数学习过程分为两个阶段. 阶段1的目的是学习具有领域不变特性的隐含层权重 \mathbf{W} , 利用自编码极限学习机模型 ELM-AE 对源域和目标域的数据进行无监督重构, 使得学习到的 ELM-AE 输出层权重 \mathbf{B} 带有领域间的共有特征, 这样 $\mathbf{W} = \mathbf{B}^T$ 也能充分提取源域和目标域之间的共同不变特征. 阶段2旨在学习具有迁移能力的输出层权重 β , 根据源域样本的分类损失、目标域样本的流形正

则损失以及领域间的联合概率分布差异损失3项构成目标函数, 通过求解该目标函数即可获得具有迁移能力的分类器输出层权重. 值得注意的是, 在整个参数学习过程中目标域的标签信息未知. 因此, 在计算条件概率分布差异时可采用文献[9]提出的伪标签迭代机制来代替目标域的真实标签. 在测试过程中, 直接将目标域的测试样本输入 UDAC-ELM 分类器中即可得到最终的预测标签.

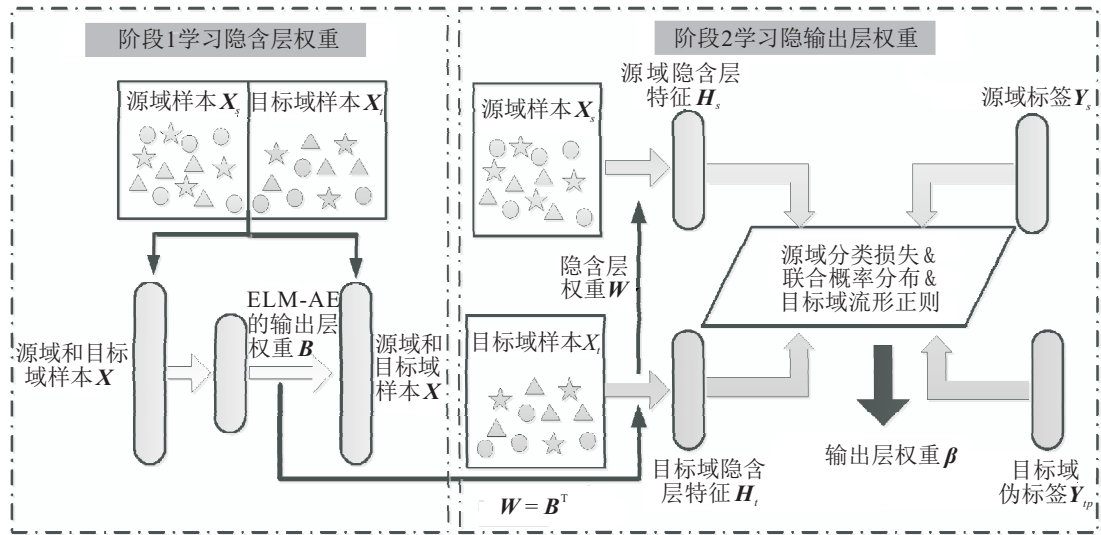


图1 UDAC-ELM的系统结构框图

2.2 ELM的领域不变隐含层权重学习

传统极限学习机的隐含层权重是随机选取的, 文献[3]用严格的理论证明了只要隐含层的激活函数是有界非常量连续分段函数, 随着隐含层神经元数量的增加, ELM可以逐渐逼近任何连续函数. 尽管随机选取的隐含层权重具有强大的函数拟合能力, 但是在跨领域分类任务中却缺乏较好的知识迁移能力. 因此, 在 UDAC-ELM 中利用 ELM-AE 的无监督重构学习来获得具有领域不变性能的隐含层权重矩阵. 采用 ELM-AE 学习 ELM 分类器隐含层权重具有以下优点: 1) 可以在无监督的情况下学习到输入数据的非线性结构特征; 2) 通过对输入数据进行重构学习可以迫使模型参数能有效提取数据的结构特征; 3) 当 ELM-AE 模型同时对不同领域数据进行最小误差重构时, 学习到的模型参数具有提取领域间共同不变特征的能力; 4) 整个参数训练过程中不需要反复迭代优化, 因而比较快速高效.

在 ELM-AE 的训练过程中, 首先随机选取正交的权重矩阵来计算隐含层特征, 正交的矩阵可以将输入数据映射到一个随机子空间中, 以此增强模型的泛化能力^[21]; 其次, 为了获得能提取领域间共同特征的参数, 令模型的目标输出和输入数据相等, 其中

$\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_t]$ 表示源域和目标域样本数据的组合; 然后, 利用式(3)计算得到 ELM-AE 输出层的权重矩阵 \mathbf{B} , 因为 \mathbf{B} 是由源域和目标域数据共同重构学习得到的, 所以能够有效提取源域和目标域数据之间的共同特征; 最后, 将 $\mathbf{W} = \mathbf{B}^T$ 作为跨领域 ELM 分类器的隐含层权重矩阵, 该矩阵不仅可以充分拟合数据的非线性结构特征, 而且具有较好的领域不变性能.

2.3 具有迁移能力的 ELM 输出层权重学习

根据第 2.2 节求得的 ELM 的领域不变隐含层权重 \mathbf{W} 可计算出源域和目标域相应的隐含层特征

$$\mathbf{H}_s = g(\mathbf{X}_s, \mathbf{W}), \tag{4}$$

$$\mathbf{H}_t = g(\mathbf{X}_t, \mathbf{W}), \tag{5}$$

其中 $g(\cdot)$ 表示隐含层激活函数. 为进一步提升 ELM 分类器的跨领域知识迁移能力, UDAC-ELM 在输出层权重的学习过程中加入联合分布适配正则项和目标域流形正则项, 从而使得输出层权重具有较好的迁移性能.

2.3.1 源域分类损失

在无监督跨领域分类器学习过程中, 源域样本包含所有的标签信息. 因此, 可以充分利用这些带标签的源域样本来学习跨领域分类器参数. 给定源域样

本数据与标签 $\{\mathbf{X}_s, \mathbf{Y}_s\}$, 根据第1节中的极限学习机算法, 分类器输出层权重的损失函数可表示为

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \|\mathbf{Y}_s - \mathbf{H}_s \beta\|^2, \quad (6)$$

其中 \mathbf{H}_s 为源域的隐含层特征表达. 为了方便对目标函数的求解, 将式(6)转换为如下形式:

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{1}{2} \text{tr}((\mathbf{H}_s \beta - \mathbf{Y}_s)^T \mathbf{C}_s (\mathbf{H}_s \beta - \mathbf{Y}_s)). \quad (7)$$

其中: $\text{tr}(\cdot)$ 为矩阵的迹, $\mathbf{C}_s = \text{diag}(C, C, \dots, C) \in R^{n_s \times (n_s)}$ 为所有源域样本的惩罚系数矩阵.

2.3.2 联合分布适配

减小领域间数据的概率分布差异是领域适应问题中最重要的解决方案之一. 最大均值差异(MMD)是一种非常有效的概率分布差异度量方法, 利用MMD算法联合匹配领域间的边缘概率分布和条件概率分布使许多跨领域迁移学习取得了较好的结果^[8,19]. 因此, 在UDAC-ELM分类器的输出层权重学习过程中加入联合概率分布适配正则项来进一步提升其迁移性能. 源域和目标域的边缘概率分布差异用MMD度量可表示为

$$D^2(D_s, D_t)_m = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} f(x_{si}) - \frac{1}{n_t} \sum_{i=1}^{n_t} f(x_{ti}) \right\|^2. \quad (8)$$

其中: $f(x_{si}) = h(x_{si})\beta$, $f(x_{ti}) = h(x_{ti})\beta$, $h(x_{si})$ 和 $h(x_{ti})$ 分别表示源域和目标域样本的隐含层特征. 将式(8)转换成矩阵形式, 领域间的边缘概率分布差异损失函数为

$$\text{tr}(\beta^T \mathbf{H}^T \mathbf{M}_m \mathbf{H} \beta). \quad (9)$$

其中: $\mathbf{H} = [\mathbf{H}_s^T, \mathbf{H}_t^T]^T \in R^{(n_s+n_t) \times L}$ 表示源域与目标域的隐含层特征表达, \mathbf{M}_m 为边缘MMD矩阵, 有

$$(M_m)_{ij} = \begin{cases} \frac{1}{n_s n_s}, & x_i, x_j \in D_s; \\ \frac{1}{n_t n_t}, & x_i, x_j \in D_t; \\ -\frac{1}{n_s n_t}, & \text{otherwise.} \end{cases} \quad (10)$$

其中 D_s 和 D_t 分别表示源域和目标域. 同样地, 利用MMD度量来计算领域间的条件概率分布差异

$$D^2(D_s, D_t)_c = \sum_{k=1}^K \left\| \frac{1}{n_s^{(k)}} \sum_{y_{si}=k} f(x_{si}) - \frac{1}{n_t^{(k)}} \sum_{y_{ti}=k} f(x_{ti}) \right\|^2. \quad (11)$$

其中: K 表示样本的类别数, $n_s^{(k)}$ 和 $n_t^{(k)}$ 表示第 k 类源域和目标域样本的个数. 将式(11)转换成矩阵形式,

领域间的条件概率分布差异损失函数为

$$\text{tr}(\beta^T \mathbf{H}^T \mathbf{M}_c \mathbf{H} \beta). \quad (12)$$

其中: $\mathbf{M}_c = \sum_{k=1}^K \mathbf{M}_c^{(k)}$ 为条件MMD矩阵, $\mathbf{M}_c^{(k)}$ 的计算过程如下:

$$(M_c)_{ij} = \begin{cases} \frac{1}{n_s^{(k)} n_s^{(k)}}, & x_i, x_j \in D_s^{(k)}; \\ \frac{1}{n_t^{(k)} n_t^{(k)}}, & x_i, x_j \in D_t^{(k)}; \\ -\frac{1}{n_s^{(k)} n_t^{(k)}}, & \begin{cases} x_i \in D_s^{(k)}, x_j \in D_t^{(k)}; \\ x_j \in D_s^{(k)}, x_i \in D_t^{(k)}; \end{cases} \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

其中: $D_s^{(k)} = \{x_i : x_i \in D_s \wedge y(x_i) = k\}$ 表示源域中属于类别 k 的样本集合, $y(x_i)$ 为 x_i 的真实标签, $n_s^{(k)} = |D_s^{(k)}|$; 同理, $D_t^{(k)} = \{x_j : x_j \in D_t \wedge \bar{y}(x_j) = k\}$ 表示目标域中属于类别 k 的样本集合, $\bar{y}(x_j)$ 为 x_j 的伪标签, $n_t^{(k)} = |D_t^{(k)}|$. 值得注意的是, 由于目标域样本中没有标签信息, 本文采用伪标签方法计算领域间的条件概率分布差异.

2.3.3 流形正则

在机器学习领域中, 有一个著名的平滑流形假设: 当两个样本点 x_1 和 x_2 在几何空间中很相近时, 它们的条件概率分布 $P(y|x_1)$ 和 $P(y|x_2)$ 也是相似的, 该假设可以保证分类器具有较为平滑的预测能力^[4]. 采用流形正则惩罚项能够在无监督场景下获得数据的潜在信息, 从而增强跨领域分类器在目标域中的分类性能. 对于目标域样本来说, 流形正则项可表示为

$$\min_{\beta} \frac{1}{2} \sum_{i,j=1}^{n_t} \alpha_{ij} \|h(x_{ti})\beta - h(x_{tj})\beta\|^2. \quad (14)$$

其中: $h(x_{ti})$ 和 $h(x_{tj})$ 表示目标域样本的隐含层特征, α_{ij} 代表 x_{ti} 与 x_{tj} 之间的相似性, 其计算方法如下:

$$\alpha_{ij} = \begin{cases} 1, & x_{ti} \in \mathbf{N}_p(x_{ti}) \text{ or } x_{tj} \in \mathbf{N}_p(x_{tj}); \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

其中: $\mathbf{N}_p(x_{ti})$ 表示 x_{ti} 的 p -近邻集合. 将式(14)转换成矩阵形式, 可表示为

$$\min_{\beta} \text{tr}(\beta^T \mathbf{H}_t^T \mathbf{G}_t \mathbf{H}_t \beta). \quad (16)$$

其中 $\mathbf{G}_t = \mathbf{D} - \mathbf{A}$ 表示目标域数据的拉普拉斯矩阵, $\mathbf{A} = [\alpha_{ij}]_{n_t \times n_t}$, \mathbf{D} 为对角矩阵, 其对角元素计算为 $D_{ii} = \sum_{i=1}^{n_t} \alpha_{ij}$.

2.4 目标函数求解

将式(7)、(9)、(12)和(16)联合,可得到求解跨领域分类器输出层权重的目标函数

$$l = \min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{1}{2} \text{tr}((\mathbf{H}_s \beta - \mathbf{Y}_s)^T \mathbf{C}_s (\mathbf{H}_s \beta - \mathbf{Y}_s)) + \frac{\lambda}{2} \text{tr}[\beta^T \mathbf{H}^T (\mathbf{M}_m + \mathbf{M}_c) \mathbf{H} \beta] + \frac{\gamma}{2} \text{tr}(\beta^T \mathbf{H}_t^T \mathbf{G}_t \mathbf{H}_t \beta), \quad (17)$$

其中 λ 和 γ 分别表示联合分布匹配正则项和流形正则项的惩罚系数. 为了方便求解目标函数,给出以下几项:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_s \\ \mathbf{0}_{n_t \times d_2} \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_s & \mathbf{0}_{n_s \times n_t} \\ \mathbf{0}_{n_t \times n_s} & \mathbf{0}_{n_t \times n_t} \end{bmatrix},$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0}_{n_s \times n_s} & \mathbf{0}_{n_s \times n_t} \\ \mathbf{0}_{n_t \times n_s} & \mathbf{G}_t \end{bmatrix}.$$

其中: d_2 为输出数据的维度, $\mathbf{0}_{n_s \times n_t}$ 为一个 n_s 行 n_t 列并且元素都为0的矩阵. 将式(17)表示的目标函数改写成最终表达形式

$$l = \min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{1}{2} \text{tr}((\mathbf{H} \beta - \mathbf{Y})^T \mathbf{C} (\mathbf{H} \beta - \mathbf{Y})) + \frac{\lambda}{2} \text{tr}[\beta^T \mathbf{H}^T (\mathbf{M}_m + \mathbf{M}_c) \mathbf{H} \beta] + \frac{\gamma}{2} \text{tr}(\beta^T \mathbf{H}^T \mathbf{G} \mathbf{H} \beta). \quad (18)$$

令目标函数关于 β 的梯度为0,可得

$$\nabla l = \beta + \mathbf{H}^T \mathbf{C} (\mathbf{H} \beta - \mathbf{Y}) + \gamma \mathbf{H}^T \mathbf{G} \mathbf{H} \beta + \lambda \mathbf{H}^T (\mathbf{M}_m + \mathbf{M}_c) \mathbf{H} \beta. \quad (19)$$

根据隐含层ELM隐含层神经元个数 L 与样本量 $n_s + n_t$ 的大小关系,求解输出层权重 β 存在的两种情况,当 $L \leq n_s + n_t$ 时,有

$$\beta^* = [\mathbf{I}_L + \mathbf{H}^T (\mathbf{C} + \lambda (\mathbf{M}_m + \mathbf{M}_c) + \gamma \mathbf{G}) \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{C} \mathbf{Y}; \quad (20)$$

当 $L > n_s + n_t$ 时,有

$$\beta^* = \mathbf{H}^T [\mathbf{I}_{n_s+n_t} + (\mathbf{C} + \lambda (\mathbf{M}_m + \mathbf{M}_c) + \gamma \mathbf{G}) \mathbf{H} \mathbf{H}^T]^{-1} \mathbf{C} \mathbf{Y}. \quad (21)$$

其中 \mathbf{I}_L 和 $\mathbf{I}_{n_s+n_t}$ 分别表示维度为 L 和 $n_s + n_t$ 的单位矩阵.

2.5 算法步骤

输入:源域和目标域的输入样本 $\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_t]$,源域样本标签 \mathbf{Y}_s ,ELM隐含层神经元个数 L ,惩罚系

数 C 、 λ 和 γ ,最大迭代次数 T ;

输出:跨领域ELM分类器 $\{\mathbf{W}, \beta\}$,目标域样本的预测标签.

step 1: 令ELM-AE的输出与输入为 \mathbf{X} ,根据式(3)求出ELM-AE的输出层权重矩阵 \mathbf{B} ;

step 2: 令 $\mathbf{W} = \mathbf{B}^T$ 为跨领域分类器隐含层权重,根据式(4)和(5)计算源域和目标域隐含层特征 \mathbf{H}_s 和 \mathbf{H}_t ;

step 3: 令条件MMD矩阵 \mathbf{M}_c 为0,分别计算边缘MMD矩阵 \mathbf{M}_m 和拉普拉斯矩阵 \mathbf{G}_t ;

step 4: 根据式(20)或(21),计算ELM分类器输出层权重 β ;

step 5: 利用所得的分类器参数 \mathbf{W} 和 β ,预测目标域数据的伪标签 \mathbf{Y}_{tp} ;

step 6: 令 $t = 1$,重复step 6 ~ step 8;

step 7: 计算边缘MMD矩阵 \mathbf{M}_m 、条件MMD矩阵 \mathbf{M}_c 和拉普拉斯矩阵 \mathbf{G}_t ;

step 8: 根据式(20)或(21),求 β ;

step 9: 利用求得的ELM分类器参数 \mathbf{W} 和 β ,预测目标域数据的伪标签 \mathbf{Y}_{tp} ;

step 10: $t = t + 1$,直到 $t = T$.

3 实验结果与分析

3.1 实验数据集

为了验证UDAC-ELM的领域适应能力,使用4组基准数据集来构造领域适应任务,分别是:USPS+MNIST和Office+Caltech-256,一共构造14组迁移任务.USPS和MNIST都是应用较为广泛的字符数据集.USPS中包含7291幅训练图片和2007幅测试图片,每张都是 16×16 像素的灰度图像.MNIST中含有60000幅训练样本和10000幅测试样本,每张都是 8×28 像素的灰度图像.这两个数据集的分布是不同的,但是它们共享10个类别.将所有图像统一表示为 16×16 像素的灰度图像,并且分别从MNIST和USPS中随机抽取2000张和1800张图像构造两组领域适应任务USPS vs MNIST和MNIST vs USPS.

Office为视觉领域适应中常用的基准对象识别数据集,它由3个对象领域组成,在电商图片Amazon、网络摄像头拍摄的低分辨率图片Webcam、数字单反相机拍摄的高分辨率图片DSLR,一共有31个类别,合计4652张图片.Caltech-256是对象识别常用的标准数据集,它包含256个类别共30607幅图像.在实验中采用文献[9]发布的Office+Caltech10数据集,共有4个领域:C(Caltech-256)、A(Amazon)、W(Webcam)和D(DSLR).随机选择两个不同领域,分

别作为源领域和目标领域,可以构建 $4 \times 3 = 12$ 组领域适应任务,如 C vs A、C vs W、C vs D、...、D vs W. 数据集的具体描述如表1所示.

表1 实验数据集

数据集	类型	样本数	特征维数	类别数	子集
USPS	digit	1800	256	10	USPS
MNIST	digit	2000	256	10	MNIST
Office	object	1410	800	10	A, W, D
Caltech-256	object	1123	800	10	C

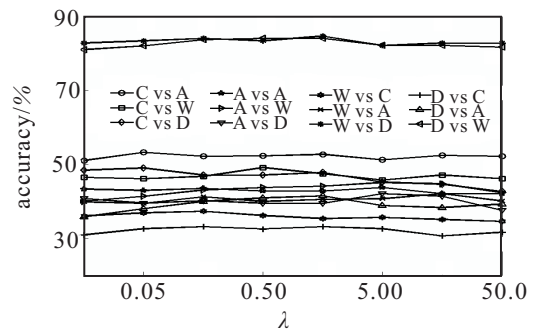
3.2 参数分析

为考察参数 λ 、 γ 、 C 和 T 对 UDAC-ELM 分类精度的影响,分别对所有字符数据集和对象识别数据集的 14 组领域适应任务进行实验,绘制如图 2~图 5 所示的各参数对跨领域分类精度影响曲线. 在实验过程中, λ 的取值范围为 $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$, γ 的取值范围为 $\{10^{-6}, 5 \times 10^{-6}, 10^{-5}, 5 \times 10^{-5}, 10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}\}$, C 的取值范围为 $\{0.0001, 0.001, 0.01, 0.1, 1, 10\}$, 每组任务共迭代 $T = 10$ 次. 对于 λ 、 γ 和 C 来说,在考察某一参数对分类精度的影响时,其他两个参数保持不变. 此外,在绘制分类精度随迭代次数变化的折线图时,为了能清晰地可视化曲线的变化趋势,将 14 组迁移任务按分类精度的范围拆分为两组显示.

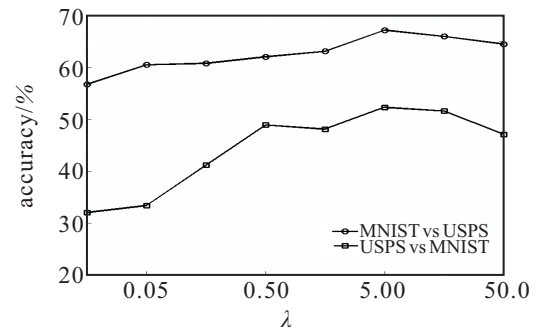
联合概率分布正则项系数 λ 对目标域分类精度的影响情况如图 2 所示. 可以看出,在对象识别数据集上曲线的总体走势比较平缓. 当 λ 取值很小时,概率分布适配的作用不明显,因此在目标域上的分类结果总体较低;当 λ 逐渐增大时,分类性能有所提升并且较为稳定. 在字符数据集中跨领域分类表现对 λ 相对比较敏感,随着 λ 的增大,分类精度逐渐提升;当 λ 大于 5 之后,分类精度趋于稳定并随着 λ 的增大略有下降趋势. 因此,对于 Office+Caltech10 和 MNIST+USPS 数据集, λ 的取值分别为 0.5 和 5.

流形正则项惩罚系数对分类性能的影响曲线如图 3 所示. 可以发现,无论是对象识别数据集还是字符数据集,在目标域中的分类精度都是先小幅上升后大幅下降. 这表明增加流形正则项可以有效增强分类器对于目标域的适应能力,但是当 γ 逐渐增大时,会很大程度上削弱目标函数中其他项的作用,从而导致跨领域分类正确率的下降. 实验表明,在 Office+Caltech10 和 MNIST+USPS 数据集中, γ 的取值分别为 10^{-4} 和 5×10^{-5} 时能获得比较满意的跨领域分类精度.

C 是极限学习机预测误差项的惩罚系数,主要用

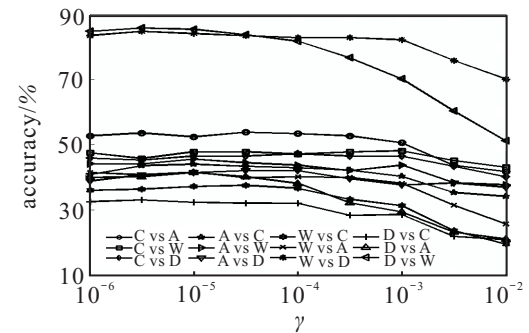


(a) Office+Caltech10 数据集

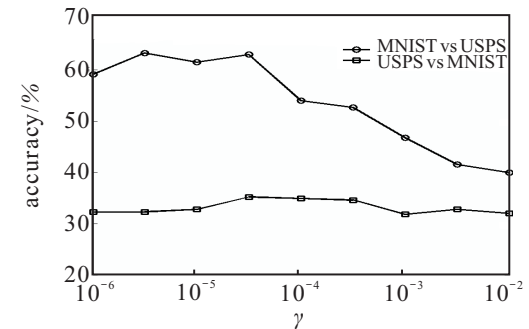


(b) MNIST+USPS 数据集

图2 分类精度与 lambda 间关系



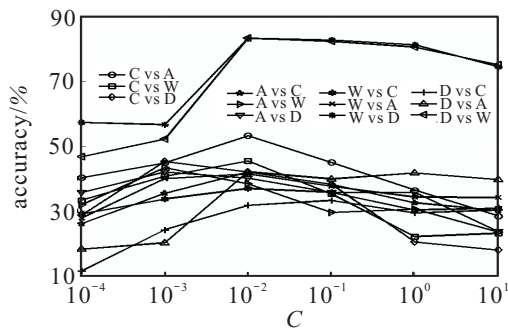
(a) Office+Caltech10 数据集



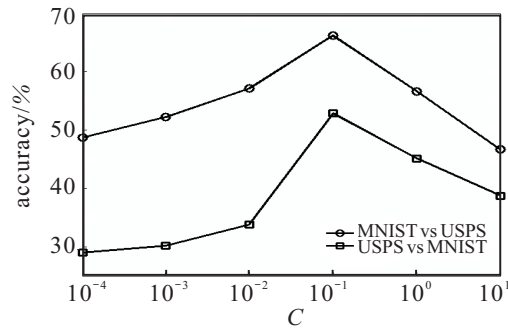
(b) MNIST+USPS 数据集

图3 分类精度与 gamma 间关系

来平衡 ELM 分类器的训练误差和正则项的影响. 如图 4 所示,在每组学习任务中,曲线的走势基本上都是先上升后下降. 这是因为当 C 过小或过大时,都会偏重某一项的功能而弱化另一项的作用,不能达到很好的平衡效果. 对于不同的分类任务来说, C 的取值会有所差异,在 Office+Caltech10 的 12 组迁移任务中, C 取 0.01 或 0.1 可以取得较好的分类效果;在 MNIST+USPS 中, C 取 0.1 可获得最佳分类精度.



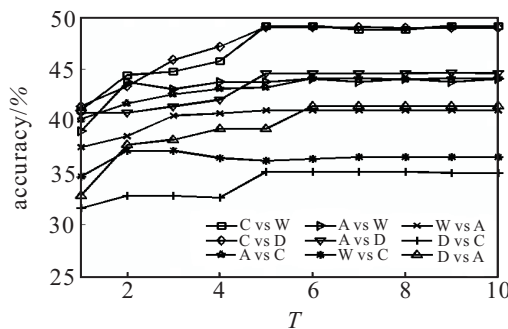
(a) Office+Caltech10 数据集



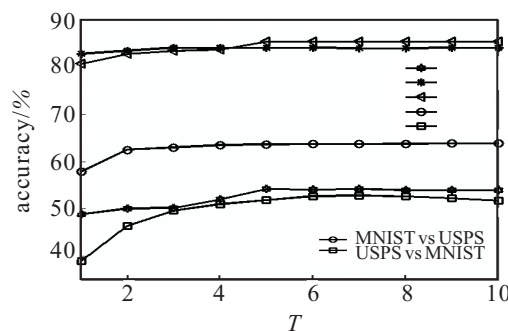
(b) MNIST+USPS 数据集

图4 分类精度与C间关系

由于在计算条件概率分布时没有目标域的标签信息,所以采用伪标签迭代方法来逐步提升分类效果.如图5所示,随着迭代次数的增加,分类精度逐步提升,当迭代到第7次左右时,分类正确率基本不再发生变化.



(a) Office+Caltech10 数据集



(b) MNIST+USPS 数据集

图5 分类精度与T间关系

3.3 对比实验

为验证UDAC-ELM的有效性,将其与8种算法进行对比,分别是:1)不添加迁移算法的传统极限

学习机ELM^[11];2)领域适应特征增广极限学习机AELM^[16];3)半监督的源(目标)领域适应ELM,选取10个样本对分类器进行微调,分别为DAELM-S(10)^[17]和DAELM-T(10)^[17];4)使用联合概率分布匹配正则项对ELM输出层参数进行适配的JUC-SDEL^M^[20];5)基于流形正则与边缘(条件)概率分布的交叉领域极限学习机CDEL^M-M(C)^[19];6)随机选取隐含层参数的跨领域ELM无监督领域适应分类器UDAC-ELMRS.其中ELM、CDEL^M-M、CDEL^M-C、DAELM-S(10)和DAELM-T(10)的结果均来自文献[19],AELM和JUC-SDEL^M的结果分别来自文献[16,20].对于UDAC-ELMRS和UDAC-ELM,隐含层神经元个数L的取值范围为[500,2000],最大迭代次数取10.在Office+Caltech10数据集中,参数的取值如下: $\lambda = 0.5, \gamma = 10^{-4}, C = \{0.01, 0.1\}$;在MNIST+USPS数据集中,各参数取值为 $\lambda = 5, \gamma = 5 \times 10^{-5}, C = 0.1$.UDAC-ELMRS的ELM分类器隐含层权重是随机选取的,未使用具有领域不变特性的隐含层参数.

由表2可以看出:1)UDAC-ELM在14组跨领域分类任务中平均分类精度为51.97%,高于其他所有对比算法,并且有8组迁移任务取得最高分类精度,说明其具有较强的知识迁移能力.2)DAELM-S(10)和DAELM-T(10)算法在实现过程中需要用到部分目标域的标签信息,尽管如此,其分类效果仍然低于无监督的JUC-SDEL^M、CDEL^M-M(C)、UDAC-ELMRS和UDAC-ELM,这表明采用概率分布匹配思想构造自适应分类器的方法优于使用少量标签信息微调分类器参数的做法.3)AELM是基于特征变换的领域适应算法,它利用ELM-AE模型对样本特征进行重构与异域增广,从而增强领域间的适配性.从实验结果来看,AELM获得较高的跨领域分类结果,这表明ELM-AE模型的重构学习能有效地刻画各领域的分布特征.4)与仅匹配边缘分布或条件分布的CDEL^M-M和CDEL^M-C相比,联合匹配边缘和条件分布的UDAC-ELMRS和UDAC-ELM在大多数组迁移任务中具有较高的分类精度.5)UDAC-ELMRS的平均分类结果比JUC-SDEL^M有较大提升,说明目标域的流形正则约束能有效加强分类器在目标域样本中的识别能力.6)相较于UDAC-ELMRS,UDAC-ELM在目标域上的分类精度提升了2.08%,主要是因为UDAC-ELM不仅对ELM输出层参数赋予迁移性能,而且利用ELM-AE的重构学习保证了其隐含层权重的领域不变特性.

表2 分类精度对比

单位: %

tasks	ELM	AELM	DAELM-S(10)	DAELM-T(10)	JUC-SDELM	CDELM-M	CDELM-C	UDAC-ELMRS	UDAC-ELM
C vs A	52.30	53.13	49.19	51.48	49.66	52.07	56.28	51.67	54.18
C vs W	46.78	49.49	42.31	44.79	39.97	51.05	50.98	46.56	49.15
C vs D	41.40	50.96	43.82	43.22	44.20	45.86	43.82	46.5	47.78
A vs C	42.56	41.14	41.83	42.37	42.38	42.33	43.17	43.63	44.08
A vs W	38.64	35.25	39.53	39.61	38.85	42.85	40.75	40.68	42.68
A vs D	39.49	36.94	40.76	42.31	43.82	45.86	41.78	43.54	44.59
W vs C	34.28	34.11	34.32	33.67	29.54	30.31	34.44	35.00	37.13
W vs A	37.27	38.93	38.12	38.83	35.47	39.83	39.58	40.25	41.75
W vs D	80.25	89.81	82.68	83.62	81.85	81.15	83.06	82.96	84.08
D vs C	34.28	33.83	34.32	33.67	27.56	30.31	34.44	33.74	35.08
D vs A	34.97	33.09	39.02	36.64	37.04	35.72	39.54	36.74	41.75
D vs W	81.02	80.33	83.39	83.15	74.81	81.76	84.34	83.56	85.42
U vs M	29.71	57.77	45.69	31.96	40.48	46.53	45.89	50.50	52.85
M vs U	25.42	62.33	63.01	31.74	62.83	62.47	42.57	63.15	67.11
average	44.17	49.79	48.43	45.50	46.31	49.15	48.62	49.89	51.97

4 结论

极限学习机可以看作一种由两层网络参数组成的分类器模型,并且具有训练速度快、泛化能力强等优点. 通过对其网络参数进行自适应调整,可以构造出具有迁移能力的跨领域极限学习机分类器. 但是,已有的方法只针对ELM的输出层参数进行了自适应学习,为进一步增强其迁移能力,提出一种能对ELM分类器的两层网络参数均赋予领域适应能力的方法. 主要思路为:在对ELM输出层权重采取联合概率分布匹配和目标域流形正则学习的同时,还利用ELM-AE模型对源域和目标域数据进行最小误差重构学习,从而训练得到能提取领域间共同特性的网络参数. 将该网络权重的转置作为跨领域ELM分类器的隐含层参数,便可以有效提取出领域间的共同不变特征. 实验结果表明,本文算法在多组迁移任务中具有良好的跨领域分类效果.

参考文献(References)

- [1] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feedforward neural networks[C]. Proceedings of IEEE International Conference on Neural Networks. Piscataway: IEEE Press, 2004: 985-990.
- [2] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1): 489-501.
- [3] 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.

- [4] Huang G, Song S, Gupta J N D, et al. Semi-supervised and unsupervised extreme learning machines[J]. IEEE Transactions on Cybernetics, 2014, 44(12): 2405-2417.
- [5] Yu Q, Miche Y, Eirola E, et al. Regularized extreme learning machine for regression with missing data[J]. Neurocomputing, 2013, 102(1): 45-51.
- [6] Zong W, Huang G B, Chen Y. Weighted extreme learning machine for imbalance learning[J]. Neurocomputing, 2013, 101(1): 229-242.
- [7] 刘志刚, 许少华, 杜娟, 等. 一种混合优化的结构自适应极限过程神经网络及应用[J]. 控制与决策, 2018, 33(7): 1335-1340.
(Liu Z G, Xu S H, Du J, et al. Structure-adapted extreme process neural network and application based on hybrid optimization[J]. Control and Decision, 2018, 33(7): 1335-1340.)
- [8] 孟娟, 胡谷雨, 潘志松, 等. 领域适应学习算法研究与展望[J]. 计算机科学, 2015, 42(10): 7-12.
(Meng J, Hu G Y, Pan Z S, et al. Research and perspective on domain adaption algorithms[J]. Computer Science, 2015, 42(10): 7-12.)
- [9] Long M S, Wang J M, Ding G G, et al. Transfer feature learning with joint distribution adaptation[C]. Proceedings of IEEE Conference on Computer Vision. Piscataway: IEEE Press, 2013: 2200-2207.
- [10] Gong B, Shi Y, Sha F, et al. Geodesic flow kernel for unsupervised domain adaptation[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012: 2066-2073.
- [11] 臧绍飞, 程玉虎, 王雪松. 基于最大分布加权均值

嵌入的领域适应学习[J]. 控制与决策, 2016, 31(11): 2083-2089.
(Zang S F, Cheng Y H, Wang X S. Domain adaptation learning based on maximum distribution weighted mean discrepancy[J]. Control and Decision, 2016, 31(11): 2083-2089.)

[12] Long M S, Wang J M, Ding G G, et al. Adaptation regularization: A general framework for transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5): 1076-1089.

[13] 许敏, 王士同, 顾鑫. TL-SVM: 一种迁移学习算法[J]. 控制与决策, 2014, 29(1): 141-146.
(Xu M, Wang S T, Gu X. TL-SVM: A transfer learning algorithm[J]. Control and Decision, 2014, 29(1): 141-146.)

[14] Liu Y, Zhang L, Deng P L, et al. Common subspace learning via cross-domain extreme learning machine[J]. Cognitive Computation, 2017, 9(4): 555-563.

[15] Chen Y M, Song S J, Li S, et al. Domain space transfer extreme learning machine for domain adaptation[J]. IEEE Transactions on Cybernetics, 2018, DOI: 10.1109/TCYB.2018.2816981.

[16] Uzair M, Mian A. Blind domain adaptation with augmented extreme learning machine features[J]. IEEE Transactions on Cybernetics, 2017, 47(3): 651-660.

[17] Zhang L, Zhang D. Domain adaptation extreme learning machines for drift compensation in E-nose systems[J]. IEEE Transactions on Instrumentation and Measurement, 2015, 64(7): 1790-1801.

[18] 许凤晖, 慕晓冬, 柴栋, 等. 基于极限学习机参数迁移的域适应算法[J]. 自动化学报, 2018, 44(2): 311-317.
(Xu S H, Mu X D, Chai D, et al. Domain adaptation algorithm with ELM parameter transfer[J]. Acta Automatica Sinica, 2018, 44(2): 311-317.)

[19] Li S, Song S J, Huang G, et al. Cross-domain extreme learning machines for domain adaptation[J]. IEEE Transactions on Systems Man and Cybernetics: Systems, 2017, DOI: 10.1109/TSMC.2017.2735997.

[20] Zhang B Y, Liu Y Y, Yuan H W, et al. A joint unsupervised cross-domain model via scalable discriminative extreme learning machine[J]. Cognitive Computation, 2018, 10(4): 1-14.

[21] Kasun L L C, Zhou H M, Huang G B, et al. Representational learning with ELMs for big data[J]. IEEE Intelligent Systems, 2013, 28(6): 31-34.

作者简介

王雪松(1974—), 女, 教授, 博士生导师, 从事人工智能、机器学习及模式识别等研究, E-mail: wangxuesongcumt@163.com;

赵季娟(1994—), 女, 硕士生, 从事迁移学习的研究, E-mail: 1379709024@qq.com;

程玉虎(1973—), 男, 教授, 博士生导师, 从事机器学习、模式识别与智能系统等研究, E-mail: chengyuhu@163.com.

(责任编辑: 孙艺红)

下 期 要 目

基于模式运动的一类生产过程调节性能与聚类参数关系..... 徐正光, 等

控制输入受限的准全控制利用率有限时间稳定控制..... 蒲 明, 等

基于突发事件的任务计划动态调整模型及算法..... 孙 鹏, 等

基于图正则自适应联合协同表示的高光谱图像分类..... 李冬青, 等

基于面板数据的灰色曲率关联模型..... 吴鸿华, 等

含执行机构未知动态的液压伺服系统输出反馈控制..... 那 靖, 等

基于智能优化算法的Pendubot轨迹规划与控制方法设计..... 王乐君, 等

基于过滤模型的聚类算法..... 邱保志, 等

基于改进火力分配模型的综合防空火力智能优化分配..... 孙海文, 等

一种具有噪声干扰的MIMO非线性时变系统自适应控制..... 张 超, 等

基于密度的模糊代表点聚类算法..... 周 洁, 等