滕少华, 吴泽锋, 滕璐瑶, 等. 置信样本选择与差异性特征增强的域适应[J]. 广东工业大学学报, 2025, 42(3): 12–26. doi: 10.12052/gdutxb.230188.
Teng Shaohua, Wu Zefeng, Teng Luyao, *et al*. Confidence sample selection and specific feature enhancement for domain adaptation[J]. Journal of Guangdong University of Technology, 2025, 42(3): 12–26. doi: 10.12052/gdutxb.230188.

置信样本选择与差异性特征增强的域适应

滕少华[1]，吴泽锋[1]，滕璐瑶[2]，张　巍[1]，曾　莹[1]
（1. 广东工业大学　计算机学院, 广东　广州 510006; 2. 广州番禺职业技术学院　信息工程学院, 广东　广州 511483）

摘要: 域适应学习因其有效减少域差异实现标签传播被广泛应用。目前，大多数域适应学习仅通过线性判别分析增强线性数据的判别性而忽略了现实世界存在非线性数据；同时，这些方法未考虑目标域的低置信样本对训练过程的负影响。因此，本文提出一种新颖的置信样本选择与差异性特征增强的域适应框架(Confidence Sample Selection and Specific Feature Enhancement, CSS-SFE)。首先，该框架通过最小最大原则选择高置信的目标样本来辅助训练，减少不正确伪标签影响；其次，权衡类散点矩阵和邻居散点矩阵的贡献来增强线性和非线性数据集的特征，提高数据集的判别性；接着，分别为源域和目标域的样本学习不同投影矩阵来保持各自的差异性特征，防止源域和目标域的样本区分性降低；再次，进一步应用边缘分布对齐和条件分布对齐减少域分布差异；最后，在多个基准数据集上进行的广泛实验证明该方法优于目前的方法。

关键词: 高置信样本选择；差异性特征增强；域适应；标签传播
中图分类号: TP181　　**文献标志码:** A　　**文章编号:** 1007–7162(2025)03–0012–15

Confidence Sample Selection and Specific Feature Enhancement for Domain Adaptation

Teng Shaohua[1], Wu Zefeng[1], Teng Luyao[2], Zhang Wei[1], Zeng Ying[1]
(1. School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China; 2. School of Information Engineering, Guangzhou Panyu Polytechnic, Guangzhou 511483, China)

Abstract: For domain shifts, domain adaptation (DA) promotes label propagation by reducing distributional differences. In recent work, DA only enhances the discriminability of linear data by linear discriminant analysis, which ignores real-world nonlinear data. Meanwhile, these methods fail to consider the negative impact of the target low-confidence samples on the training process. Therefore, confidence sample selection and specific feature enhancement for domain adaptation (CSS-SFE) is proposed. Firstly, the framework selects target high-confidence samples through the min-max principle to reduce the impact of incorrect pseudo-label during training; Secondly, the class scatter matrix and neighbor scatter matrix are balanced to enhance the features of linear and nonlinear datasets so that the discriminability of the samples is improved; Thirdly, the framework maintains source and target specific features by learning different projection matrices, which prevents the discriminability of the samples decreasing; Fourthly, the marginal and conditional distribution alignment is further applied to reduce the domain distribution discrepancy; Finally, extensive experiments on several benchmark datasets demonstrate the superiority of CSS- SFE over state-of-the-art methods.

Key words: confidence sample selection; specific feature enhancement; domain adaptation; label propagation

传统机器学习算法通常基于一个假设，即训练数据集(源域)和测试数据集(目标域)服从独立同分布。在这一假设下，已经取得了许多有效的研究成果。然而，由于真实数据往往涉及不同的环境、采集

收稿日期：2023-11-21　**录用日期：**2024-03-22　**网络首发日期：**2024-08-02
基金项目：国家自然科学基金资助项目(61972102)；广州市科技计划项目(2023A04J1729)
作者简介：滕少华(1962–)，男，教授，博士生导师，CCF杰出会员，主要研究方向为数据挖掘、协同计算、模式识别，E-mail：shteng@gdut.edu.cn
通信作者：曾莹(1988–)，女，硕士研究生，主要研究方向为优化调度、协同计算、模式识别，E-mail：zengying@gdut.edu.cn

设备和跨平台风格，源域和目标域数据可能不服从独立同分布。因此，不同域非独立同分布的情况成为亟待解决的机器学习问题。近年来，域适应学习在处理非独立同分布数据方面引起了广泛关注。域适应学习通过利用源域和目标域的样本特征进行分布对齐，从而使得源域的标签知识能够被迁移到目标域的无标签数据上。这种方法有效地降低了对目标域进行重复、昂贵的人工标注的需求。目前，一些经典的域适应学习方法包括迁移成分分析(Transfer Component Analysis, TCA)[1]和联合分布对齐(Joint Distribution Adaptation, JDA)[2]。这些方法主要关注于边缘分布对齐和条件分布对齐，以减少域间的位移。

在域适应中，线性判别分析(Linear Discriminant Analysis, LDA)[3]被广泛应用于增强源域和目标域样本的判别性，从而使源域训练的分类器能够为目标域样本生成更可靠的伪标签。可靠的伪标签不仅能更好地指导条件分布对齐，而且有助于标签的正向传播。目前存在多种关于判别性的域适应方法。Li等[4]提出了域不变和类判别特征学习(Domain Invariant and Class Discriminative Feature Learning, DICD)，通过LDA增强源域和目标域样本的可区分性来提高域适应性。Zhang等[5]提出判别联合概率的最大均值差异(Discriminative Joint Probability Maximum Mean Discrepancy, DJP-MMD)，通过提高联合概率分布的可迁移性和判别性来增强域适应效果。Teng等[6]提出的增量置信样本分类(Incremental Confidence Samples Into Classification, ICSC)方法则通过增量标签的方式减少误差累积并提高判别性。此外，Meng等[7]提出了双层自适应和判别知识迁移(Dual-level Adaptive and Discriminative Knowledge Transfer, DLAD)方法，通过无标签目标数据促进分类器的学习。这些方法的共同目标是通过判别性特征学习来增强域适应的效果，从而在目标域上可靠地传递标签信息。

尽管这些方法获得了良好的性能，但忽略了以下3个问题：

(1) 这些方法忽略了目标域中可能存在的低置信样本，这可能导致标签不正确传播，并在迭代过程中不断积累。一些域适应学习方法[3-4]利用源域的分类器为目标域样本生成伪标签，然后利用这些伪标签来指导判别性分析和域的分布对齐。由于域分布差异，源域的分类器可能无法可靠地分类目标域样本，从而导致目标域中产生不可靠的伪标签。这些低置信样本可能会误导线性判别分析和域的分布对齐。更进一步，这些样本的不正确伪标签可能通过迭代过程不断传播，造成错误标签的累积。

(2) 这些方法可能无法有效提高非线性数据集的判别性，甚至可能导致非线性数据集的判别性更差。LDA在同类样本服从非独立同分布时可能会失败，如图1所示，类1样本服从非独立同分布，LDA最小化类内散点矩阵会将样本投影到子空间2中，造成类1和类2判别性低。在非线性数据集中，LDA可能会牺牲服从非独立同分布类别的判别性，来增强服从独立同分布类别的判别性。Zhu等[8]提出邻居线性判别分析(Neighborhood Linear Discriminant analysis, NLDA)，通过邻居散点矩阵来增强非线性数据集的判别性，最大化邻居间散点矩阵使得投影后子类判别性最大化。如图1所示，NLDA将类1看成2个邻居集合(子类)，将其投影到子空间2中，使得子类分别距离类2最远，从而提高非线性数据集的判别性。值得注意的是，NLDA仅考虑子类投影后的紧凑性和邻居间不同类别样本的邻居集合的分离程度，投影后属于同类的子类较分散，这抑制了服从独立同分布的类别判别性，且未考虑类别间的判别性，以至于投影后的类别判别性可能不足。NLDA可能会牺牲类别判别性和独立同分布类别的判别性，来增强服从非独立同分布子类的判别性。

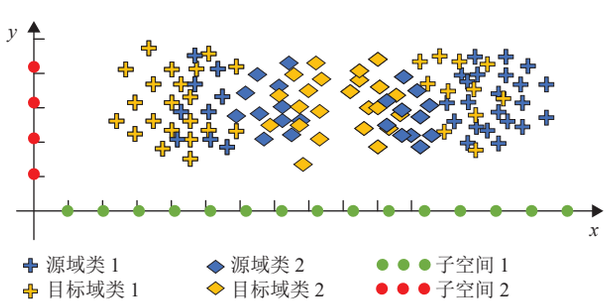


图 1　非线性数据集在LDA和NLDA下的映射图

Fig.1　Mapping of nonlinear datasets under LDA and NLDA

(3) 这些方法忽视了不同域的判别性是有差异的。目前的域适应学习方法[6-7]通常采用相同的投影矩阵来学习源域和目标域的投影子空间，主要关注域间的一致性信息，而忽略了域间的差异性信息。由于存在域差异，不同域的样本判别性可能是不同的。在学习域间一致性的过程中，判别性可能会因此而损失。这种过于严格的一致性约束导致域对齐后样本判别性不足，且对标签的传播不利。域差异的程度不同，域间差异性学习的忽视可能导致样本的判别性不同程度降低。

为了解决上述提到的问题，本文提出了一个新

的域适应框架，即置信样本选择与差异性特征增强。该框架包括3个主要部分：(1) 置信样本选择，选择高置信目标样本进行训练；(2) 差异性特征增强，权衡类散点矩阵和邻居散点矩阵，以及差异性学习增强不同域样本的判别性特征；(3) 跨域对齐和知识迁移，进行边缘分布对齐和条件分布对齐减少域差异，进一步进行标签传播。置信样本选择为差异性特征增强、跨域对齐和知识迁移选择出可靠的样本进行训练。差异性特征增强提高样本的判别性，有利于促进跨域对齐以及知识迁移。跨域对齐可以减少域偏移的影响。本文的主要贡献如下：

(1) 置信样本选择采用伪标签选择策略，在目标域中选择高置信样本。仅高置信的目标样本会被训练，从而有效减少了迭代过程中的误差积累。

(2) 差异性特征增强通过权衡类散点矩阵和邻居散点矩阵增强样本判别性，且差异性学习利用双投影矩阵保留源域和目标域样本判别性。

(3) 跨域对齐和知识迁移减少域的边缘分布差异和条件分布差异，使知识从源域向目标域迁移。

(4) 在多个基准数据集上进行的大量实验表明，所提出的方法优于其他近年的无监督域适应方法。

1 相关工作

1.1 目标样本选择

Wang等[9]提出平衡分布适应(Balanced Distribution Adaptation, BDA)，该方法衡量边缘分布对齐和条件分布对齐的重要性，从而适应不同数据的分布情况。由于BDA没有考虑低置信样本对条件分布对齐的负面影响，标签的传播不可靠。Wang等[10]提出了简易迁移学习(Easy Transfer Learning, Easy-TL)，以降低模型的复杂度并获得竞争力的性能。但是Easy-TL的分类器仅通过源域数据结构来预测目标样本，忽略了目标域原有的数据结构，可能会影响目标样本伪标签的准确性。Peng等[11]提出主动迁移学习(Active Transfer Learning, ATL)，通过重权重避免噪声样本的负面影响。ATL削弱源域噪声样本，但没有选择目标域的高置信样本进行辅助训练。Zhang等[5]提出判别联合概率最大均值差异(DJP-MMD)，用联合分布代替边缘分布和条件分布对齐。DJP-MMD进一步减小了边缘和条件分布对齐的近似误差，不过，未进一步对目标样本进行筛选，可能会令联合分布存在较大的误差。Tian等[12]提出类质心匹配和局部流行学习(Class Centroid Matching and Local Manifold Self-learning, CMMS)，通过匹配目标域和源域类质心来减少域分布差异。CMMS通过k-means聚类目标样本，被错误聚类的样本可能会影响类质心匹配。Xiao等[13]提出用于无监督视觉域适应的标签解耦分析(Label Disentangled Analysis for Unsupervised Visual Domain Adapta-tion, LDADA)，减少标签对特征学习的负面影响。但是，LDADA将伪标签视为样本部分特征进行分布对齐，忽略了伪标签的不可靠性。Meng等[7]提出双层自适应和判别知识迁移(DLAD)，解决训练的分类器可能过度拟合源域的情况。DLAD充分考虑了源域数据结构和目标域数据结构，但目标域分类器的训练受目标低置信样本的影响。这些方法主要通过减少噪声样本来提高样本可区分性和促进分布对齐。值得注意的是，这些方法选择所有目标样本进行训练，而其中包含的低置信样本可能会导致错误，进一步，这些错误会在迭代过程中不断累积。

1.2 线性判别性

Ghifary等[3]提出散点成分分析(Scatter Component Analysis, SCA)，从类的层次提高样本判别性并且从域的层次上减少域差异来促进标签传播。SCA仅从线性数据的角度来学习，忽略了对非线性数据的适用性。Li等[4]提出域不变和类判别特征学习(DICD)，考虑域间的差异性，且分别为源域和目标域样本学习判别性。但是，DICD也存在没有考虑数据集的非线性特性的情况。Teng等[6]提出了增量置信样本分类 (ICSC)，采用增量标记且自适应地调整类内和类间判别性的重要性。ICSC进一步增强了样本的判别性，不过，其仅从线性数据集的角度上考虑改进的判别性学习。Li等[14]提出基于标签校正的渐进分布对齐(Progressive Distribution Aligment Based on Label Correction, PDALC)，利用伪标签校正机制来学习判别子空间。PDALC每次迭代都能够获得更准确的伪标签来指导判别性和域适应学习。Lu等[15]提出了判别不变对齐域适应(Discriminative Invariant Alignment, DIA)，利用域的可判别性和几何结构来处理类质心偏移问题。DIA在学习判别性子空间时保持原始数据的结构。Zhao等[16]提出带密度峰的判别几何和统计对齐(Discriminant Gometrical and Statistical Alignment with Density Peaks, DGSA)，选择密度峰来减少干扰实例的影响。DGSA减少噪声样本的影响同时强化样本的判别特征。PDALC、DIA、DGSA都利用线性判别分析方法来强化样本特征。样本判别性的提高有利于域分布对齐，能更可靠地进行标签传播。然而，当数据集是非线性的时候，这些

方法表现可能不具有优势。

1.3　域间样本一致性与差异性

在子空间学习方面，Wang等[17]提出了流行嵌入分布对齐(Manifold Embedded Distribution Alignment, MEDA)，利用数据的流行结构进行动态分布对齐。MEDA分别为源域和目标域学习流行结构防止数据失真。值得注意的是，MEDA对源域和目标域学习统一的投影矩阵，保留域间的一致性，减少域分布差异。而源域和目标域各自特有的流行结构，可能由于差异性没有被学习而无法保留。Zhang等[18]提出联合判别分布适应和流形正则化(Joint Discriminative Distribution Adaptation and Manifold Regularization, DDAMR)，保持样本的最近邻关系。MEDA流行学习平等看待所有邻居样本，而DDAMR进一步增强了样本近邻的关系。但是这些流行学习表现可能会随着差异性的忽略而降低。Yang等[19]提出李群流行分析(Lie Group Manifold Analysis, LGMA)，首次尝试用李代数流形空间进行域适应。LGMA从另一个新颖的角度来分析数据的流行结构。和MEDA、DDMAR一样，LGMA未考虑域差异的存在，其源域和目标域的流行结构是有差异的。Sun等[20]提出双图联合自适应和特征选择(Joint Adaptive Dual Graph and Feature Selection, ADGFS)，可自适应地更新相似性矩阵。ADGFS防止了高维度数据算法复杂度高的情况，同时也能够在迭代中更好地学习数据的流行结构。源域和目标域的流行结构的差异可能会在ADGFS学习域间一致性中丢失，所以自适应更新相似性矩阵受影响。

2　置信样本选择与差异性特征增强学习框架

2.1　符号和问题定义

本文所提方法的流程图如图2所示，本小节使用的主要符号如表1所示，域适应(Domain Adaptation, DA)问题如下。

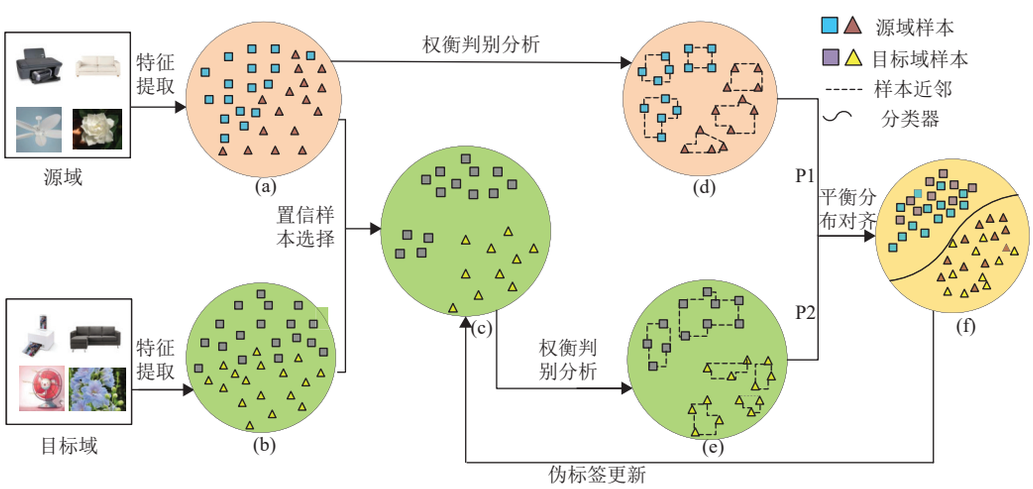


图 2　本文算法的流程图

Fig.2　Flowchart of the algorithm in the paper

$\mathcal{P}_{\boldsymbol{X}_\mathrm{s},\boldsymbol{Y}_\mathrm{s}}$是源域的联合分布。$\mathcal{P}_{\boldsymbol{X}_\mathrm{t},\boldsymbol{Y}_\mathrm{t}}$是目标域的联合分布。由于域分布不同($\mathcal{P}_{\boldsymbol{X}_\mathrm{s},\boldsymbol{Y}_\mathrm{s}} \neq \mathcal{P}_{\boldsymbol{X}_\mathrm{t},\boldsymbol{Y}_\mathrm{t}}$)，在源域上训练的分类器难以准确分类目标域样本。为了解决这个问题，域适应学习域分布对齐子空间，将源域的知识迁移到目标域。具体来说，$\boldsymbol{X}_\mathrm{s}=\{x_{\mathrm{s},i}\}_{i=1}^{n_\mathrm{s}}$是源域的样本，$\boldsymbol{Y}_\mathrm{s}=\{y_{\mathrm{s},i}\}_{i=1}^{n_\mathrm{s}}$是源域的样本标签以及$\boldsymbol{X}_\mathrm{t}=\{x_{\mathrm{t},i}\}_{i=1}^{n_\mathrm{t}}$是目标域样本。域适应是找到一个投影矩阵$\boldsymbol{P}$，将源域特征空间$\boldsymbol{X}_\mathrm{s}$和目标域特征空间$\boldsymbol{X}_\mathrm{t}$投影到子空间中$\boldsymbol{Z}_\mathrm{s}=\boldsymbol{P}^\mathrm{T}\boldsymbol{X}_\mathrm{s}$以及$\boldsymbol{Z}_\mathrm{t}=\boldsymbol{P}^\mathrm{T}\boldsymbol{X}_\mathrm{t}$。然后，在源域的样本上训练分类器$f(\boldsymbol{Z}_\mathrm{s})$来对目标域$\boldsymbol{Z}_\mathrm{t}$进行更准确的分类。

2.2　置信样本选择

目前的域适应学习[7]采用伪标签来指导线性判别分析和条件分布对齐。由于域之间的分布差异，源域的分类器可能会为目标样本生成不可靠的伪标签，从而抑制域的分布对齐的过程。因此，为了克服这一问题，应该选择目标域中的高置信样本进行训练，以确保更可靠的伪标签用于学习线性判别分析和条件分布对齐。

在进行置信样本选择时，需要综合考虑目标样本的数据结构。首先，利用源域的分类器和目标域的分类器对目标域样本进行预测。其次，从这2个分类

表1 主要符号
Table 1 Main notation

符号	表示的意义
$\mathcal{P}_{\boldsymbol{X}_\mathrm{s},\boldsymbol{Y}_\mathrm{s}}/\mathcal{P}_{\boldsymbol{X}_\mathrm{t},\boldsymbol{Y}_\mathrm{t}}$	源域/目标域的联合分布
$\boldsymbol{P}_1/\boldsymbol{P}_2$	源域/目标域的投影矩阵
m	数据集的维度
d	数据集降维后的维度
$n_\mathrm{s}/n_\mathrm{t}$	源域/目标域样本的数量
$\boldsymbol{Y}_\mathrm{s}$	源域真实标签矩阵
$\mathbb{R}$	实数集矩阵
$\widehat{\boldsymbol{Y}}_\mathrm{t}$	目标域伪标签矩阵
$\boldsymbol{X}_\mathrm{s}$	源域样本集合
$\boldsymbol{X}_\mathrm{t}$	目标域样本集合
$\widehat{\boldsymbol{X}}_\mathrm{t}$	目标域高置信样本集合
$\boldsymbol{A}_{\mathrm{TDA\_S}}/\boldsymbol{A}_{\mathrm{TDA\_T}}$	源域/目标域权衡判别分析矩阵
$\boldsymbol{A}_\mathrm{sb}/\boldsymbol{A}_\mathrm{tb}$	源域/目标域类间散点矩阵
$\boldsymbol{A}_\mathrm{sw}/\boldsymbol{A}_\mathrm{tw}$	源域/目标域类内散点矩阵
$\boldsymbol{A}_\mathrm{sn}/\boldsymbol{A}_\mathrm{tn}$	源域/目标域邻居散点矩阵
$\mathbf{Z}_\mathrm{s}/\mathbf{Z}_\mathrm{t}$	映射后的源域/目标域样本集合

器中选择具有一致伪标签预测的样本。接着，采用最小最大原则，从双分类器预测伪标签一致的目标样本中筛选出高置信度的样本。在此过程中，高置信样本的数量应根据类别比例进行选择，以防止类别不平衡的情况。具体步骤如下。

首先，采用源域的数据结构信息来预测目标样本。具体来说，本文方法通过计算目标样本到源域类质心$\boldsymbol{u}_{\mathrm{s},c}$的距离来进行伪标签的预测。通过这个方法，本文得到目标样本到类别C的概率矩阵$\boldsymbol{p}_\mathrm{s}(y|x_\mathrm{t})$。在源域的预测中，伪标签$\hat{y}_\mathrm{s}$为概率矩阵中概率最高$p_{\hat{y}_\mathrm{s}}$的类别。通过式(1)~(2)表示：

$$\boldsymbol{p}_\mathrm{s}(y|x_\mathrm{t})=\frac{\exp(-\|\boldsymbol{P}^\mathrm{T}x_\mathrm{t}-\boldsymbol{P}^\mathrm{T}\boldsymbol{u}_{\mathrm{s},c}\|_\mathrm{F}^2)}{\sum_{c=1}^{C}\exp(-\|\boldsymbol{P}^\mathrm{T}x_\mathrm{t}-\boldsymbol{P}^\mathrm{T}\boldsymbol{u}_{\mathrm{s},c}\|_\mathrm{F}^2)} \tag{1}$$

$$(\hat{y}_\mathrm{s},p_{\hat{y}_\mathrm{s}})=\operatorname{argmax}\boldsymbol{p}_\mathrm{s}(y|x_\mathrm{t}) \tag{2}$$

其次，采用目标域数据结构信息来预测目标样本。具体来说，采用聚类计算目标域的类质心$\boldsymbol{u}_{\mathrm{t},c}$。然后，概率矩阵$\boldsymbol{p}_\mathrm{t}(y|x_\mathrm{t})$通过计算目标样本到$C$类中心的距离得到的。在目标域的预测中，伪标签$\hat{y}_\mathrm{t}$为概率矩阵中概率最高$p_{\hat{y}_\mathrm{t}}$的类别，由式(3)~(4)表示：

$$\boldsymbol{p}_\mathrm{t}(y|x_\mathrm{t})=\frac{\exp(-\|\boldsymbol{P}^\mathrm{T}x_\mathrm{t}-\boldsymbol{P}^\mathrm{T}\boldsymbol{u}_{\mathrm{t},c}\|_\mathrm{F}^2)}{\sum_{c=1}^{C}\exp(-\|\boldsymbol{P}^\mathrm{T}x_\mathrm{t}-\boldsymbol{P}^\mathrm{T}\boldsymbol{u}_{\mathrm{t},c}\|_\mathrm{F}^2)} \tag{3}$$

$$(\hat{y}_\mathrm{t},p_{\hat{y}_\mathrm{t}})=\operatorname{argmax}\boldsymbol{p}_\mathrm{t}(y|x_\mathrm{t}) \tag{4}$$

最后，采用最小最大原则来选择高置信样本。由于域差异的存在，仅研究源域的结构信息可能导致较大的分类错误。因此，Wang等[21]提出基于伪标签选择的结构化预测(Structured Prediction Based Selective Pseudo-labeling, SPL)，引入了目标域分类器，该分类器利用目标域的结构信息进行无监督聚类。在SPL中，选择了2个分类器中预测概率较高的那个作为目标样本的标签。当2个分类器预测的类别不一致时，仅选择概率较高的标签可靠性不足。为了最大化伪标签的可靠性，2个分类器的预测标签应该一致，并且都有较高的预测概率。首先，在初步选择被2个域分类器标记为同一类的样本后，按照2个分类器对同一样本预测概率的较低值进行排序，排名靠前的样本表示2个分类器对该样本都有较高的预测值，因此选择这些样本作为高置信样本参与域适应学习，即采用最小最大原则。具体来说，高置信样本从伪标签预测一致的样本中挑选，高置信样本的数量采用类别比例$\left(vn_\mathrm{t}^c\right)/T$来防止类别不平衡，其中$v$，$T$和$n_\mathrm{t}^c$分别代表当前迭代次数、总迭代次数和第$C$类的数量。然后，采用最小最大原则来选择高置信样本$\hat{x}_\mathrm{t}$，即从伪标签预测一致且预测概率较低的样本中，按概率对样本进行排序，每个类中选择最高的$\left(vn_\mathrm{t}^c\right)/T$数量的置信样本组成高置信目标样本集合，高置信目标样本可以通过式(5)获得：

$$\begin{aligned}(\widehat{\boldsymbol{X}}_\mathrm{t},\widehat{\boldsymbol{Y}}_\mathrm{t})=\bigcup_{c=1}^{c}\operatorname*{argmax}_{p_i}\Bigg\{&x_i|x_i\in\boldsymbol{X}_\mathrm{t}^c(\hat{y}_\mathrm{s}=\hat{y}_\mathrm{t}),\\&p_i\in\min_{\hat{y}_\mathrm{s}=\hat{y}_\mathrm{t}}(p_{\hat{y}_\mathrm{s}},p_{\hat{y}_\mathrm{t}}),1\leqslant i\leqslant\frac{vn_\mathrm{t}^c}{T}\Bigg\}\end{aligned} \tag{5}$$

通过高置信样本的选择，目标域不再是所有样本参与训练，有效降低了低置信目标样本所带来的影响。接下来，将高置信样本辅助学习判别性分析和分布对齐。

2.3 差异性特征增强

2.3.1 权衡判别分析

域适应方法[22-23]采用线性判别分析(LDA)增强源域和目标域的判别性。通过提高伪标签的准确性促进条件分布对齐。LDA的定义如下

定义1 线性判别分析(LDA) 假设$\boldsymbol{X}\in\mathbb{R}^{m\times n}$是一个样本集，其标签为$\boldsymbol{Y}\in\mathbb{R}^n$，线性判别分析定义为

$$\max_{\boldsymbol{P}}\boldsymbol{A}=\max_{\boldsymbol{P}}\mathrm{Tr}(\boldsymbol{P}^\mathrm{T}(\boldsymbol{A}_\mathrm{b}-\boldsymbol{A}_\mathrm{w})\boldsymbol{P}) \tag{6}$$

式中：$\boldsymbol{A}_\mathrm{b}=\sum_{c=1}^{C}n^{(c)}(m^{(c)}-\overline{m})(m^{(c)}-\overline{m})^\mathrm{T}$是类间散点矩阵，$m^c$是第$c$类的质心，$\overline{m}$是所有样本的质心。$\boldsymbol{A}_\mathrm{w}=\sum_{c=1}^{C}\boldsymbol{X}^{(c)}\boldsymbol{H}^{(c)}(\boldsymbol{X}^{(c)})^\mathrm{T}$是类内散点矩阵。$\boldsymbol{H}^{(c)}=\boldsymbol{I}^{(c)}-(\mathbf{1}^{(c)}(\mathbf{1}^{(c)})^\mathrm{T})/n^{(c)}$是第$C$类数据的中心矩阵。$\boldsymbol{I}^{(c)}$是单位

矩阵。$\mathbf{1}^{(c)}$是全为τ的列向量。

LDA最小化类内散点矩阵和最大化类间散点矩阵，使得样本投影后同一种类别数据的投影点尽可能接近，而不同类别中心之间的距离尽可能大，从而增强样本的判别性。而类内散点矩阵最小化仅在类内样本服从独立同分布时有效，因此，当类内样本服从非独立同分布时，邻居线性判别分析被提出，其定义如下。

定义2　邻居线性判别分析(NLDA)　假设$\boldsymbol{X}\in\mathbb{R}^{m\times n}$是一个样本集，其标签为$\boldsymbol{Y}\in\mathbb{R}^{n}$，邻居线性判别分析定义为

$$\boldsymbol{A}_n=\sum_{i=1}^{n}\sum_{x_j\in\mathrm{KNN}(x_i),y_i=y_j}(x_j-\overline{m}_i)(x_j-\overline{m}_i)^{\mathrm{T}}-\sum_{i=1}^{n}\sum_{x_j\in\mathrm{KNN}(x_i),y_i\neq y_j}(\overline{m}_i-\overline{m}_j)(\overline{m}_i-\overline{m}_j)^{\mathrm{T}}\tag{7}$$

式中：第一项是邻居内散点矩阵，x_j是样本i的最近邻，$\overline{m}_i$是样本i最近邻集合的质心。最小化邻居内散点矩阵让子类内的样本在投影后更接近。第二项是邻居间散点矩阵，样本j是样本i的最近邻且类别不同，$\overline{m}_j$是样本j的最近邻且类别相同的集合质心，通过最大化邻居间散点矩阵，让子类样本在投影后和最近不同类的样本距离更远，从而增强非线性数据集判别性。

值得注意的是，NLDA仅考虑邻居间不同类别的最近邻集合的差异，并未考虑整体类别差异，且对于线性数据来说，NLDA未考虑类内子类的聚合。为了最大化样本的判别性，本文提出了权衡判别分析。

定义3　权衡判别分析(Trade-off Discriminant Analysis, TDA)　假设$\boldsymbol{X}\in\mathbb{R}^{m\times n}$是一个样本集，其标签为$\boldsymbol{Y}\in\mathbb{R}^{n}$，权衡判别分析定义为

$$\max_{\boldsymbol{P}}\boldsymbol{A}_{\mathrm{TDA}}=\max_{\boldsymbol{P}}\mathrm{Tr}(\boldsymbol{P}^{\mathrm{T}}(\beta(\boldsymbol{A}_{\mathrm{b}}-\boldsymbol{A}_{\mathrm{w}})-\theta\boldsymbol{A}_{\mathrm{n}})\boldsymbol{P})\tag{8}$$

式中：$\boldsymbol{A}_{\mathrm{b}}$和$\boldsymbol{A}_{\mathrm{w}}$是类间散点矩阵和类内散点矩阵，$\boldsymbol{A}_{\mathrm{n}}$是邻居散点矩阵。$\beta$和$\theta$是超参数，用于权衡线性和非线性数据的可区分性。

面对数据集复杂的分布情况时，可以通过权衡类散点矩阵和邻居散点矩阵来增强样本判别性：(1) 当数据集是线性分布时，数据集通过最大化类间散点矩阵和最小化类内散点矩阵投影后，不同类别间距离最大化和同类别样本距离最小化，样本判别性被提高。最大化邻居间散点矩阵使得邻居间不同类距离更远，最小化邻居散点矩阵使得子类更紧凑且保持局部关系，防止特征空间被扭曲，进一步增强数据集的判别性。(2) 当数据集是非线性分布时，通过最大化邻域间散点矩阵将样本投影邻居间不同类别距离更远的子空间上，使得子类具有判别性，且最小化邻居内散点矩阵使得子类内样本更聚集。非线性数据集包含服从非独立同分布的类别，也可能包含服从独立同分布的类别，如图1所示，应该最大化邻居间散点矩阵保持子类的判别性的同时，最小化类内散点矩阵使服从独立同分布类别的样本紧密聚合，从而所有样本判别性被最大化。由于数据分布情况不同，可以通过调节类散点矩阵和邻居散点矩阵的超参数让样本投影后判别性最大。

权衡判别分析提高线性和非线性数据集的判别性。值得注意的是，由于域差异，源域和目标域的判别性往往不同。尽管增强了数据的判别特征，但也可能因为仅保留域间的一致性信息而丢失。因此，进一步提出了差异性学习。

2.3.2　差异性学习

PDALC仅使用一个统一的投影矩阵为源域和目标域学习判别性分布对齐子空间，其保留域间样本的一致性是为了减少域分布差异。值得注意的是，不同域中样本的判别性是不同的，域间样本的差异性被忽略会影响样本的判别性。尽管学习了判别性分析，但也会因为忽略域间样本差异性而丢失。为了提高源样本和目标样本的判别性，应学习域间样本的差异性。具体来说，源域和目标域分别学习投影矩阵$\boldsymbol{P}_1$和$\boldsymbol{P}_2$。$\boldsymbol{P}_1$和$\boldsymbol{P}_2$松弛仅学习域间的一致性信息的约束。源域和目标域的判别性在子空间中保留。式(7)应改写为

$$\max_{\boldsymbol{P}_1}\boldsymbol{A}_{\mathrm{TDA\_S}}=\max_{\boldsymbol{P}_1}\mathrm{Tr}(\boldsymbol{P}_1^{\mathrm{T}}(\beta(\boldsymbol{A}_{\mathrm{sb}}-\boldsymbol{A}_{\mathrm{sw}})-\theta\boldsymbol{S}_{\mathrm{sn}})\boldsymbol{P}_1)\tag{9}$$

$$\max_{\boldsymbol{P}_2}\boldsymbol{A}_{\mathrm{TDA\_T}}=\max_{\boldsymbol{P}_2}\mathrm{Tr}(\boldsymbol{P}_2^{\mathrm{T}}(\mu(\boldsymbol{A}_{\mathrm{tb}}-\boldsymbol{A}_{\mathrm{tw}})-\eta\boldsymbol{A}_{\mathrm{tn}})\boldsymbol{P}_2)\tag{10}$$

式中：$\boldsymbol{A}_{\mathrm{TDA_S}}$是权衡源域类散点矩阵和邻居散点矩阵，$\boldsymbol{A}_{\mathrm{TDA_T}}$是权衡目标域类散点矩阵和邻居散点矩阵；$\boldsymbol{A}_{\mathrm{sb}}$和$\boldsymbol{A}_{\mathrm{sw}}$为源域类散点矩阵；$\boldsymbol{A}_{\mathrm{sn}},\boldsymbol{A}_{\mathrm{tn}}$是源域和目标域的邻居散点矩阵；$\beta$和$\theta$是权衡超参数；$\boldsymbol{A}_{\mathrm{tb}}$和$\boldsymbol{A}_{\mathrm{tw}}$是目标域类散点矩阵；$\mu$和$\eta$是权衡超参数。

同时，差异性学习也要学习域间的一致性信息，以减少域间差异。具体来说，应使2个投影矩阵最小化，以调整域分布

$$\min_{\boldsymbol{P}_1,\boldsymbol{P}_2}\|\boldsymbol{P}_1-\boldsymbol{P}_2\|_{\mathrm{F}}^2\tag{11}$$

差异性学习可以在学习判别子空间的时候，保持源域和目标域各自的判别性，而同时也学习域间

的一致性知识，减少域分布差异。其不会由于仅学习域间一致性知识而导致分布对齐后的样本判别性不足，可以进一步促进跨域对齐和知识迁移。

2.4 跨域对齐与知识迁移

通过置信样本选择和特异性特征增强，源域和目标域的样本判别性提高，有助于获取更精确的伪标签。进一步，为了将源域知识迁移到目标域，本文采用边缘分布和条件分布对齐较少域分布差异。因为样本的判别性提高，通过跨域对齐，知识可以更好地迁移到目标域。由于不同数据集的差异，边缘分布和条件分布的重要性也应得到强调。因此，采用了平衡分布自适应权衡边缘分布和条件分布，如式(12)所示：

$$\min_{\boldsymbol{P}_1,\boldsymbol{P}_2}(1-\tau)D(\mathcal{P}(\boldsymbol{X}_\mathrm{s}),\mathcal{P}(\widehat{\boldsymbol{X}}_\mathrm{t}))+\tau D(\mathcal{P}(y_\mathrm{s}|\boldsymbol{X}_\mathrm{s}),\mathcal{P}(y_\mathrm{t}|\widehat{\boldsymbol{X}}_\mathrm{t})) \quad (12)$$

式中：τ是权衡参数，$D(\mathcal{P}(\boldsymbol{X}_\mathrm{s}),\mathcal{P}(\widehat{\boldsymbol{X}}_\mathrm{t}))$是源域和目标域边缘分布对齐，$D(\mathcal{P}(y_\mathrm{s}|\boldsymbol{X}_\mathrm{s}),\mathcal{P}(y_\mathrm{t}|\widehat{\boldsymbol{X}}_\mathrm{t}))$是对齐源域和目标域条件分布。上式可写成

$$\begin{aligned}\min_{\boldsymbol{P}_1,\boldsymbol{P}_2}(1-\tau)\|\frac{1}{n_\mathrm{s}}\boldsymbol{P}_1^\mathrm{T}x_i-\frac{1}{n_\mathrm{t}}\sum_{x_j\in\boldsymbol{X}_\mathrm{t}}\boldsymbol{P}_2^\mathrm{T}\widehat{x}_j\|_\mathrm{F}^2+\\ \tau\sum_{c=1}^{C}\|\frac{1}{n_\mathrm{s}^{(c)}}\boldsymbol{P}_1^\mathrm{T}x_i-\frac{1}{n_\mathrm{t}^{(c)}}\boldsymbol{P}_2^\mathrm{T}\widehat{x}_j\|_\mathrm{F}^2\end{aligned} \quad (13)$$

式中：$x_i\in\boldsymbol{X}_\mathrm{s}^{(c)}$和$\widehat{x}_j\in\widehat{\boldsymbol{X}}_\mathrm{t}^{(c)}$表示源域和高置信目标域中属于$C$类的样本；$n_\mathrm{s}^{(c)}$和$\widehat{n}_\mathrm{t}^{(c)}$表示源域和高置信目标域中$C$类样本的数量。将式(13)简写为

$$\min_{\boldsymbol{P}_1,\boldsymbol{P}_2}\mathrm{Tr}\left(\begin{bmatrix}\boldsymbol{P}_1^\mathrm{T} & \boldsymbol{P}_2^\mathrm{T}\end{bmatrix}\begin{bmatrix}\boldsymbol{M}_\mathrm{s} & \boldsymbol{M}_\mathrm{st}\\ \boldsymbol{M}_\mathrm{ts} & \boldsymbol{M}_\mathrm{t}\end{bmatrix}\begin{bmatrix}\boldsymbol{P}_1\\ \boldsymbol{P}_2\end{bmatrix}\right) \quad (14)$$

$$\begin{aligned}&\boldsymbol{M}_\mathrm{s}=\boldsymbol{X}_\mathrm{s}((1-\tau)\boldsymbol{L}_\mathrm{s}+\tau\sum_{c=1}^{C}\boldsymbol{L}_\mathrm{s}^{(c)})\boldsymbol{X}_\mathrm{s}^\mathrm{T},\\ &\boldsymbol{L}_\mathrm{s}=\frac{1}{n_\mathrm{s}^2}\mathbf{1}_\mathrm{s}\mathbf{1}_\mathrm{s}^\mathrm{T},\\ &(\boldsymbol{L}_\mathrm{s}^{(c)})_{ij}=\begin{cases}\dfrac{1}{(n_\mathrm{s}^{(c)})^2},x_i,x_j\in\boldsymbol{X}_\mathrm{s}^{(c)}\\ 0,\text{otherwise}\end{cases}\end{aligned} \quad (15)$$

$$\begin{aligned}&\boldsymbol{M}_\mathrm{t}=\widehat{\boldsymbol{X}}_\mathrm{t}\left((1-\tau)\boldsymbol{L}_\mathrm{t}+\tau\sum_{c=1}^{C}\boldsymbol{L}_\mathrm{t}^{(c)}\right)\widehat{\boldsymbol{X}}_\mathrm{t}^\mathrm{T},\\ &\boldsymbol{L}_\mathrm{t}=\frac{1}{n_\mathrm{t}^2}\mathbf{1}_\mathrm{t}\mathbf{1}_\mathrm{t}^\mathrm{T},\\ &(\boldsymbol{L}_\mathrm{t}^{(c)})_{ij}=\begin{cases}\dfrac{1}{(\widehat{n}_\mathrm{t}^{(c)})^2},\widehat{x}_i,\widehat{x}_j\in\widehat{\boldsymbol{X}}_\mathrm{t}^{(c)}\\ 0,\text{otherwise}\end{cases}\end{aligned} \quad (16)$$

$$\begin{aligned}&\boldsymbol{M}_\mathrm{st}=\boldsymbol{X}_\mathrm{s}\left((1-\tau)\boldsymbol{L}_\mathrm{st}+\tau\sum_{c=1}^{C}\boldsymbol{L}_\mathrm{st}^{(c)}\right)\widehat{\boldsymbol{X}}_\mathrm{t}^\mathrm{T},\\ &\boldsymbol{L}_\mathrm{st}=-\frac{1}{n_\mathrm{s}\widehat{n}_\mathrm{t}}\mathbf{1}_\mathrm{s}\mathbf{1}_\mathrm{t}^\mathrm{T},\\ &(\boldsymbol{L}_\mathrm{st}^{(c)})_{ij}=\begin{cases}-\dfrac{1}{n_\mathrm{s}^{(c)}\widehat{n}_\mathrm{t}^{(c)}},x_i\in\boldsymbol{X}_\mathrm{s}^{(c)},\widehat{x}_j\in\widehat{\boldsymbol{X}}_\mathrm{t}^{(c)}\\ 0,\text{otherwise}\end{cases}\end{aligned} \quad (17)$$

$$\begin{aligned}&\boldsymbol{M}_\mathrm{ts}=\widehat{\boldsymbol{X}}_\mathrm{t}\left((1-\tau)\boldsymbol{L}_\mathrm{ts}+\tau\sum_{c=1}^{C}\boldsymbol{L}_\mathrm{ts}^{(c)}\right)\boldsymbol{X}_\mathrm{s}^\mathrm{T},\\ &\boldsymbol{L}_\mathrm{ts}=-\frac{1}{n_\mathrm{s}\widehat{n}_\mathrm{t}}\mathbf{1}_\mathrm{s}\mathbf{1}_\mathrm{t}^\mathrm{T},\\ &(\boldsymbol{L}_\mathrm{ts}^{(c)})_{ij}=\begin{cases}-\dfrac{1}{n_\mathrm{s}^{(c)}\widehat{n}_\mathrm{t}^{(c)}},x_j\in\boldsymbol{X}_\mathrm{s}^{(c)},\widehat{x}_i\in\widehat{\boldsymbol{X}}_\mathrm{t}^{(c)}\\ 0,\text{otherwise}\end{cases}\end{aligned} \quad (18)$$

本文框架基于置信样本选择、特异性特征增强和跨域对齐为源域和目标域学习判别域不变表示$(\boldsymbol{Z}_\mathrm{s},\boldsymbol{Z}_\mathrm{t})$。接着，选择了1-最近邻分类器(1-Nearest neighbor, 1-NN)作为基分类器，其能够利用源域的样本$\boldsymbol{Z}_\mathrm{s}$和源域的标签$\boldsymbol{Y}_\mathrm{s}$来训练1-NN。然后，通过1-NN分类目标域样本$\boldsymbol{Z}_\mathrm{t}$，将源域的知识迁移到目标域。因此，可以获得更可靠的伪标签去减少人工标注的成本。最后在迭代过程中，伪标签又能指导特异性特征增强和跨域对齐。

2.5 总目标函数

为了控制目标投影矩阵的复杂度，本文限制$\mathrm{Tr}\left(\boldsymbol{P}_2^\mathrm{T}\boldsymbol{P}_2\right)$较小。具体来说，本文的目标是通过求解式(19)找到2个投影$\boldsymbol{P}_1$和$\boldsymbol{P}_2$学习域不变表示。将式(9)~(11)和(13)整合为整体目标函数：

$$\max_{\boldsymbol{P}_1,\boldsymbol{P}_2}\frac{\mathrm{Tr}\left(\begin{bmatrix}\boldsymbol{P}_1^\mathrm{T} & \boldsymbol{P}_2^\mathrm{T}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{\mathrm{TDA\_S}} & 0\\ 0 & \boldsymbol{A}_{\mathrm{TDA\_T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{P}_1\\ \boldsymbol{P}_2\end{bmatrix}\right)}{\mathrm{Tr}\left(\begin{bmatrix}\boldsymbol{P}_1^\mathrm{T} & \boldsymbol{P}_2^\mathrm{T}\end{bmatrix}\begin{bmatrix}\boldsymbol{M}_\mathrm{s}+\lambda\boldsymbol{I} & \boldsymbol{M}_\mathrm{st}-\lambda\boldsymbol{I}\\ \boldsymbol{M}_\mathrm{ts}-\lambda\boldsymbol{I} & \boldsymbol{M}_\mathrm{t}+(\lambda+\gamma)\boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{P}_1\\ \boldsymbol{P}_2\end{bmatrix}\right)} \quad (19)$$

式中：$\boldsymbol{I}\in\mathbb{R}^{d\times d}$是同一矩阵，$\gamma$和$\lambda$是正则化参数。

2.6 优化

为了便于优化式(19)，将$[\boldsymbol{P}_1\boldsymbol{P}_2]$简写为$\boldsymbol{W}$，即$\boldsymbol{W}=[\boldsymbol{P}_1\boldsymbol{P}_2]$，然后，目标函数和相应的约束条件可以改写为

$$\max_{\boldsymbol{W}}\frac{\mathrm{Tr}\left(\boldsymbol{W}^\mathrm{T}\begin{bmatrix}\boldsymbol{A}_{\mathrm{TDA\_S}} & 0\\ 0 & \boldsymbol{A}_{\mathrm{TDA\_T}}\end{bmatrix}\boldsymbol{W}\right)}{\mathrm{Tr}\left(\boldsymbol{W}^\mathrm{T}\begin{bmatrix}\boldsymbol{M}_\mathrm{s}+\lambda\boldsymbol{I} & \boldsymbol{M}_\mathrm{st}-\lambda\boldsymbol{I}\\ \boldsymbol{M}_\mathrm{ts}-\lambda\boldsymbol{I} & \boldsymbol{M}_\mathrm{t}+(\lambda+\gamma)\boldsymbol{I}\end{bmatrix}\boldsymbol{W}\right)} \quad (20)$$

请注意，目标函数目的是学习$\boldsymbol{W}$，对$\boldsymbol{W}$的规模调整是不变的。因此，将目标函数重写为

$$\max_{\boldsymbol{W}}\mathrm{Tr}\left(\boldsymbol{W}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{A}_{\mathrm{TDA\_S}} & 0\\ 0 & \boldsymbol{A}_{\mathrm{TDA\_T}}\end{bmatrix}\boldsymbol{W}\right),$$

$$\text{s.t. }\mathrm{Tr}\left(\boldsymbol{W}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{M}_{\mathrm{s}}+\lambda\boldsymbol{I} & \boldsymbol{M}_{\mathrm{st}}-\lambda\boldsymbol{I}\\ \boldsymbol{M}_{\mathrm{ts}}-\lambda\boldsymbol{I} & \boldsymbol{M}_{\mathrm{t}}+(\lambda+\gamma)\boldsymbol{I}\end{bmatrix}\boldsymbol{W}\right)=1 \quad (21)$$

为求得满足约束的最优解，通过拉格朗日函数求解模型

$$\boldsymbol{L}=\mathrm{Tr}\left(\boldsymbol{W}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{A}_{\mathrm{TDA\_S}} & 0\\ 0 & \boldsymbol{A}_{\mathrm{TDA\_T}}\end{bmatrix}\boldsymbol{W}\right)+\mathrm{Tr}\left(\left(\boldsymbol{W}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{M}_{\mathrm{s}}+\lambda\boldsymbol{I} & \boldsymbol{M}_{\mathrm{st}}-\lambda\boldsymbol{I}\\ \boldsymbol{M}_{\mathrm{ts}}-\lambda\boldsymbol{I} & \boldsymbol{M}_{\mathrm{t}}+(\lambda+\gamma)\boldsymbol{I}\end{bmatrix}\boldsymbol{W}-\boldsymbol{I}\right)\boldsymbol{\Phi}\right) \quad (22)$$

通过设置导数$\partial L/\partial W=0$可以得到

$$\begin{bmatrix}\boldsymbol{A}_{\mathrm{TDA\_S}} & 0\\ 0 & \boldsymbol{A}_{\mathrm{TDA\_T}}\end{bmatrix}\boldsymbol{W}=\begin{bmatrix}\boldsymbol{M}_{\mathrm{s}}+\lambda\boldsymbol{I} & \boldsymbol{M}_{\mathrm{st}}-\lambda\boldsymbol{I}\\ \boldsymbol{M}_{\mathrm{ts}}-\lambda\boldsymbol{I} & \boldsymbol{M}_{\mathrm{t}}+(\lambda+\gamma)\boldsymbol{I}\end{bmatrix}\boldsymbol{W}\boldsymbol{\Phi} \quad (23)$$

式中：$\boldsymbol{\Phi}=\mathrm{diag}(\lambda_1,\cdots,\lambda_k)$是前$k$个特征值；$\boldsymbol{W}=[\boldsymbol{W}_1,\cdots,\boldsymbol{W}_k]$包含相应的特征向量，可通过广义特征值分解进行解析求解。得到变换矩阵$\boldsymbol{W}$后，就可以得到$\boldsymbol{P}_1$和$\boldsymbol{P}_2$。为了更好地说明问题，CSS-SFE算法过程如算法1所示。

算法1　CSS-SFE算法

1) 输入：$\boldsymbol{X}_{\mathrm{s}},\boldsymbol{X}_{\mathrm{t}}$：源域和目标域样本；

$\boldsymbol{Y}_{\mathrm{s}}$：源域的标签；

$\beta,\mu,\theta,\eta,\tau$：超参数；$\lambda,\gamma$：正则化参数。

2) 初始化：

(a) $\boldsymbol{A}_{\mathrm{TDA_S}},\boldsymbol{A}_{\mathrm{TDA_T}},\boldsymbol{M}_{\mathrm{s}},\boldsymbol{M}_{\mathrm{t}},\boldsymbol{M}_{\mathrm{st}}$和$\boldsymbol{M}_{\mathrm{ts}}$根据式(9), (10), (15), (16), (17)和(18)进行初始化；

利用$\boldsymbol{X}_{\mathrm{s}}$和$\widehat{\boldsymbol{X}}_{\mathrm{t}}$训练获取伪标签$\widehat{\boldsymbol{Y}}_{\mathrm{t}}$；

(b) 求解方程 (23)中的广义特征分解问题，选择前k个特征值对应的k个特征向量作为变换矩阵$\boldsymbol{W}$，获取源域和目标域各自的变化矩阵$\boldsymbol{P}_1$和$\boldsymbol{P}_2$；

(c) 将原始数据映射到相应的子空间，得到变换后的表示：$\boldsymbol{Z}_{\mathrm{s}}=\boldsymbol{P}_1^{\mathrm{T}}\boldsymbol{X}_{\mathrm{s}},\boldsymbol{Z}_{\mathrm{t}}=\boldsymbol{P}_2^{\mathrm{T}}\boldsymbol{X}_{\mathrm{t}}$；

(d) 在$\{\boldsymbol{Z}_{\mathrm{s}},\boldsymbol{Y}_{\mathrm{s}}\}$上训练分类器$f$，以更新目标域$\widehat{\boldsymbol{Y}}_{\mathrm{t}}=f(\boldsymbol{Z}_{\mathrm{t}})$中的伪标签；

(e) 根据式(15)、(16)、(17)和(18)更新$\boldsymbol{M}_{\mathrm{s}},\boldsymbol{M}_{\mathrm{t}},\boldsymbol{M}_{\mathrm{st}}$和$\boldsymbol{M}_{\mathrm{ts}}$；

(f) 根据式(10)更新$\boldsymbol{A}_{\mathrm{TDA_T}}$。

(g) $v=v+1$

3) 直至$v=T$

4) 输出：$\boldsymbol{P}_1,\boldsymbol{P}_2$：投影矩阵；

$\widehat{\boldsymbol{Y}}_{\mathrm{t}}$：目标域样本$\widehat{\boldsymbol{X}}_{\mathrm{t}}$的伪标签

2.7　时间复杂度

本小节将分析算法1的时间复杂度。优化过程中各部分的复杂度如下：

(1) 构造矩阵$\boldsymbol{A}_{\mathrm{TDA_S}}$和$\boldsymbol{A}_{\mathrm{TDA_T}}$的复杂度分别为$O(2n_{\mathrm{s}}^2+n_{\mathrm{s}}k^2)$和$O(2n_{\mathrm{t}}^2+n_{\mathrm{t}}k^2)$；

(2) 构造矩阵$\boldsymbol{M}_{\mathrm{s}},\boldsymbol{M}_{\mathrm{t}},\boldsymbol{M}_{\mathrm{st}},\boldsymbol{M}_{\mathrm{ts}}$的复杂度为$O(n^2)$；

(3) 求解广义特征值分解问题的复杂度为$O(m^3)$。

假设T为迭代次数，则该算法的总体复杂度$O(T(n_{\mathrm{s}}k^2+n_{\mathrm{t}}k^2+2n_{\mathrm{s}}^2+3n_{\mathrm{t}}^2+2n^2+m^3))\approx O(T(nk^2+n^2+m^3))$。

3　实验验证

3.1　数据集

在本节中，采用了5个基于领域适应的流行数据集：Office + Caltech10 (decaf)[24]、Office31 (ResNet50)[25]、ImageCLEF-DA (ResNet50)[26]、COIL20 [27]和Office-Home[28]。表2简要列出了这5个公共数据集。其中，特征对应维度m，样本数对应$n_{\mathrm{s}}/n_{\mathrm{t}}$，类别数对应$C$。

表 2　数据集概览

Table 2　Overview of data sets

数据集	域	特征	类别数	样本数
Office+ Caltech10	A, C, D, W	Decaf6 (4 096)	10	1 123/958/295/157
Office31	A, D, W	ResNet50 (2 048)	31	2 817/498/795
ImageCLEF-DA	C, I, P	ResNet50 (2 048)	12	600/600/600
COIL20	COIL1, COIL2	SURF (1 024)	20	720/720
Office-Home	Ar, Cl, Pr, Rw	ResNet50 (2 048)	65	2 421/4 379/4 428/4 357

Office+Caltech10(decaf)[24]所述数据集被广泛用于域适应。该数据集包含4个域：加州理工学院(C)、亚马逊(A)、网络摄像头(W)和数码单反相机(D)。Office+Caltech10 数据集是通过合并Office-31 (由 A、W和D中的31个类组成)和Caltech-256 (由C中的256个类组成)创建的。即使是同一类别，各域的数据分布也不尽相同。

Office31(ResNet50)[25]数据集包含31个对象类别，分为3个领域：亚马逊(A)、数码单反相机(D)和网络摄像头(C)。这些类别中的物体通常出现在办公室环境中，如笔记本电脑、键盘和文件柜。亚马逊每个类别平均包含90张图片，共计2 817张图片。数码单反相机领域包含498张高分辨率图片(4 288×2 848)，每个类别包含5个对象，每个对象平均从3个不同的视角拍摄。网络摄像头包含795幅低分辨率图像(640×480)。

ImageCLEF-DA (ResNet50)[26]数据集是ImageCLEF-DA领域适应性挑战赛的基准数据，该挑战赛有3个领域：Caltech-256 (C)、ImageNet ILSVRC 2012 (I)和Pascal VOC 2012 (P)。每个领域有12个类别，每个类别有50幅图像。

COIL20[27]由哥伦比亚大学维护数据库包含20个不同的物体。每个物体以5°为增量，水平旋转360°从72个不同角度拍摄。因此，每个物体可获得72幅图像，共计1 440幅图像。数据集包括2个子集，其中第1个子集包括10个物体的720张未经处理的图像，第2个子集包括10个不同主体的720张经过处理的图像。

Office-Home[28]数据集包含65种不同对象，共有30 475个原始样本，由4个子域组成：Art(Ar)，Clipart(Cl)，Product(Pr)和Real-World(Rw)。这些子域的大小分别为2 427，4 365，4 439和4 357。

3.2 对比方法

为了便于比较，本文介绍了8种浅层的迁移学习域适应方法和3种深度的迁移学习域适应方法：

平衡分布域适应(BDA)[9]，着眼于平衡边缘分布和条件分布的相对重要性。BDA的优势在于能够处理不同数据集上不同分布的情况。

流形嵌入分布对齐域适应(MEDA)[17]旨在在格拉斯曼流形中学习域不变分类器，并实现结构风险最小化。

容易迁移学习(EasyTL)[10]，其利用域内编码来探索域内结构。

可区分联合概率最大均值差异的域适应(JPDA)[5]，该文提出用联合概率来替代域适应中常用的最大均值差异(Maximum Mean Discrepancy, MMD)实现边缘分布和条件分布对齐。

联合区分子空间学习和无监督迁移分类(Joint Distinct Subspace Learning and Unsupervised Transfer Classification, JDSC)[23]旨在减少几何和统计位移，然后学习自适应分类器，以提高伪标签的准确性。

用于无监督视觉域适应的标签解耦分析(LDADA)[13]，利用样本移除特征和标签重构特征，以防止伪标签误导域适应。

将增量置信样本纳入分类的域适应(ICSC)[6]，动态评估转移和区分的重要性以及用增强方式选择置信样本。

双层自适应和判别域适应(DLAD)[7]，通过考虑2个领域的结构风险最小化(Structural Risk Minimization, SRM)和执行加权分布自适应来实现域自适应分类器。DLAD促进了联合分类器的学习。

可迁移对抗训练(Transferable Adversarial Training, TAT)[29]，该方法生成可迁移的示例以填补源领域和目标领域之间的差距，并通过对这些可迁移示例进行对抗训练，使深度分类器在这些示例上进行一致的预测。

用于图像分类的深度子域自适应网(Deep SubDomain Adaptation Network, DSAN)[30]，通过在不同领域的特定于领域的层激活的相关子域分布上对齐，利用局部最大均值差异(Local Maximum Mean Difference, LMMD)来学习一个迁移网络。

用于领域自适应的最大密度散度(Adversarial Tight Match, ATM)[31]，提出了一种新颖的距离损失，称为最大密度散度(Maximum Density Divergence, MDD)，用于量化分布差异。

3.3 结果与讨论

如表3~7所示，本文的方法在实验中与几种最先

表3 在Office + Caltech 10 (decaf)数据集上的分类精度

Table 3 Classification accuracy on Office + Caltech 10 (decaf) dataset %

源域	目标域	BDA	MEDA	EasyTL	JPDA	JDSC	LDADA	ICSC	DLAD	Our
C	A	89.56	93.40	91.86	89.98	**94.10**	92.38	93.32	93.63	92.59
C	W	85.42	**95.60**	83.73	83.39	88.10	88.14	93.22	90.51	**96.95**
C	D	87.26	91.10	89.17	87.26	94.30	94.90	92.36	93.63	**96.18**
A	C	83.08	87.40	84.42	82.90	89.10	89.23	88.78	89.23	**89.58**
A	W	83.39	88.10	86.44	82.71	89.20	91.86	89.15	96.27	**98.31**
A	D	82.80	88.10	89.17	82.80	91.10	93.63	**96.18**	94.96	**96.18**
W	C	83.44	**93.20**	78.27	80.77	88.40	87.80	83.70	87.80	90.20
W	A	88.84	**99.40**	89.35	88.10	92.60	92.80	88.41	92.90	93.42
W	D	**100.00**	99.40	99.36	**100.00**	**100.00**	**100.00**	99.36	99.36	**100.00**
D	C	84.06	87.50	81.03	81.92	88.00	89.14	82.90	85.57	**89.76**
D	A	92.28	93.20	89.56	90.40	**94.00**	93.01	89.67	92.90	93.63
D	W	99.66	97.60	96.95	**100.00**	99.70	99.32	98.98	99.32	**100.00**
平均值		88.32	92.80	88.28	87.52	92.40	92.68	91.34	93.00	**94.73**

进的域适应方法进行了性能比较。表格中用粗体和下划线分别表示最高精度和次高精度。从表中可以看出本文的方法取得了良好的性能。与基准方法BDA 相比，有了显著的改进。

在Office+Caltech10 (decaf)的实验结果如表3所示，本文的方法在12个任务中有8个任务获得了最佳性能，在3个任务中获得了次佳性能。本文的分类准确率达到94.73%。具体而言，在任务C→D和任务A→W中分类准确率分别达到96.18%和98.31%，比BDA分别高出8.92、14.92个百分点。

在Office31 (ResNet50)的实验结果如表4所示，在6个任务中，本文的方法在4个任务中表现最佳。具体而言，本文方法、ICSC和DLAD的分类准确率分别为90.15%、86.07%和89.20%。

在ImageCLEF-DA (ResNet50)的实验结果如表5所示，在6个任务中，本文的方法有4个任务取得了最佳性能，有2个任务取得了次佳性能。在实验中，本文方法的平均准确率为91.35%。与BDA和 ICSC相比，分别提高了7.22、2.52个百分点。

在COIL20的实验结果如表6所示，从实验结果可看出，本文的方法平均准确率非常接近100%。相对于几种最先进的域适应方法，本文方法取得了有效的改进。具体而言，与EasyTL相比，本文方法提高了17.08个百分点；与DLAD相比，本文方法提高了2.91个百分点。

表 4　在Office31(ResNet50)数据集上的分类精度

Table 4　Classification accuracy on Office31 (ResNet50) dataset　　%

源域	目标域	BDA	MEDA	EasyTL	JPDA	JDSC	LDADA	ICSC	DLAD	Our
A	D	81.33	91.16	85.10	82.13	92.77	87.55	87.80	93.57	**94.78**
A	W	83.27	91.19	83.80	86.04	92.33	86.92	87.80	**93.46**	92.83
D	A	69.01	75.22	71.80	71.07	75.58	73.38	74.10	75.08	**77.85**
D	W	98.99	98.87	95.10	97.74	98.87	97.61	95.00	**99.12**	97.61
W	A	68.76	75.22	69.60	68.51	75.12	73.06	74.30	73.94	**77.85**
W	D	99.80	**100.00**	96.80	99.20	99.80	99.60	97.40	**100.00**	**100.00**
平均值		83.53	88.61	83.70	84.11	89.08	86.35	86.07	89.20	**90.15**

表 5　在ImageCLEF-DA(ResNet50)数据集上的分类精度

Table 5　Classification accuracy on ImageCLEF-DA (ResNet50) dataset

源域	目标域	BDA	MEDA	EasyTL	JPDA	JDSC	LDADA	ICSC	DLAD	Our
C	I	90.83	93.50	91.50	88.33	**95.50**	90.33	91.30	92.80	94.67
C	P	77.66	80.37	77.70	72.42	81.22	77.50	78.70	78.80	**81.90**
I	C	94.33	96.50	96.00	90.83	96.67	95.17	94.70	95.70	**97.17**
I	P	75.80	**81.73**	78.70	75.97	81.56	78.83	80.90	79.80	81.56
P	C	87.00	95.00	95.00	82.00	95.83	93.50	94.70	95.30	**97.50**
P	I	79.17	94.17	90.30	78.83	93.17	91.17	92.70	92.80	**95.33**
平均值		84.13	90.21	88.20	81.40	90.66	87.75	88.83	89.20	**91.35**

表 6　在COIL20数据集上的分类精度

Table 6　Classification accuracy on COIL20 dataset　　%

源域	目标域	BDA	MEDA	EasyTL	JPDA	JDSC	LDADA	ICSC	DLAD	Our
COIL1	COIL2	97.22	92.78	83.75	92.08	92.08	90.42	89.72	96.67	**100**
COIL2	COIL1	96.81	93.89	81.11	89.86	90.83	88.89	90.28	96.53	**99.03**
平均值		96.69	93.33	82.43	90.97	91.46	89.65	90.00	96.60	**99.51**

在Office-Home的实验结果如表7所示，在12个任务中，本文方法有5个任务取得了最佳性能，同时有4个任务取得了次佳性能。平均而言，相对于传统域适应学习DLAD，本文方法提升了0.17个百分点。与深度域适应学习方法TAT、DSAN、ATM相比，分别提高了3.74、2.06和1.76个百分点。

表 7 在Office-Home数据集上的分类精度

Table 7 Classification accuracy on Office-Home dataset %

源域	目标域	EasyTL	JPDA	LDADA	ICSC	DLAD	TAT	DSAN	ATM	Our
Ar	Cl	52.80	46.35	49.76	51.70	**57.70**	51.60	54.40	52.40	54.73
Ar	Pr	72.10	60.60	69.38	71.30	**78.10**	69.50	70.80	72.60	76.93
Ar	Rw	75.90	67.62	76.18	75.70	**80.70**	75.40	75.40	78.00	78.91
Cl	Ar	55.00	50.52	53.65	62.00	60.90	59.40	60.40	61.10	**64.98**
Cl	Pr	65.90	62.81	64.36	70.70	73.40	69.50	67.80	72.00	**76.50**
Cl	Rw	67.60	62.59	66.24	70.70	74.00	68.60	68.00	72.60	**79.09**
Pr	Ar	54.40	51.79	54.02	62.40	63.10	59.50	62.60	59.50	**65.97**
Pr	Cl	46.90	47.72	46.12	50.00	53.10	50.50	**55.90**	52.00	54.14
Pr	Rw	74.70	72.09	73.61	76.00	80.30	76.80	78.50	79.10	**80.67**
Rw	Ar	63.80	59.99	66.79	68.20	70.30	70.90	**73.80**	73.30	67.37
Rw	Cl	52.30	49.99	52.07	52.40	59.50	56.60	**60.60**	58.90	55.56
Rw	Pr	78.00	74.34	78.28	79.00	82.60	81.60	83.10	**83.40**	81.23
平均值		63.28	58.87	62.54	65.84	69.50	65.83	67.61	67.91	**69.67**

根据实验结果观察，得出以下结论：

模型的有效性：与BDA相比，本文的方法在5个数据集上的准确率都有明显提高。置信样本选择减少了低置信样本的影响，促进判别性学习和域的分布对齐。差异性特征增强提高了源域和目标域的样本的判别性，适用于线性和非线性数据集，目标域样本可以更容易分类。然后利用分布对齐将源域知识传播到目标域上。

与核方法对比的优缺点：由于域差异，核方法一般结合域适应方法一起使用。在上面的对比方法中，MEDA, JDSC和DLAD都采用了RBF核函数增强其模型，而本文精度高于采用核方法的域适应学习。核函数在处理非线性关系方面具有优势，它可以将数据映射到高维空间，更好地捕捉非线性关系。然而，核函数需要计算高维空间中的内积，这会导致计算复杂度的增加。此外，在不同的数据集上选择适当的核函数是一个挑战，而核函数的模型通常被认为是“黑盒”，很难解释模型学到的特征和关系。相比之下，本文方法通过权衡判别分析学习了局部关系，同样适应于非线性数据。在域适应学习中，知识迁移依赖于数据特征，而本文方法保留了数据的局部关系，提供了更强的可解释性。然而，需要注意的是，不同数据集的局部关系是不同的，在本文方法中需要考虑邻居样本的个数选项，以便更好地适应非线性数据。

3.4 参数敏感度分析

为了研究参数的影响，进行了参数敏感度分析实验，在固定其他参数的同时改变一个参数的值。

参数μ和η的敏感性：如图3(a)和图3(b)所示，μ是目标域类散点矩阵的超参数，η是目标域邻居散点矩阵的超参数。在数据集decaf、office31、COIL取得最高精度时，η均大于μ，证明对于非线性数据集(在3.5节中对数据集进行可视化，证明其线性与非线性数据)来说，邻居散点矩阵更有利于增强样本的判别性。而在数据集ImageClef取得最高精度时，η小于μ，证明对于线性数据集来说，类散点矩阵更有利于增强样本的判别性。

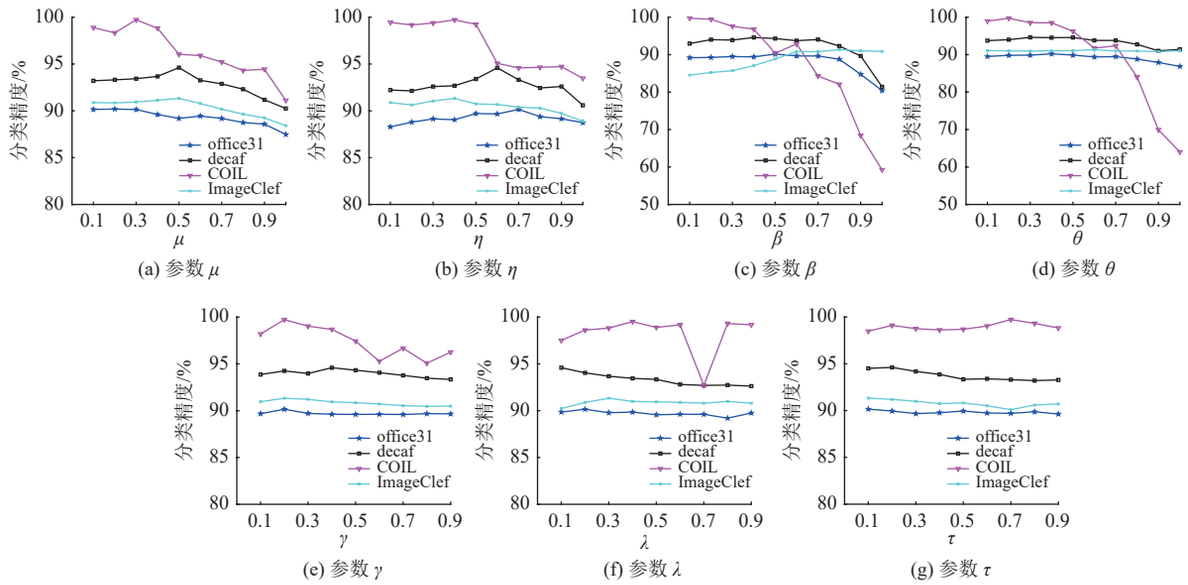
参数β和θ的敏感性：如图3(c)和图3(d)所示，β是源域类散点矩阵的超参数，θ是源域邻居散点矩阵的超参数。在所有数据集上，其趋势和μ, η一致。值得注意的是，β和θ参数在不同数据集上比较敏感，这可能是由于较大的β和θ会导致模型过度拟合于源域数据，而不利于泛化到目标域。

参数γ和λ的敏感性：如图3(e)和图3(f)所示，$\gamma \in [0.1, 0.9]$ 和$\lambda \in [0.1, 0.9]$ 被视为正则化参数。COIL的精度随正则化参数的变化而波动，其他数据集则对正则化参数不敏感。

参数τ的敏感性：如图3(g)所示，分类准确率随τ的增大而变化，其是权衡边缘分布对齐和条件分布对齐的相对重要性参数。表明边缘和条件分布对齐在不同数据集下的相对重要性不同。

3.5 可视化

为了验证本文的方法，本文选择了公共数据集进行可视化。由于数据集中包含许多类别，为了更清晰地观察，本文仅选择了部分类别进行可视化。以下是可视化的结果：(1) 图4(a)展示了在decaf数据集中DSLR的4个类别的可视化。在这个可视化中，类别2样本服从非独立同分布，有2个子类分布在类别1的两侧。而类别1、类别3和类别4样本服从独立同分布。



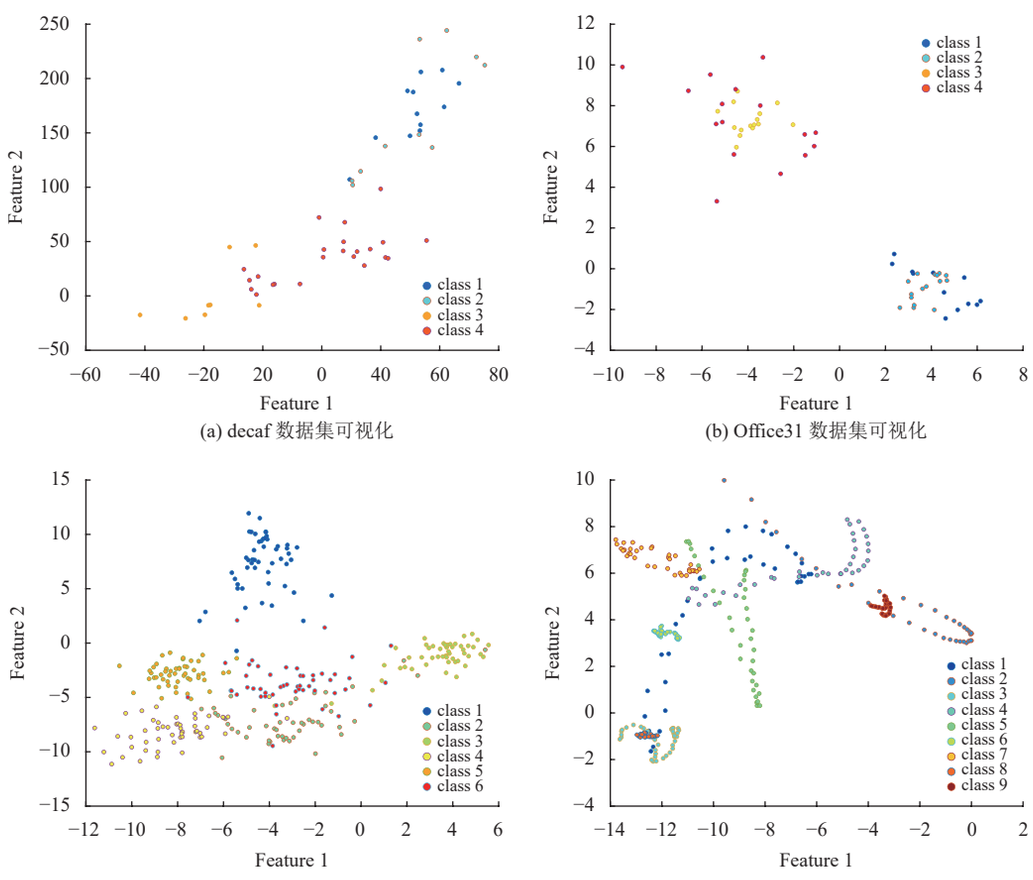
(a) 参数 μ　(b) 参数 η　(c) 参数 β　(d) 参数 θ

(e) 参数 γ　(f) 参数 λ　(g) 参数 τ

图 3　参数敏感性分析图

Fig.3　Parameter sensitivity analysis diagram

(2) 图4(b)展示了Office31数据集中DSLR的4个类别的可视化。该数据集中的类别1和类别4样本服从非独立同分布。(3) 图4(c)展示了ImageCLEF-DA数据集ImageNet ILSVRC 2012(I)的6个类别的可视化。该数据集的类内样本均服从独立同分布。(4) 图4(d)展示了COIL1数据集9个类别的可视化，存在独立同分布



(a) decaf 数据集可视化　(b) Office31 数据集可视化

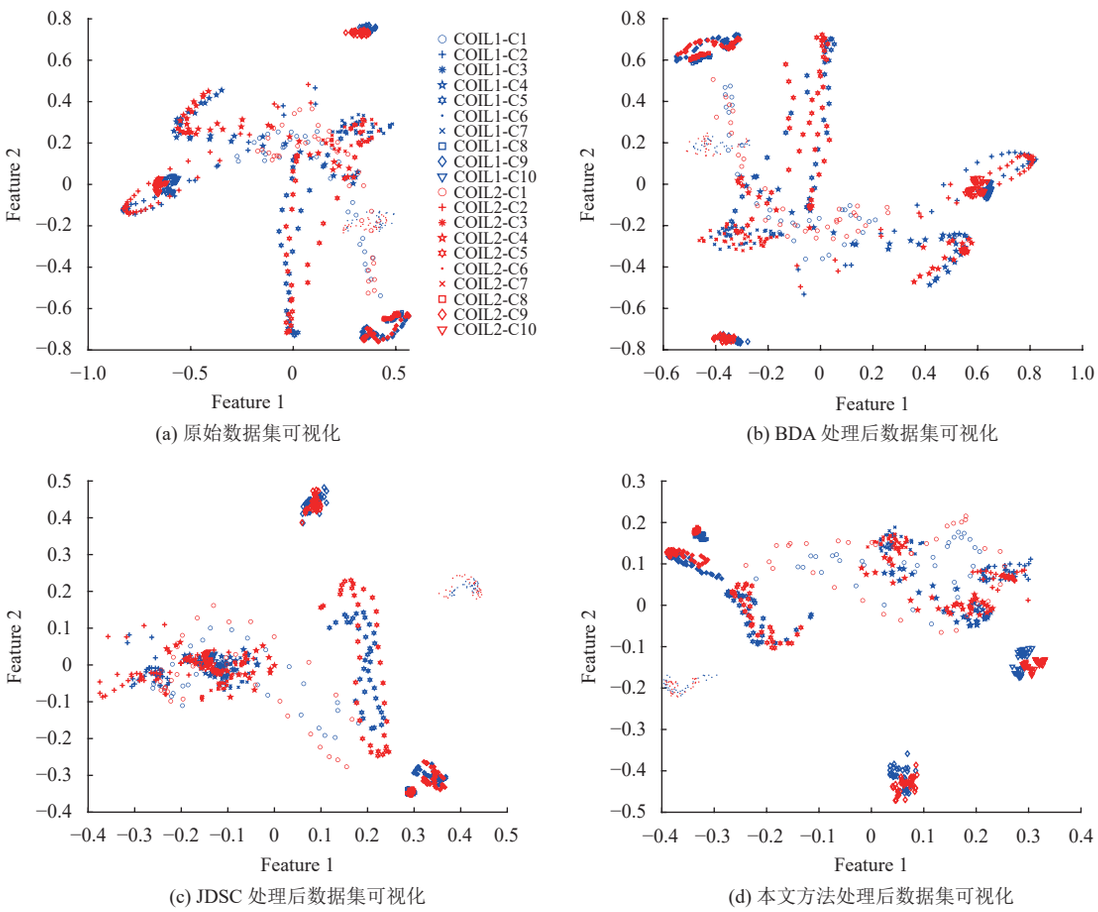
(c) ImageCLEF-DA 数据集可视化　(d) COIL1 数据集可视化

图 4　4个公共数据集部分类别的可视化

Fig.4　Visualization of some classes of four public datasets

的类别和非独立同分布的类别。因此，本文提出权衡判别分析来增强具有不同分布的数据集的判别性。

接下来，展示原始数据和处理后的COIL20数据集的数据分布。如图5所示，分别展示了未经处理和处理后的COIL20的散点图(图中C1~C10分别对应不同类别)。从图5(a)中可以观察到，对于未经处理的数据，源域(COIL1)和目标域(COIL2)存在分布差异且样本分布分散。经过BDA和JDSC处理的COIL数据集如图5(b)和图5(c)所示，源域和目标域的分布差异被减小。然而，由于样本没有被聚类，其分布相对分散。这些样本判别性不足，将影响知识的迁移过程。经过本文的方法处理后，如图5(d)所示，样本得到了有效的聚合，其判别性提高。通过跨域对齐以及知识迁移，标签能更可靠地从源域传播到目标域。



(a) 原始数据集可视化　(b) BDA 处理后数据集可视化

(c) JDSC 处理后数据集可视化　(d) 本文方法处理后数据集可视化

图 5　数据集COIL未经处理和处理后的可视化

Fig.5　Visualization of the data set COIL unprocessed and processed

3.6　消融实验

为了证明本文方法(CSS-SFE)中每个部分在模型中的重要性，对该方法进行了4个额外变体的研究。这些变体包括：(1) CSS-SFE-CSM，移除了类散点矩阵(Class Scatter Matrix, CSM)。该变体仅考虑非线性数据的可区分性。(2) CSS-SFE-NSM，移除了邻居散点矩阵(Neighborhood Scatter Matrix, NSM)。该变体仅考虑线性数据集的可判别性。(3) CSS-SFE-CSS，移除了置信样本选择。(4) CSS-SFE-SFE，移除了差异性学习。TDA代表权衡判别分析，CSS代表置信样本选择，SFE代表差异性学习。在4个数据集(Office-Caltech10, Office31, ImageCLEF-DA, COIL20)中进行实验，结果如表8所示。

根据表8，可以得出以下结论。

(1) 权衡判别分析增强样本的判别性。在decaf, Office31和COIL数据集中，使用CSS-SFE-CSM的分类精度高于使用CSS-SFE-NSM。然而，在ImageCLEF-DA数据集上，CSS-SFE-NSM表现更为优越。当两者结合使用时，所有数据集都取得了更高的分类精度。权衡判别分析的使用有助于增强样本的判别性，特别是进行跨域对齐和知识迁移时，样本更容易获得可靠的伪标签。

(2) 置信样本选择有效地防止了低置信样本的负面影响。CSS-SFE-CSS在多个数据集上的分类精

表 8　消融实验的精度结果

Table 8　The accuracy of ablation experiment　　%

方法/精度		CSS-SFE-CSM	CSS-SFE-NSM	CSS-SFE-CSS	CSS-SFE-SFE	CSS-SFE
组成成分	CSM		√			
	NSM	√				
	TDA			√	√	√
	CSS	√	√		√	√
	SFE	√	√	√		√
数据集	Office+Caltech10	93.53	92.55	92.97	92.78	**94.73**
	Office31	89.15	88.47	89.67	87.32	**90.15**
	ImageCLEF-DA	88.90	90.60	89.82	88.54	**91.35**
	COIL20	98.54	97.50	98.61	99.38	**99.51**

度均低于本文方法。尽管差异性特征增强、跨域对齐和知识迁移能够提升域适应学习分类的准确率，但是低置信样本的参与仍可能对这些增强特征和对齐过程产生负面影响，导致整体精度下降。

(3) 差异性学习保持不同域样本的判别性。CSS-SFE-SFE在大部分数据集上表现较差，这一现象可以归因于域差异导致源域和目标域的判别性存在区别。当在同一投影矩阵上学习源域和目标域时，由于这些差异被忽略，源域和目标域在投影空间中的判别性未能得到有效保留，导致样本的判别性不足。这也解释了为何CSS-SFE-SFE的性能较差。

4　总结

本文提出了一种新的无监督域适应框架，被称为置信样本选择与差异性特征增强的域适应(CSS-SFE)。通过对数据的可视化分析，证明了数据存在线性和非线性的情况。差异性特征增强在提高线性和非线性数据的判别性方面发挥了有效作用，而置信样本选择则减少了低置信样本对模型性能的负面影响。在对比实验中，本文方法凸显了其在域适应任务上的优越性，并通过消融实验验证了置信样本选择和差异性特征增强的有效性。这一研究强调了置信样本选择在促进判别性分析和域的分布对齐方面的积极作用，同时指出了差异性特征增强在权衡线性和非线性数据集方面的有效性。未来的研究可以在本文的基础上深入探讨差异性特征对域适应的积极影响。

参考文献：

[1] PAN S J, TSANG F W, KWOK J T, *et al*. Domain adaptation via transfer component analysis [J]. Journal of IEEE Transactions on Neural Networks and Learning Systems, 2011, 22: 199-210.

[2] LONG M S, WANG J M, DING G G, *et al*. Transfer feature learning with joint distribution adaptation[C]//2013 International Conference on Computer Vision. Sydney: IEEE, 2013: 2200-2207.

[3] GHIFARY M, BALDUZZI D, KLEIJN W B, *et al*. Scatter component analysis: a unified framework for domain adaptation and domain generalization [J]. Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39: 1414-1430.

[4] LI S, SONG S J, HUANG G, *et al*. Domain invariant and class discriminative feature learning for visual domain adaptation [J]. Journal of IEEE Transactions on Image Processing, 2018, 27: 4260-4273.

[5] ZHANG W, WU D R. Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation[C]//2020 International Joint Conference on Neural Networks. Glasgow, United Kingdom: IEEE, 2020, 1-8.

[6] TENG S H, ZHENG Z F, WU N Q, *et al*. Domain adaptation via incremental confidence samples into classification [J]. Journal of International Journal of Intelligent Systems, 2022, 37: 365-385.

[7] MENG M, LAN M C, YU J, *et al*. Dual-level adaptive and discriminative knowledge transfer for cross-domain recognition [J]. Journal of IEEE Transactions on Multimedia, 2023, 25: 2266-2279.

[8] ZHU F, GAO J B, YANG J, *et al*. Neighborhood linear discriminant analysis [J]. Journal of Pattern Recognition, 2022, 123: 108422.

[9] WANG J D, CHEN Y Q, HAO S J, *et al*. Balanced distribution adaptation for transfer learning[C]// 2017 International Conference on Data Mining. New Orleans: IEEE, 2017: 1129-1134.

[10] WANG J D, CHEN Y Q, YU H, *et al*. Easy transfer learning by exploiting intra-domain structures[C]//2019 International Conference on Multimedia and Expo. Shanghai: IEEE, 2019: 1210-1215.

[11] PENG Z H, ZHANG W, HAN N, *et al*. Active transfer learning [J]. Journal of IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30: 1022-1036.

[12] TIAN L, TANG Y Q, HU Y Q, *et al*. Domain adaptation by class centroid matching and local manifold self-learning [J]. Journal of IEEE Transactions on Image Processing, 2020, 29: 9703-9718.

[13] XIAO N, ZHANG L, XU X, *et al*. Label disentangled analysis for unsupervised visual domain adaptation [J]. Journal of Knowledge-Based Systems, 2021, 229: 107309.

[14] LI Y, LI D S, LU Y W, *et al*. Progressive distribution alignment based on label correction for unsupervised domain adaptation[C]//2021 International Conference on Multimedia and Expo. Shenzhen: IEEE, 2021: 1-6.

[15] LU Y W, LI D S, WANG W J, *et al*. Discriminative invariant alignment for unsupervised domain adaptation [J]. Journal of IEEE Transactions on Multimedia, 2022, 24: 1871-1882.

[16] ZHAO J C, LI L S, DENG F, *et al*. Discriminant geometrical and statistical alignment with density peaks for domain adaptation [J]. Journal of IEEE Transactions on Cybernetics, 2022, 52: 1193-1206.

[17] WANG J D, FENG W J, CHEN Y Q, *et al*. Visual domain adaptation with manifold embedded distribution alignment[C]// 2018 Multimedia Conference on Multimedia Conference. Seoul: ACM, 2018: 402-410.

[18] ZHANG W, LI C, TENG S H. Joint discriminative distribution adaptation and manifold regularization for unsupervised domain adaptation[C]// 24th International Conference on Computer Supported Cooperative Work in Design. Dalian: IEEE, 2021: 226-231.

[19] YANG H W, HE H, ZHANG W Z, *et al*. Lie group manifold analysis: an unsupervised domain adaptation approach for image classification [J]. Journal of Applied Intelligence, 2022, 52: 4074-4088.

[20] SUN J, WANG Z H, WANG W, *et al*. Joint adaptive dual graph and feature selection for domain adaptation [J]. Journal of IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32: 1453-1466.

[21] WANG Q, BRECKON. Unsupervised domain adaptation via structured prediction based selective pseudo-labeling[C]// 2020 Applications of Artificial Intelligence Conference. New York: AAAI, 2020: 6243-6250.

[22] ZHANG J, LI W Q, OGUNBONA P. Joint geometrical and statistical alignment for visual domain adaptation[C]//2017 Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 5150-5158.

[23] SARAY S N, TAHMORESNEZHAD J. Joint distinct subspace learning and unsupervised transfer classification for visual domain adaptation [J]. Journal of Signal Image Video Process, 2021, 15: 279-287.

[24] GONG B Q, SHI Y, SHA F, *et al*. Geodesic flow kernel for unsupervised domain adaptation[C]// 2012 Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012: 2066-2073.

[25] SAENKO K, KULIS B, FRITZ M *et al*. Adapting visual category models to new domains[C]// 11th European Conference on Computer Vision, Heraklion. Greece: Springer, 2010: 213-226.

[26] WANG Q, BRECKON T P. Unsupervised domain adaptation via structured prediction based selective pseudo-labeling[C]//The Thirty-Fourth Conference on Artificial Intelligence. New York: AAAI, 2020: 6243-6250.

[27] NENE S A, NAYAR S K, MURASE H. Columbia object image library (COIL-100) [R]. New York: Columbia University, 1996.

[28] VENKATESWARA H, EUSEBIO J, CHAKRABORTY S, *et al*. Deep hashing network for unsupervised domain adaptation[C]// 2017 Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 5018-5027.

[29] LIU H, LONG M S, WANG J M, *et al*. Transferable adversarial training: a general approach to adapting deep classifiers[C] //2019 Proceedings of the 36th International Conference on Machine Learning. California: PMLR, 2019: 4013-4022.

[30] ZHU Y C, ZHUANG F Z, WANG J D, *et al*. Deep subdomain adaptation network for image classification [J]. Journal of IEEE Transactions on Neural Networks and Learning Systems, 2020, 32: 1713-1722.

[31] LI J J, CHEN E, DING Z M, *et al*. Maximum density divergence for domain adaptation [J]. Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43: 3918-3930.

（责任编辑：张玮欣　英文审核：熊荣斌）