DOI:10.19651/j.cnki.emt.2211612

基于欧式距离对偶的对抗性无监督域适应算法研究*

宗子杨[1] 何 军[1] 宦 海[1] 李庆勇[2]
(1.南京信息工程大学人工智能学院 南京 210044;2.南京信息工程大学电子与信息工程学院 南京 210044)

摘 要: 在无监督域适应中,对抗性训练框架验证了双分类器差异度量对迁移学习的重要性。经典的 UDA 算法采用类内差异来度量双分类器的距离,例如 L-1 范数和 Kullback-Leibler 散度。本文从几何角度出发,考虑欧式空间中的双分类器的分布,以及传统的双分类器算法的不足,提出了一种新的欧式对偶度量,并将其纳入到对抗性的 UDA 框架中。欧式对偶度量能够有效扩大双重分类器在假设空间中的分布。另外,本文也为欧式对偶度量的理论误差上届提供了理论依据。在公共 UDA 数据集上的实验表明,欧式对偶对抗算法在小规模数据集 Digits、中规模数据集 Office-31 和大规模数据集 VisDA 的平均准确率分别为 98.3%、87.8%和 81.7%,很大程度上优于其他具有类内差异的双分类器 UDA 方法,并取得了与最先进方法相当的结果。

关键词: 域自适应;迁移学习;机器学习;欧式距离
中图分类号: TP181 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Learning on the Euclidean discrepancy dual for unsupervised domain adaptation

Zong Ziyang[1] He Jun[1] Huan Hai[1] Li Qingyong[2]
(1. School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China;
2. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Recently, the adversarial training framework of maximizing and minimizing the discrepancy between bi-classifier has been proved effective in unsupervised domain adaptation (UDA). Classical UDA approaches usually choose to use some simple intra-class discrepancies to measure the difference between the bi-classifier, such as L-1 norm and Kullback-Leibler divergence. From a geometric point of view, this work designs a novel European dual difference by considering the distribution of dual classifiers in the European Space and combining the defects of the classical dual classifier algorithm, and combines it into this adversarial UDA framework. This novel discrepancy can effectively distinguish the two probabilities predicted by the bi-classifier whether they are close in determinacy or in uncertainty. In addition, we also provide theoretical support to prove the upper bound of the theoretical error of the metric. Experiments on the public UDA dataset show that the average accuracy of the European-style dual adversarial algorithm in the small-scale dataset Digits, the medium-scale dataset Office-31, and the large-scale dataset VisDA are 98.3%, 87.8%, and 81.7%, which outperforms other 2-classifier UDA methods with intra-class variance and achieves results comparable to state-of-the-art methods.

Keywords: domain adaptation;transfer learning;machine learning;Euclidean metric

0 引 言

随着深度学习[1]在人工智能领域中的不断应用,监督学习在图像分类领域中取得了极佳的效果。监督学习主要依赖于大量标注的数据,而这种大量标注的数据需要耗费极大的人成本。为了节约相应的成本,无监督学习应运而生,如何对现实世界中的无标签的图像进行分类[2]成为越来越受到关注的话题。目前主流无监督学习主要分为两类,即依赖数据内部的结构信息进行分类训练和从有标注的源域习得知识对目标域进行分类。其中,这种学习源域到目标域映射的无监督学习方法被称为域自适应(unsupervised domain adaptation,UDA)。

收稿日期:2022-10-06
* 基金项目:国家自然科学基金(62001238)项目资助

最近，基于对抗性双分类器的无监督域适应方法在域适应方面取得了巨大成功，非常流行。这些方法中的大多基于类内差异，比如 Saito 等[3]提出了 MCD（maximum classifier discrepancy）度量不同域之间的域间隙并通过减小域间隙来对齐源域和目标域的样本特征，Wang 等[4]在 MCD 的基础上引入了特征鉴别对齐 DFA（discriminative feature alignment，DFA）。他们工作的核心是最大化不同分类器之间的差异，以尽可能精准的度量源域和目标域的间隙。Saito 等[3]采用包括一个特征生成器和两个分类器的架构，先通过最大化双分类器之间的差异来探索分类器的决策边界，而后通过最小化双分类器之间的差异促使特征生成器生成更具辨别力的特征。受 Saito 等的启发，Li 等[5]对双分类器之间距离对度量进行改进，利用概率之间的内积进一步提高分类结果。

这种双分类器的结构能够学习到将源域和目标域的映射到同一特征空间的特征生成器，并且得到有效鉴别无标签图像的分类器，但是这样的“有效”并不鲁棒，一旦改变实验设置的网络参数或随机数，结果就不再稳健。这是因为模型在对特征生成器参数进行反向传播优化的过程中，两个分类器之间的对抗尤为重要。如果双分类器在模型计算之初出现如[0.33，0.34，0.33]和[0.34，0.33，0.33]这样的分布，那么对于特征生成器来说它的可优化空间极小，因为分类器趋于相同，没有对抗的空间，显然在模型计算之初这样的一致性分布置信度并不高。

为了解决这个问题，Shannon 等[6]提出的信息熵理论指出，信息的确定性越高，熵越小。为了提高模型性能，Lee 等[7]使用目标域的熵最小化作为正则化项，使目标域分类器的输出具有确定性，同时试图通过熵的引入使得分类器对于图像自身的信息有指向性地趋于不一致。但这忽略了一个问题，熵的引入确实能够在训练过程中利用图像自身提取的特征信息将分类器分开，但是在训练早期特征生成器的置信度并不高，对结果仍然会造成影响。

从几何角度看，MCD 在训练特征生成器时使用曼哈顿距离[8]所度量的分类器差异进行反向传播，这个度量并不是欧氏空间[9]中两点最直接的度量，且由于可度量空间过短，当双分类器预测概率均匀分布而训练尚未完全时，分类器已满足优化目标，导致反向传播过程中无法更有效地对分类器进行充分对抗训练。

针对以上问题，本文设计了一种基于双分类器的欧式对偶度量方法（Euclidean discrepancy dual，EDD），一方面通过采用欧氏距离度量，更直接地度量双分类器之间的差异，同时以对偶的度量方式来替代过短的优化距离，再结合信息熵等正则化方法进一步提高结果的确定性。本文提出的方法，也可以与其他基于度量的域自适应的方法比如多分类器方法结合，进一步帮助其他方法提升效率。

1　欧式空间对偶差异及其理论支撑

1.1　方　　法

本文所使用的 EDD 网络结构参考了 MCD 所使用的网络结构和训练方法，具体的网络框架如图 1 所示。给定一个有 m 个带标签样本的源域 $S=\{X^S,Y^T\}_m$ 和一个有 n 个不带标签样本的目标域 $T=\{X^T\}_n$。域自适应的目标是学习从源域迁移到目标域的域不变表示。

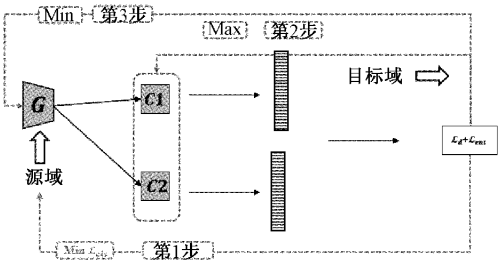

图 1　EDD 网络结构

EDD 网络结构由一个特征生成器 G 和两个分类器 $C1$，$C2$ 组成。其中，特征生成器 G 用于生成源域和目标域的特征，分类器 $C1$，$C2$ 接受特征生成器生成的样本特征，并对其进行分类，输出最终分类结果。

EDD 网络结构包括了 3 个反向传播的过程：

1）“Min-第 1 步”为第 1 次反向传播，第 1 次反向传播监督学习源域上的带标签的样本，以初始化特征生成器 G 和分类器 $C1$，$C2$ 的参数，获得目标任务的源样本的信息；

2）“Max-第 2 步”为第 2 次反向传播，仅通过优化分类器 $C1$，$C2$ 的参数来最大化双分类器对目标域样本测试的结果之间的差异，使得分类器的分布贴近源域和目标域之间的决策边界；

3）“Min-第 3 步”为第 3 次反向传播过程，仅通过优化生成器 G 的参数来最小化分类器之间的分布，目的在于，通过调整生成器 G 的参数来使得差异极大的分类器意见趋于一致，以此对齐源域和目标域的特征的分布。

可见，在整个框架中，对结果影响最大的部分，在于分类器之间差异的度量，这关系到算法是否能够将网络参数优化到最理想的状态。

同时，分类器对目标域样本的预测结果的分布也极为重要。假如双分类器在训练早期出现如[0.33，0.34，0.33]和[0.34，0.33，0.33]这样的分布，显然这样的分布置信度并不高，而对算法的优化目标而言，却是比较理想的结果。如此会使得损失函数在训练早期就收敛，而网络参数却没有得到最大程度的优化。

为了最大程度利用图形处理器（graphics processing unit，GPU）的训练资源，提升算法效率，提高域自适应的迁移结果。本文提出了一种用于度量分类器差异的欧式对偶度量，这种度量既可以充分利用反向传播的资源，也能防止双分类器在训练早期的收敛。

1.2 欧式空间对偶差异(EDD)

如图 2 所示,从几何角度来观察这个问题。图 2 展示的是在一个二分类问题中,双分类器之间的差异在欧氏空间中的情况。假设双分类器分别为 C_1 和 C_2,则当使用曼哈顿距离时,在欧氏空间中度量双分类器之间的差异所经过的路径为 a_1+a_2。实际上在欧氏空间中双分类器之间的最短路径为 a_3,即双分类器之间差异的 L-2 范数。L-1 范数在无形中对算法在反向传播的优化过程中造成一定资源的浪费。推广到更高维度的欧式空间,L-2 范数相较于 L-1 范数更具有优势。

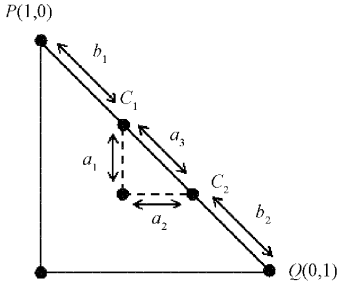

图 2 不同度量方式的对比

而对于当双分类器预测概率均匀分布而训练不完全时已满足优化目标,导致反向传播下陷入局部最优点的情况,可以从对偶的角度来考虑这个问题。

可以考虑通过优化分类器之间差异以外的部分,即在优化过程中以最大化 b_1+b_2 代替最小化 a_3,以对偶逼迫分类器向理想的方向优化。请注意,图中展现的只是双分类器在二维欧式空间中的情况,而实际的分类器可能是 n 维,所以欧式对偶并不是简单的 $\sqrt{2}-a_3$。

对此,本文设计了一个新的欧式对偶度量用于度量分类器之间的差异以反映域间隙。假设 n 维双分类器的预测向量分别为 P 和 Q,有:

$$d(P,Q)=\sum \| p_i-q_j \|_2-\sum \| p_k-q_k \|_2 \quad (1)$$

其中,$P=\{p_1,p_2,\cdots,p_n\}$ 且 $Q=\{q_1,q_2,\cdots,q_n\}$,$i,j,k\in\{1,2,\cdots,3\}$。

需要注意的是,MCD 通过最小化双分类器的曼哈顿距离使双分类器相似,它只考虑分类器之间的相似性。当概率分布接近均匀分布时,分类器也可能相似,此时,分类器之间的差异很差。同样的情况也出现在很多与双分类器相关的工作中。本文一方面通过采用欧氏距离度量,更直接地度量双分类器之间的差异,同时以对偶的度量方式来替代过短的优化距离,再结合信息熵等正则化方法进一步提高结果的确定性。

1.3 理论支撑

在这项工作中,本文提出了一个用于更精准度量源域 S 和目标域 T 之间域间隙的新度量 $d(P,Q)$。本文严格证明了 $d(P,Q)$ 的迁移误差上界限。

为了度量 S 和 T 之间的域间隙,本文引入 David[10] 关于假设空间的理论。

引理 1:对于任意一个假设空间 H 有 $h\in H$,则任意两个分布 D 和 D' 之间的距离被定义为:

$$d_{H\Delta H}(D,D')=2\sup[\Pr_D[I(h)]-\Pr_{D'}[I(h)]] \quad (2)$$

由此可知,对于 $\forall h,h'\in H$,S 和 T 之间的域间隙为:

$$d_{H\Delta H}(S,T)=2\sup[\Pr_S[I(h,h')]-\Pr_T[I(h,h')]] \quad (3)$$

假定 U_S 和 U_T 是无标签样本,其中 U_T 有 m' 个样本且 $\forall\delta\in(0,1)$,$h\in H$,目标域的误差上界为:

$$\varepsilon_T(h)\leqslant\varepsilon_S(h)+\frac{1}{2}d_{H\Delta H}(U_S,U_T)+4\sqrt{\frac{2d\log(2m')+\log(2/\delta)}{m'}}+\lambda \quad (4)$$

由此可知,当 $h,h'\in H$ 对于有标签的源域和无标签的目标域误差上界为:

$$\varepsilon_T(h)\leqslant\varepsilon_S(h)+\frac{1}{2}\sup|\hat{d}_S(h,h')-\hat{d}_T(h,h')|+4\sqrt{\frac{2d\log(2m')+\log(2/\delta)}{m'}}+\lambda=\varepsilon_S(h)+\frac{1}{2}\sup\left|\begin{bmatrix}1 & -1\end{bmatrix}\begin{bmatrix}\hat{d}_S(h,h')\\ \hat{d}_T(h,h')\end{bmatrix}\right|+4\sqrt{\frac{2d\log(2m')+\log(2/\delta)}{m'}}+\lambda \quad (5)$$

其中,$\hat{d}_T(h,h')$ 表示双分类器在目标域上的距离,而 $\hat{d}_S(h,h')$ 表示双分类器在源域上的距离。

1.4 信息熵正则化方法

本文采用信息熵作为正则化方法辅助训练,以进一步提升训练结果的确定性。已知信息熵越大,结果确定性越低,信息熵越小,确定性越高。通过对分类器输出概率的分布计算信息熵,并将其添加到损失函数中进行反向传播,能够迫使网络参数向着更理想的方向优化。

在第 1 步中,监督学习,不需要正则化方法辅助。

在第 2 步中,我们将信息熵最大化,使分类器的结果趋于均匀分布,并鼓励分类器预测概率趋向于多样化。

在第 3 步中,最小化信息熵,以提升预测概率的确定性,提高预测结果的置信度,进一步减小源域和目标域之间的域间隙。

其中,$\ell_{ent}(X^T)$ 是条件熵,用于衡量分类的确定性,条件熵定义为:

$$\ell_{ent}=-E\sum p_i(\hat{y}^T\mid X^T)\log(p_i(\hat{y}^T\mid X^T)) \quad (6)$$

2 训练过程

2.1 算 法

如图 1 所示,ℓ_{cls} 表示特征生成器 G 和两个分类器 $C1$,$C2$ 的交叉熵损失,ℓ_d 表示两个分类器 $C1$,$C2$ 的差异,ℓ_{ent}

表示两个分类器 $C1$,$C2$ 的信息熵损失。

首先,在源域样本上训练生成器和分类器,并学习映射 $f_s:X^S\rightarrow Y^S$。为了使模型对源数据进行正确分类,优化目标是使分类损失最小化。$C_i(G(X^S))$ 是源域样本的基本分类器输出的概率分布,Y^S 是源域样本的真实标签。

如图 1 所示,灰色反向传播箭头指向所有模块,这表明需要初始化所有网络参数。目标函数如下:

$$\min_{G,C}\quad \ell_{cls}(X^S,Y^S)=-\mathrm{E}\left[\frac{1}{n}\cdot y_s(\log(C1(G(X_s)))+\log(C2(G(X_s))))\right] \tag{7}$$

其次,固定生成器并训练分类器网络。本文使用欧式对偶度量来优化分类器,并最大化分类器之间的差异,从而确保基本分类器的多样性。如图 1 所示,蓝色反向传播箭头仅指向分类器模块,表示生成器网络参数未更新。

$$\max_{C1,C2}\quad \ell_{adv}(Y^T)=\mathrm{E}[d(C1,C2)]+\ell_{ent}(C1,C2) \tag{8}$$

经过上述训练后,目标样本大多位于决策边界附近。为了鼓励生成器生成远离决策边界的显著特征,对生成器进行训练,以最小化欧式对偶度量。同时,最大化目标样本和输出之间的信息熵。如图 1 所示,图中的橙色表示固定分类器网络参数未更新。

$$\min_{G}\quad \ell_{adv}(Y^T)=\mathrm{E}[d(C1,C2)]+\ell_{ent}(C1,C2) \tag{9}$$

在测试过程中,对所有分类器的输出进行平均,以获得目标域数据的标签。

$$p(y\mid x_i)=\frac{1}{n}\sum p_i(y\mid x_i) \tag{10}$$

2.2　训练步骤

如图 1 所示,具体的训练步骤分为 3 步:

1)第 1 步:模型预训练

首先通过对标记的源域样本的监督学习来训练特征生成器和双分类器,以确保特征生成器获得特定任务的显着特征。训练双分类器 $C1$,$C2$ 和生成器 G 来最小化 Softmax 交叉熵,优化目标如下:

$$\min_{G,C1,C2}\quad \ell_{cls}(X^S,Y^S) \tag{11}$$

2)第 2 步:训练双分类器

固定特征生成器 G 的参数,并使用未标记的目标域样本来训练双分类器 C_1,C_2 以增加类间差异。优化目标如下:

$$\min_{C1,C2}\quad -\ell_d(X^T)-\ell_{ent}(X^T) \tag{12}$$

3)第 3 步:训练生成器

固定分类器 $C1$,$C2$ 的参数,使用未标记的目标域样本来训练特征生成器 G。目标域特征对类别相似的源域特征是封闭的,远离决策边界,更具判别力。优化目标如下:

$$\min_{G}\quad \ell_d(X^T)+\ell_{ent}(X^T) \tag{13}$$

在训练过程中,循环往复训练以上 3 个步骤,使得双分类器 $C1$,$C2$ 和生成器 G 的参数达到最理想的情况。

3　实验设计与结果分析

3.1　数据集

为了验证本文提出的 EDD 度量的广泛性、鲁棒性以及有效性,本文分别针对小规模的简单数据集 Digits、中等规模复杂数据集 Office-31 以及大规模复杂数据集 VisDA 进行了实验与测试。同时为了验证本文所使用的正则化方法对结果的影响,本文选择在小规模数据集 Digits 上进行了消融实验。

1)Digits 数据集:本文在 SVHN[11]、USPS[12]、MNIST 上进行了实验,以评估本文方法的有效性。如图 3 所示,SVHN(S)是一个由门号的真实街景组成的彩色数字图像裁剪数据集;MNIST(M)由现实世界中的彩色图像组成,相对模糊;USPS(U)由手写数字组成。这 3 个数据集有 10 个类别,但它们的分布不同。对于 SVHN(S)、MNIST(M)和 USPS(U)的样本,设置了 3 组域适应任务:S→M、U→M 和 M→U。

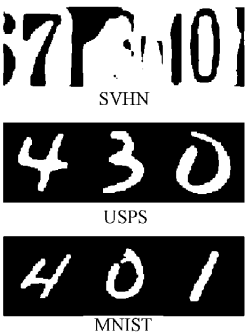

图 3　Digits 图例

2)Office-31 数据集[13]:Ofiice-31 包含 31 类图片,共 4 652 个样本,是测试域适应的通用数据集。如图 4 所示,该数据集的图像来自 3 个不同的数据域,包括亚马逊网站 Amazon(A)采集的样本、计算机摄像头 Webcam(W)采集的样本和单镜头采集的样本 DSLR(D)。对于 Amazon(A)、Webcam(W)和 DSLR(D)的数据样本,设置了 6 组域适应任务:A→D、D→A、A→W、W→A、D→W 和 W→D。

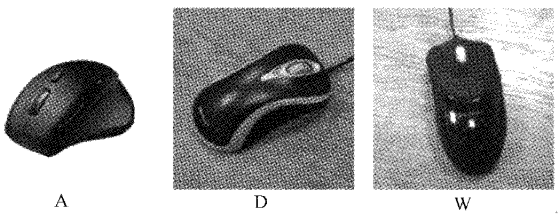

图 4　Office-31 图例

3)VisDA 数据集:如图 5 所示,VisDA 代表最大的跨域对象分类,包含 Train、Validation 和 Test 的 12 个类别中超过 280 K 的图像。对于 VisDA 设置了 Train→Validation 的任务。

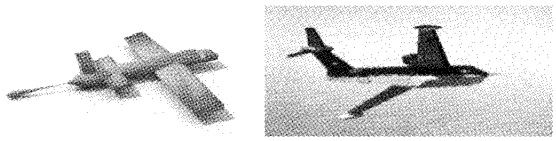

图 5 VisDA 图例

3.2 训练细节

本文使用的实验框架为 Pytorch 深度学习框架,操作系统为 Ubuntu16.04,GPU 为英伟达 GTX2080Ti,显存为 12 G。

对于 Digits 数据集,本文采用 MCD 中使用的卷积神经网络(convolutional neural networks,CNN)架构进行特征生成,并使用全连接层作为分类器。

对于中型和大型数据集,本文采用 ResNet[14] 作为 VisDA 数据集(ResNet-101)和 Office-31 数据集(ResNet-50)的特征生成器。对于分类器,本文为 VisDA 使用三个全连接层,为 Office-Home 使用两个全连接层。模型采用随机梯度下降(stochastic gradient descen,SGD)优化,批量大小设置为 32。

本文在所有实验中均设置了 Dropout 层来防止过拟合,同时为了减少梯度爆炸的影响,对所有实验使用了批次归一化层。

本文在数字任务中将学习率设置为 1×10^{-2},在 VisDA 任务和 Office-31 任务中设置为 1×10^{-3}。

另外,本文采用学习率衰减[15] 的方法来优化学习率并将权重衰减设置为 5×10^{-4}。

3.3 消融研究

为了验证本文所使用的正则化方法边缘熵对实验结果的影响,以本文提出的欧式对偶度量作为对抗训练的基础对 Digits 数据集的 3 项任务进行了 4 组实验的对比:1)仅使用欧式对偶度量;2)仅在第 2 步训练过程中添加信息熵正则化项;3)仅在第 3 步训练过程中添加信息熵正则化项;4)同时在第 2 步和第 3 步添加信息熵正则化项。

实验的对比结果展示在表 1 中,从结果可以看出,使用信息熵作为正则化项辅助训练,能够提升算法整体训练效果。同时在不使用信息熵的情况下,本文所设计的欧式对偶度量仍然能保持具有竞争力的效果。其中,算法 2 和算法 3 差距不大,表明在整个算法流程中,第 2 步和第 3 步的对结果的影响同样重要。而算法 4 相较于算法 1 整体平均值提升了 2.1%,可见信息熵对于结果的提升具有一定帮助。

表 1 信息熵对 EDD 的影响 %

算法	S→M	M→U	U→M	均值
1	96.4	96.7	96.9	96.7
2	97.1	97.1	97.3	97.2
3	97.9	97.0	97.6	97.5
4	98.8	97.6	98.6	98.3

3.4 算法性能对比

为了验证欧式对偶度量对域适应的有效性,本文分别对小规模数据集 Digits、中规模数据集 Office-31 和大规模数据集 VisDA 进行了性能测试并与其他先进方法进行了对比。

1)Digits 数据集性能对比

本文记录了 Digits 数据集在第 150ep 的结果,3 项任务的测试结果和对比方法的结果被展示在表 2 中,Source 代表仅使用源域监督训练的模型在目标域上测试得到的结果。其中,欧式对偶度量对 Digits 的 3 组任务均取得了最好的结果,并且在平均准确度上优于以前的工作。与基线方法 MCD 相比,通过引入欧式对偶度量,S→M、U→M 和 M→U 分别提升了 2.2%,4.5%和 1.1%,平均结果提升了 2.7%,分类精度得到了大幅度的提高。对于 U→M 任务,提升较为明显,原因在于 MNIST 的数据集规模远远大于 USPS,所以传统方法习得的源域知识较少,不足以支撑较好的域间隙的度量,而 EDD 度量能够更准确的估计域间隙。

表 2 EDD 在数字数据集上的对比试验 %

算法	S→M	M→U	U→M	均值
Source	67.1	76.7	63.4	69.1
MMD[16]	71.1	81.1	—	76.1
DANN[17]	76.0	85.1	73.2	78.1
ADDA[18]	76.0	89.4	90.1	95.2
GTA[19]	92.4	92.8	90.8	92
MCD[13]	96.2	96.5	94.1	95.6
EDD	**98.8**	**97.6**	**98.6**	**98.3**

由于 Digits 数据集整体结果较好,更明显的差异需要在更大的数据集上进一步比较。

2)Office-31 数据集性能对比

本文记录了 Office-31 数据集在第 100ep 的结果,六项任务的测试结果和对比方法的结果被展示在表 3 中,ResNet50 代表仅使用源域监督训练的模型在目标域上测试得到的结果。实验中使用最小化熵和伪标签作为正则化方法。与其他领域适应方法相比,本文的方法达到了最高的平均准确率,并且 5 个任务中均表现最佳,A→W 也表现良好。由此可见本文的方法对中等数据集同样能够保持具有竞争力的结果。

3)VisDA 数据集性能对比

本文记录了 VisDA 数据集在第 100ep 的结果,12 个类的测试结果和对比方法的结果被展示在表 4 中,ResNet101 代表仅使用源域监督训练的模型在目标域上测试得到的结果。本文对 VisDA 任务执行了 30 个 epoch 的实验,实验中使用最小化熵和伪标签作为正则化方法。与其他方法相比,可以观察到在 VisDA 中引入欧式对偶度量后,准确率

表 3　EDD 在 Office-31 数据集上的对比试验　　%

算法	A→D	A→W	D→A	D→W	W→A	W→D	均值
ResNet50	68.9	68.4	62.5	96.7	60.7	99.3	76.1
DANN[16]	79.7	82.0	68.2	96.9	67.4	99.1	82.2
DAN[19]	78.6	80.5	63.6	97.1	62.8	99.6	82.2
GTA[18]	87.7	89.5	72.8	97.9	71.4	100.0	86.5
CDAN[20]	89.8	**93.1**	70.1	98.2	68.0	99.9	86.5
EDD	**90.4**	91.3	**73.7**	**98.6**	**72.6**	**100.0**	**87.8**

表 4　EDD 在 VisDA 数据集上的对比试验　　%

算法	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	均值
ResNet101	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
CDAN[11]	85.2	66.9	83.0	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	38.0	73.9
JADA[12]	91.9	78.0	81.5	68.7	90.2	84.1	84.0	73.6	88.2	67.2	79.0	38.0	77.0
SWD[13]	90.8	**82.5**	81.7	70.5	91.7	69.5	**86.3**	77.5	87.4	63.6	**85.6**	26.2	76.4
MCD[3]	87.0	60.9	83.7	64.0	88.9	79.6	84.7	76.9	**88.6**	40.3	83.0	25.8	71.9
EDD	**96.7**	81.8	**84.5**	**72.1**	**92.2**	**90.9**	86.1	**82.3**	83.4	**80.0**	85.4	**45.3**	**81.7**

得到了一定程度的提高，效果明显高于其他方法。与直接可比的基线方法 MCD 相比较，本文提出的 EDD 度量提升了 9.8%，具有非常大幅度的提升。可见，对于大规模数据集，EDD 度量可以避免 L-1 范数在某些情况下的缺陷。

4　结　　论

本文从几何角度出发，设计了一种基于欧氏距离对偶度量 EDD 的自适应算法。该算法将双分类器中的每个分量视为欧式空间中的几何点，并最大化几何点对于不同类的预测概率的差异，有效地最大化真实分类的预测概率。此外，本文还提供了理论支持，以确保新度量的有效测度，并证明该度量的理论误差的上限。实验结果表明，本文所提出的方法在小规模数据集中相比于同类型双分类器架构中的经典算法 MCD 提升了 2.7%，在大规模数据集中提升了 9.8%，提升幅度巨大。同时，在小、中、大 3 类数据集准确率分别 98.3%，87.8%，and81.7%，出色的表现相比于最近其他流行的 UDA 架构也毫不逊色。此外，最近源域隐私保护相关的技术越来越受到关注。由于双分类器在假设空间中仅为双假设形式，在源域数据不可访问的情况下搜索空间有限，仍需进一步研究。

参考文献

[1] DENG L, YU D. Deep learning: Methods and applications[J]. Journal of Foundations and trends in signal processing, 2014, 7(3-4):197-387.

[2] 赵勇，李怀宇. 基于通用距离测量的机器学习方法用于图像分类和聚类[J]. 电子测量技术，2017，40(9)：136-140.

[3] SAITO K, WATANABE K, USHIKU Y, et al. Maximum classifier discrepancy for unsupervised domain adaptation[C]. Proceedings of the Computer Vision and Pattern Recognition(CVPR). IEEE, 2018: 3723-3732.

[4] WANG J, CHEN J, LIN J, et al. Discriminative feature alignment: Improving transferability of unsupervised domain adaptation by Gaussian-guided latent alignment [J]. Pattern Recognition, 2021, 116(1):107-143.

[5] LI S, LYU F, XIE B, et al. Bi-classifier determinacy maximization for unsupervised domain adaptation [C]. Proceedings of the AAAI Conference on Artificial Intelligence(AAAI), 2021:2.

[6] SHANNON C E. A mathematical theory of communication [J]. The Bell System Technical Journal, 1948, 27(3):379-423.

[7] LEE S, KIM D, KIM N, et al. Drop to adapt: Learning discriminative features for unsupervised domain adaptation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 91-100.

[8] MALKAUTHEKAR M D. Analysis of Euclidean distance and Manhattan distance measure in Face recognition [C]. Proceedings of Conference on Computational Intelligence and Information Technology(CIIT 2013), 2013:503-507.

[9] KAKUTANI S. Some characterizations of Euclidean space [J]. Japanese Journal of Mathematics: Transactions and Abstracts, 1940, 16(1):93-97.

[10] BEN-DAVID S, BLITZER J, CRAMMER K, et al. A theory of learning from different domains[J]. Machine Learning, 2010, 79(1):151-175.

[11] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[C]. Proceedings of the IEEE, 1998: 2278-2324.

[12] HULL J J. A database for handwritten text recognition research[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 1(1):550-554.

[13] SAENKO K, KULIS B, FRITZ M, et al. Adapting visual category models to new domains[C]. Proceedings of the European Conference on Computer Vision, 2010:216-226.

[14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.

[15] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[C]. Proceedings of the International Conference on Machine Learning, 2015: 1180-1189.

[16] GRETTON A, BORGWARDT K M, RASCH M J, et al. A kernel two-sample test[J]. Journal of Machine Learning Research, 2012, 13(1):723-773.

[17] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain adversarial training of neural networks[J]. Journal of Machine Learning Research, 2016, 17(59): 1-35.

[18] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial Discriminative Domain Adaptation[C]. Proceedings of the Computer Vision and Pattern Recognition, IEEE, 2017:7167-7176.

[19] SANKARANARAYANAN S, BALAJI Y, CASTILLO C D, et al. Generate to adapt: Aligning domains using generative adversarial networks[C]. Proceedings of the Computer Vision and Pattern Recognition, IEEE, 2018:8503-8512.

[20] LONG M, CAO Z, WANG J, et al. Conditional adversarial domain adaptation[J]. Advances in Neural Information Processing Systems, 2018, 1(1): 1640-1650.

作者简介

宗子杨,硕士研究生,主要研究方向为迁移学习、域自适应。

E-mail:zongziyang@nuist.edu.cn

何军,博士,副教授,主要研究方向为机器学习、计算机视觉。

E-mail:hejun.zz@gmail.com

宦海(通信作者),博士,副教授,主要研究方向为遥感图像语义分割、图像超分辨率重建。

E-mail:haihuan@nuist.edu.cn

李庆勇,硕士研究生,主要研究方向为迁移学习、域自适应。

E-mail:qingyongli2021@gmail.com