基于生成对抗网络的无监督域适应分类模型

王格格,郭 涛,余 游,苏 菡
(四川师范大学计算机科学学院,四川成都 610101)

摘 要: 生成适应模型利用生成对抗网络实现模型结构,并在领域适应学习上取得了突破.但其部分网络结构缺少信息交互,且仅使用对抗学习不足以完全减小域间距离,从而使分类精度受到影响.为此,提出一种基于生成对抗网络的无监督域适应分类模型(Unsupervised Domain Adaptation classification model based on GAN,UDAG).该模型通过联合使用生成对抗网络和多核最大均值差异度量准则优化域间差异,并充分利用无监督对抗训练及监督分类训练之间的信息传递以学习源域分布和目标域分布之间的共享特征.通过在四种域适应情况下的实验结果表明,UDAG模型学习到更优的共享特征嵌入并实现了域适应图像分类,且分类精度有明显提高.

关键词: 生成适应模型;迁移学习;领域适应学习;生成对抗网络;多核最大均值差异;无监督学习
中图分类号: TP181 **文献标识码:** A **文章编号:** 0372-2112 (2020)06-1190-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2020.06.021

Unsupervised Domain Adaptation Classification Model Based on Generative Adversarial Network

WANG Ge-ge, GUO Tao, YU You, SU Han
(*Department of Computer Science, Sichuan Normal University, Chengdu, Sichuan* 610101, *China*)

Abstract: Generate-to-adapt model has used generative adversarial network to implement model structure and has made a breakthrough in domain adaptation learning. However, some of its network structures lack information interaction, and the ability to use only adversarial learning is not sufficient to completely reduce the inter-domain distance. In this paper, an unsupervised domain adaptation classification model based on generative adversarial network (UDAG) is proposed. This model optimizes inter-domain differences and makes full use of the information between unsupervised confrontation training and supervised classification training to learn the shared features between the source and target domain distribution. The experimental results under four domain adaptation conditions show that the UDAG model learns better shared feature embedding and implements domain adaptive classification, and the classification accuracy is significantly improved.

Key words: generate-to-adapt model; transfer learning; domain adaptation learning; generative adversarial network; MK-MMD; unsupervised learning

1 引言

在大数据时代的背景下,数据呈爆炸式增长,但大部分数据属于无标记数据.由于进行数据标定的工作代价高昂且费时,近年无监督思想在机器学习领域得到了广泛应用[1,2].此外,传统的机器学习算法通常假设训练数据和测试数据服从相同的概率分布,但实际上大多数实用场景下训练数据的概率分布和测试数据的概率分布是不同的[3],从而导致了在传统机器学习算法下训练的模型可能不再适用于实际应用中出现的情况.

针对以上问题,迁移学习(Transfer Learning,TL)方法被提出[4],领域适应学习(Domain Adaptation Learning,DAL)作为一种同构迁移学习方法[5],解决了如何在源域数据分布和目标域数据分布不同但相关的情况下,实现目标域上学习任务的问题.Ganin[6]、Ghifary[7]和Tzeng[8,9]等人利用传统DAL的思想,通过添加梯度反转层、学习共享编码或者使用预训练模型等方法将两个领域的图像特征映射到相同的特征分布上,以实现目标域的图像在源域模型上的准确分类.而生成对

收稿日期:2019-06-25;修回日期:2019-11-13;责任编辑:覃怀银
基金项目:国家自然科学基金(No.61403266);四川省可视化与虚拟现实重点实验室项目(No.KJ201419)

抗网络(Generative Adversarial Network,GAN)[10,11]作为深度学习中一种无监督生成模型,其独特的对抗训练方式能够让噪声分布逐渐向真实数据分布拟合,在图像生成方面具有良好的表现.因此,Russo[12]、Volpi[13]和Hoffman[14]等人基于GAN,通过引入域间对称映射、进行数据增强或执行重构损失和语义损失来保证源域图像和目标域图像在特征空间上的一致性.

近来,Sankaranarayanan等人[15]提出生成适应模型(Generate-To-Adapt model,GTA),通过利用特征嵌入和GAN之间的共生关系来使源域分布和目标域分布在一个联合特征空间中更接近.但由于GTA部分网络结构过于分离,缺少信息交互,并且仅使用GAN的对抗学习能力不足以优化域间差异,从而影响域适应分类效果.为此,本文提出一种基于GAN的无监督域适应分类模型UDAG,该模型联合使用GAN和多核最大均值差异(the Multiple Kernel variant of Maximum Mean Discrepancy,MK-MMD)[16]最小化域间距离,并利用GAN无监督对抗训练与分类网络监督训练之间的信息传递以学习源域分布和目标域分布之间的共享特征,并回传各个类别空间的特征梯度信号,以供其他网络学习分类信息.

2　理论基础

2.1　领域适应学习

定义1(数据集)　数据集 D 是一个二元组 $D=(X,Y)$,其中:$X=\{x_1,x_2,\cdots,x_m\}$ 是 m 个输入实例观察值集合,本文中,$\forall x_i\in X$,x_i 表示输入的一个图片实例;$Y=\{y_1,y_2,\cdots,y_n\}$ 是 n 个类别标签集合.

定义2(源域数据集和目标域数据集)　给定有标记数据集 $D_s=(X_s,Y_s)$ 和无标记数据集 $D_t=(X_t,Y_t)$,设 D_s 和 D_t 的特征空间分别为 F_s、F_t,如果满足:

①$F_s=F_t$;

②$Y_s=Y_t$;

③$P_s(x_s)\neq P_t(x_t)$,其中 P 表示数据集的边缘概率分布;

④$Q_s(y_s|x_s)=Q_t(y_t|x_t)$,其中 Q 表示数据集的条件概率分布;

则称数据集 D_s 为源域数据集,数据集 D_t 为目标域数据集.

定义3(无监督域适应分类)[17]　给定源域数据集 $D_s=(X_s,Y_s)$ 和目标域数据集 $D_t=(X_t,Y_t)$,以及学习任务 T.任务 T:利用 X_s、Y_s 和 X_t 学习一个分类器 $f:x_s\rightarrow y_s$,能够对目标域 D_t 的标签 $y_t\in Y_t$ 进行预测.称这个任务 T 为无监督领域适应分类任务.

2.2　多核最大均值差异

定义4(最大均值差异(MMD))[18]　设 H_k 是定义在拓扑空间 χ_k 上的再生核Hilbert空间(RKHS),p 和 q 分别是 χ_k 上的Borel概率度量,f 是 H_k 上的特征映射,$\mu_k(p)\in H_k$ 和 $\mu_k(q)\in H_k$ 分别是 p 和 q 在 H_k 中的平均嵌入.若对于 $\forall f\in H_k$,满足 $E_{x\sim p}f(x)=\langle f,\mu_k(p)\rangle_{H_k}$ 且 $E_{x\sim q}f(x)=\langle f,\mu_k(q)\rangle_{H_k}$,则 p 和 q 之间的MMD距离为:

$$\begin{aligned}\eta_k(p,q)&=\|\mu_k(p)-\mu_k(q)\|^2_{H_k}\\&=E_{xx'}k(x,x')+E_{yy'}k(y,y')-2E_{xy}k(x,y)\end{aligned}\tag{1}$$

其中 $x,x'\overset{\text{i.i.d.}}{\sim}p$,$y,y'\overset{\text{i.i.d.}}{\sim}q$,核函数 $k(\cdot)$ 一般选择表示无穷维的高斯核函数:

$$k(x,x')=\exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)\tag{2}$$

若令 $h_k(x,x',y,y')=k(x,x')+k(y,y')-k(x,y')-k(x',y)$,$v=[x,x',y,y']$,则有:

$$\eta_k(p,q)=E_{xx'yy'}h_k(x,x',y,y')=E_vh_k(v)\tag{3}$$

MK-MMD在原始MMD特征核 $k(x,x')$ 的基础上,使用多个不同高斯核函数 $\{k_u\}$ 的凸组合形成一个复合核函数,其可利用不同核函数来增强距离度量性能,从而能够更准确地将输入空间的值映射到RKHS以得到最优值.

定义5(多核最大均值差异(MK-MMD))　设 $\{k_u\}_{u=1}^d$ 是一组正定函数,且满足 $k_u:\chi_k\times\chi_k\rightarrow R$,则 $\exists D>0$,使得总内核函数为:

$$K:=\left\{k:k=\sum_{u=1}^d\beta_uk_u,\sum_{u=1}^d\beta_u=D,\beta_u\geq0,\forall u\in\{1,\cdots,d\}\right\}\tag{4}$$

其中系数 $\{\beta_u\}$ 的约束以保证派生的多核 k 是特征性的,d 为内核的数量.若 k 是有界的,且 $\forall k\in K$ 都与一个 RKHSH_k 唯一关联,根据定义4,p 和 q 之间的MK-MMD距离为:

$$\phi_k(p,q)=\sum_{u=1}^d\beta_u\eta_u(p,q)\tag{5}$$

其中,$\eta_u(p,q)=E_vh_u(v)$,且 $\eta_u(p,q)$ 中使用的特征核 $k_u\in K$.

2.3　GTA模型

GTA模型的学习过程由分类分支和对抗分支两个并行流共同完成:(1)分类分支通过使用源域标记数据学习一个特征提取网络到预测网络的组合,预测网络用于对源域真实数据进行分类,并将各个类别空间的梯度信息回传给特征提取网络;(2)对抗分支通过特征提取网络学习源域数据和目标域数据的共享特征并作为GAN的输入,生成器负责生成类似源的源域生成图像和目标域生成图像,判别器同时作为二分类器和多分类器,使用源域和目标域上的无标记数据进行对抗训练,并仅使用源域标签进行监督学习.最后,预测网络通过利用特征提取网络和GAN提供的知识在目标域中进行识别.图1是GTA模型的流程图.

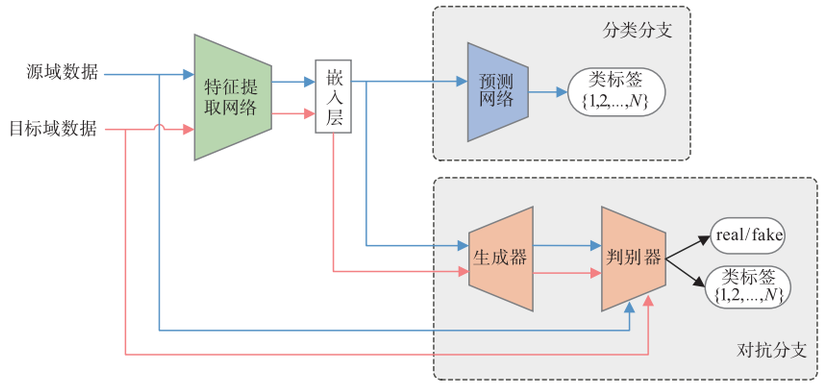


图1 GTA模型流程图

3 UDAG 模型

3.1 问题描述

通过对 GTA 模型的结构研究分析发现存在以下问题:一是预测网络和判别器虽然同时作为多分类器,但由于两者之间缺少信息交互,判别器上带有源域和目标域共享特征的生成图像信息不能直接传递给预测网络,导致分类性能只能极大程度地取决于特征提取网络的性能;二是仅使用 GAN 的对抗学习能力不足以最小化源域分布和目标域分布之间的距离,缺少一种距离度量准则以辅助 GAN 学习共享嵌入,从而改善域适应效果.

受 GTA 模型结构的启发,并针对其存在的不足,本文提出基于 GAN 的无监督域适应分类模型 UDAG,其流程图如图 2 所示.

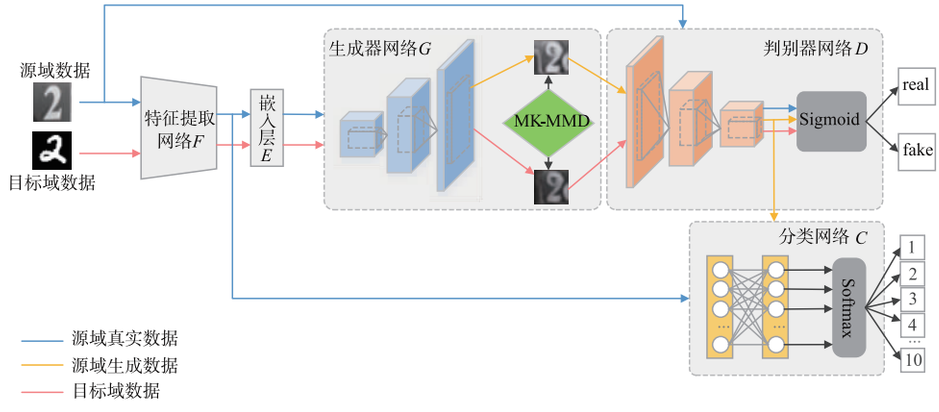


图2 UDAG模型流程图

3.2 结构流程

设 m 为源域和目标域输入数据大小,下面分别对 UDAG 模型四个网络结构的特点及作用进行介绍:

(1)F 使用卷积神经网络结构,并将源域真实数据 x_{sm} 和目标域真实数据 x_{tm} 作为其输入. 在模型训练过程中,F 能够逐渐学习到两个领域的共同特征,并产生输出 $f_{sm}=F(x_{sm})$ 和 $f_{tm}=F(x_{tm})$. 嵌入层 E 将服从标准正态分布 N(0,1) 的随机噪声 z_m,分别与 f_{sm} 和 f_{tm} 连接作为 G 的输入,即 $e_{sm}=E(f_{sm},z_m)$ 和 $e_{tm}=E(f_{tm},z_m)$.

(2)G 在与 D 进行无监督对抗训练中,逐渐生成特征分布一致的源域生成图像 $G(e_{sm})$ 及目标域生成图像 $G(e_{tm})$. MK-MMD 作用在 $G(e_{sm})$ 和 $G(e_{tm})$ 之间,在 GAN 以对抗训练拉近源域和目标域分布距离的基础上,进一步减小域间数据分布差异. 根据定义 5,MK-MMD 在 G 上的损失函数如式(6)所示. 其中,γ 为平衡系数,用于控制 MK-MMD 减小域间差异的强度.

$$L_{\mathrm{mk-mmd}}=\gamma\min(\phi_k(G(e_{sm}),G(e_{tm}))) \tag{6}$$

(3)D 接受三种输入数据来源:x_{sm}、$G(e_{sm})$ 和 $G(e_{tm})$,其目的是将输入数据 x_{sm} 判定为“真”,将输入数据 $G(e_{sm})$ 和 $G(e_{tm})$ 判定为“假”,即 $D(x_{sm})\to 1$、$D(G(e_{sm}))\to 0$、$D(G(e_{tm}))\to 0$. 真实数据 x_{sm} 的加入,能够指导 $G(e_{sm})$ 和 $G(e_{tm})$ 在训练过程中正确学习真实数据分布,避免 G 出现不收敛的情况. x_{sm}、$G(e_{sm})$ 和 $G(e_{tm})$ 在 D 上的对抗损失函数分别如式(7)~(9)

所示.

$$L_{\text{adv,sr}} = \max_{D} \frac{1}{m}\sum_{i=1}^{m}\log(D(x_{si})) \tag{7}$$

$$L_{\text{adv,sf}} = \max_{D} \frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(e_{si}))) \tag{8}$$

$$L_{\text{adv,tf}} = \max_{D} \frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(e_{ti}))) \tag{9}$$

(4) C 接收 f_{sm} 和通过 D 结构的源域生成数据 $D'(G(e_{sm}))$ 作为输入,并使用相同的源标签对其进行训练. f_{sm} 提供源域真实数据和目标域真实数据的共享特征信息,指导 C 同时对源域和目标域的数据进行正确分类;当 GAN 达到纳什均衡时,$D'(G(e_{sm}))$ 将带有源域生成数据和目标域生成数据的共享特征信息传递给 C,该信息的加入增加了 C 与 GAN 网络的信息交互,防止其在真实源域数据上训练时产生过拟合现象,提高 C 的泛化能力. 并且 C 将源域分类损失梯度信号反向传播给 D、G 以及 F,增强 D、G、F 处理分类特征信息的能力,从而进一步提高 C 在目标域上的分类能力. 四个网络模块的更新会产生相互影响,且根据反向传播算法,在同一批次的训练中,需优先更新 C,然后依次更新 D、G,F 因需要接受来自其他三个网络的回传梯度,需最后更新. f_{sm} 和 $D'(G(e_{sm}))$ 在 C 上的分类损失分别如式(10)和式(11)所示.

$$L_{\text{cls,f}} = \min_{C}\left(-\frac{1}{m}\sum_{i=1}^{m} y_i \cdot \log(C(f_{si}))\right) \tag{10}$$

$$L_{\text{cls,d}} = \min_{C}\left(-\frac{1}{m}\sum_{i=1}^{m} y_i \cdot \log(C(D'(G(e_{si}))))\right) \tag{11}$$

3.3 算法流程

UDAG 模型的整体算法流程如算法 1 所示.

算法 1　UDAG 模型训练

Input:源域数据集 $D_s=(X_s,Y_s)$,目标域数据集 $D_t=(X_t,Y_t)$,训练次数 K,批量大小 m,平衡系数 γ.

Output:UDAG 模型 Π

1. 随机初始化模型 Π 中所有网络层的参数;
2. for k in 1;K do
   - 2.1 从 D_s 中采样 m 个有标记数据 $\{x_{si},y_i\}_{i=1}^{m}$,记为 x_{sm};从 D_t 中采样 m 个无标记数据 $\{x_{ti}\}_{i=1}^{m}$,记为 x_{tm};
   - 2.2 x_{sm} 通过 F 网络计算得到 $f_{sm}=F(x_{sm})$;x_{tm} 通过 F 网络计算得到 $f_{tm}=F(x_{tm})$;
   - 2.3 随机产生 m 个噪声样本,记为 $z_m=\{z_i\}_{i=1}^{m}$;
   - 2.4 f_{sm} 和 z_m 通过嵌入层 E 得到 $e_{sm}=E(f_{sm},z_m)$,f_{tm} 和 z_m 通过嵌入层 E 得到 $e_{tm}=E(f_{tm},z_m)$;
   - 2.5 根据式(10)和式(11),计算分类器 C 的损失函数:
     $$L_C=L_{\text{cls,f}}+L_{\text{cls,d}};$$
   - 2.6 根据式(7)~(9)和式(11),计算判别器 D 的损失函数:
     $$L_D=L_{\text{cls,d}}+L_{\text{adv,sr}}+L_{\text{adv,sf}}+L_{\text{adv,tf}};$$
   - 2.7 根据式(6)、式(8)和式(11),计算生成器 G 的损失函数:
     $$L_G=L_{\text{cls,d}}+L_{\text{adv,sf}}+\gamma\cdot L_{\text{mk-mmd}};$$
   - 2.8 根据式(9)~(11),计算特征提取网络 F 的损失函数:
     $$L_F=L_{\text{cls,d}}+L_{\text{cls,f}}+L_{\text{adv,tf}};$$
   - 2.9 使用梯度下降法反向传播梯度信号,以更新 C、D、G 和 F 中的参数;
3. end for
4. 输出 UDAG 模型 Π,算法停止.

4 实验与分析

为了说明 UDAG 模型的有效性,本文进行了三种不同的实验:(1)分类准确率对比实验;(2)生成图像效果对比实验;(3)t-SNE 图可视化实验.

4.1 实验设置

4.1.1 数据集

为避免实验结果在单一数据集上出现偶然性,选择在 MNIST[19]、USPS[20] 和 SVHN[21] 三种公共数据集上进行实验验证,并设置了四种常见的域适应情况:SVHN→MNIST、MNIST→USPS、MNIST→USPS(p)和 USPS→MNIST. 其中,MNIST→USPS(p)使用文献[22]的相同设置,$A\to B$ 表示 A 作为源域、B 作为目标域的域适应学习.

4.1.2 参数设置

实验采用小批量随机梯度下降法进行训练,批量大小为 100,学习率设置为 0.0005,学习率衰减参数为 0.0001,并使用 Adam 优化器进行梯度更新,其中指数衰减参数 β_1 为 0.8,β_2 为 0.999.

4.2 分类准确率对比实验

4.2.1 实验步骤

Setp1　选取相应数据集的训练数据作为源域数据集 D_{s1} 和目标域数据集 D_{t1},并产生 4 个随机种子用以随机打乱数据集,得到源域数据集 D_{s2}、D_{s3}、D_{s4}、D_{s5} 和目标域数据集 D_{t2}、D_{t3}、D_{t4}、D_{t5};

Step2　分别使用 D_{s1} 和 D_{t1}、D_{s2} 和 D_{t2}、D_{s3} 和 D_{t3}、D_{s4} 和 D_{t4}、D_{s5} 和 D_{t5},按照算法 1 的步骤进行训练,得到 UDAG 模型 Π_1、Π_2、Π_3、Π_4 和 Π_5;

Step3　固定模型 Π_1 中的 F 网络和 C 网络分支,并将其作为测试模型 Ψ_1. 同样的步骤,获得测试模型 Ψ_2、Ψ_3、Ψ_4 和 Ψ_5;

Step4　使用测试模型 Ψ_1 对目标域数据集 D_{t1} 进行预测,计算出分类精度. 同样的步骤,计算出 Ψ_2、Ψ_3、Ψ_4 和 Ψ_5 分别在 D_{t2}、D_{t3}、D_{t4}、D_{t5} 上的分类精度;

Step5　取 5 个预测模型分类精度的平均值作为最终模型的分类精度;

Step6　根据 4.1.1 节,在四种不同域适应情况下,重复以上步骤,得到每种域适应情况下的分类精度.

表 1 UDAG 模型与其他域适应方法的分类精度(均值±方差%)对比

	SVHN→MNIST	MNIST→USPS	MNIST→USPS(p)	USPS→MNIST
DANN	76.0±1.8	-	89.4±0.2	90.1±0.8
DDC	68.1±0.3	-	79.1±0.5	66.5±3.3
DRCN	82.0±0.16	-	91.8±0.09	73.7±0.04
ADDA	76.0±1.8	-	89.4±0.2	90.1±0.8
DIFA	89.7±2.0	96.2±0.2	92.3±0.1	89.7±0.5
SBADA-GAN	76.1	97.6	-	95.0
CycADA	88.3±0.2	94.8±0.2	-	95.7±0.2
GTA	92.4±0.9	95.3±0.7	92.8±0.9	90.8±0.9
UDAG	**94.3±1.6**	**97.5±0.1**	**93.3±0.4**	**98.3±0.1**

4.2.2 结果分析

将 UDAG 模型与当前主流的其他域适应方法在 SVHN→MNIST、MNIST→USPS、MNIST→USPS(p)和 USPS→MNIST 四种域适应情况下进行分类精度比较,其实验结果如表 1 所示. 其中,分类精度最高的值用粗体表示,其次高的值用添加下划线的方式表示. 可以看出,在四种域适应情况下,UDAG 模型的分类精度都达到了最高值. 尤其是在 USPS→MNIST 的情况下,UDAG 模型的分类精度可以达到 98.3±0.1%,相较于 CycADA 的分类精度提高了 2.6% 左右,在其他三种域适应情况下,其分类精度也提高了 0.5%~1.9% 左右.

4.3 生成图像效果对比实验

4.3.1 实验步骤

Setp1 分别从 4.2.1 节中 D_{s1} 和 D_{t1} 的每个类别内随机选取 20 个数据,组成大小为 200 的源域测试数据集 D_s^* 和目标域测试数据集 D_t^*;

Step2 计算 D_s^* 和 D_t^* 之间的 MK-MMD 距离,作为未适配前的域间原始 MK-MMD 值;

Step3 按照算法 1 的步骤进行训练,当运行次数为 20、40、60 和 80 的时候,固定 F 网络和 G 网络分支,并分别作为预测模型 Ω_1、Ω_2、Ω_3 和 Ω_4;

Step4 使用 D_s^* 和 D_t^* 在预测模型 Ω_1、Ω_2、Ω_3、Ω_4 上分别产生源域生成数据集 $\hat{D}_{s1}$、$\hat{D}_{s2}$、$\hat{D}_{s3}$、$\hat{D}_{s4}$ 和目标域生成数据集 $\hat{D}_{t1}$、$\hat{D}_{t2}$、$\hat{D}_{t3}$、$\hat{D}_{t4}$;

Step5 分别使用 $\hat{D}_{s1}$ 和 $\hat{D}_{t1}$、$\hat{D}_{s2}$ 和 $\hat{D}_{t2}$、$\hat{D}_{s3}$ 和 $\hat{D}_{t3}$、$\hat{D}_{s4}$ 和 $\hat{D}_{t4}$ 计算出相应的域间 MK-MMD 值.

Step6 根据 4.1.1 节,在四种不同域适应情况下,重复以上步骤,得到每种域适应情况下的 MK-MMD 值.

4.3.2 结果分析

表 2 是 UDAG 模型生成的图像及相应的 MK-MMD 值,可以看出在四种域适应情况下,随着训练次数据的增加,源域生成图像和目标域生成图像的质量都在逐渐提高,且两个域的生成图像特征都逐渐接近于源域真实图像特征,证明 UDAG 模型在训练过程中确实学习到了域间共享特征,并生成了分布一致的源域生成图像和目标域生成图像. 与此同时,MK-MMD 值也随着运行次数呈现出不断减小的趋势,证明 UDAG 模型确实能够优化域间距离,并且可以得出结论,在相同的运行次数下,源域生成分布和目标域生成分布之间的差异性与 MK-MMD 值成正相关关系.

4.4 t-SNE 图可视化实验

4.4.1 实验步骤

Setp1 分别从 4.2.1 节中 D_{s1} 和 D_{t1} 的每个类别内随机选取 25 个数据及相应标签,组成大小为 250 的源域测试数据集 D_s^{**} 和目标域测试数据集 D_t^{**};

Step2 将 D_s^{**} 和 D_t^{**} 的数据进行归一化操作,并计算出相应的 t-SNE 值 $t1$ 和 $t2$;

Step3 使用 $t1$ 和 $t2$ 及 D_s^{**} 和 D_t^{**} 的相应标签绘制出源域和目标域未适应前的 t-SNE 图[23];

Step4 固定 4.2.1 节 UDAG 模型 Π_1 中的 F 网络和 G 网络分支作为预测模型 Γ,使用 D_s^{**} 和 D_t^{**} 在 Γ 上产生适应后的源域生成数据 D_s'' 和目标域生成数据 D_t'';

Step5 使用 D_s'' 和 D_t'' 重复 Step2、Step3,得到源域和目标域适应后的 t-SNE 图;

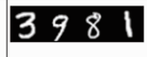
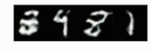
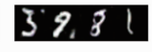
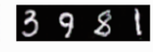
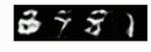
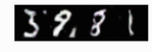
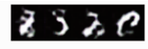
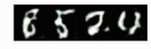
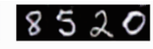
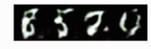
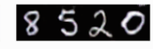
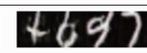
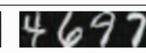
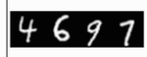
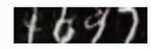
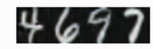
Step6 根据 4.1.1 节,在四种不同域适应情况下,重复以上步骤,得到每种域适应情况下的源域和目标域未适应前的 t-SNE 图及适应后的 t-SNE 图.

4.4.2 结果分析

图 3~6 分别是在四种域适应情况下,源域数据分布和目标域数据分布未适应前和适应后的 t-SNE 图. 可以看到,在未适应前源域数据和目标域数据散乱地分布在一起,也未观察到任何域间适应信息和分类信息. 而在使用 UDAG 模型对其进行域适应分类后,源域数据和目标域数据开始以不同的类别聚集在一起,特别是在 MNIST 与 USPS 的域适应情况下,源域数据和目标

域数据类间距离较小,类与类之间的距离较大. 而在SVHN→MNIST 的情况下,两个域的数据按类别聚集在一起的效果较差,这可能与原始数据的域间差异程度有关.

表 2　UDAG 模型生成的图像及 MK-MMD 值

		原始图像	运行20次	运行40次	运行60次	运行80次
SVHN→MNIST	源					
	目标					
	MK-MMD值	2.1966	1.5307	0.9025	0.8236	0.1659
MNIST→USPS	源					
	目标					
	MK-MMD值	0.7106	0.0176	0.0057	0.0025	0.0019
MNIST→USPS(p)	源					
	目标					
	MK-MMD值	0.6572	0.2031	0.0707	0.0555	0.0221
USPS→MNIST	源					
	目标					
	MK-MMD值	0.7401	0.0265	0.0024	0.0022	0.0018

未适应前　　适应后

图3　SVHN→MNIST的t-SNE图

未适应前　　适应后

图4　MNIST→USPS的t-SNE图

未适应前　　适应后

图5　MNIST→USPS(p)的t-SNE图

未适应前　　适应后

图6　USPS→MNIST的t-SNE图

5 结束语

为解决 GTA 模型部分网络结构缺少信息交互以及仅使用 GAN 框架减小域间差异的能力较弱,从而导致域适应分类精度不高的问题,本文提出一种基于 GAN 的无监督域适应分类模型 UDAG,该模型学习源域数据分布和目标数据分布的共享特征嵌入,并联合使用 GAN 和 MK-MMD 度量准则进一步减小域间距离. 同时,该模型还利用无监督对抗训练及监督分类训练共同构造出分类精度高的域适应分类器,在四种域适应情况下的实验结果显示了 UDAG 模型在无监督域适应分类中的优越性. 另外,对于如何减小模型中因 GAN 的对抗训练引起的波动性是本文需要进一步研究的问题.

参考文献

[1] WANG X, GUPTA A. Unsupervised learning of visual representations using videos[A]. Proceedings of the IEEE International Conference on Computer Vision[C]. Santiago, Chile: IEEE, 2015. 2794 – 2802.

[2] MAHJOURIAN R, WICKE M, ANGELOVA A. Unsupervised learning of depth and ego-motion from monocular video using 3D geometric constraints[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City, USA: IEEE, 2018. 5667 – 5675.

[3] 刘建伟,孙正康,等. 域自适应学习研究进展[J]. 自动化学报,2014,40(8):1576 – 1600.
LIU Jian-Wei, SUN Zheng-Kang, et al. Review and research development on domain adaptation learning[J]. Acta Automatica Sinica, 2014, 40(8): 1576 – 1600. (in Chinese)

[4] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345 – 1359.

[5] ROZANTSEV A, SALZMANN M, FUA P. Beyond sharing weights for deep domain adaptation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(4): 801 – 814.

[6] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[EB/OL]. https://arxiv.org/abs/1409.7495v1, 2015 – 02 – 27.

[7] GHIFARY M, KLEIJN W B, ZHANG M, et al. Deep reconstruction-classification networks for unsupervised domain adaptation[A]. European Conference on Computer Vision[C]. Amsterdam, the Netherlands: Springer, 2016. 597 – 613.

[8] TZENG E, HOFFMAN J, ZHANG N, et al. Deep domain confusion: Maximizing for domain invariance[EB/OL]. https://arxiv.org/abs/1412.3474, 2014 – 12 – 10.

[9] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial discriminative domain adaptation[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Hawaii, USA: IEEE, 2017. 7167 – 7176.

[10] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[A]. Advances in Neural Information Processing Systems[C]. Montreal, Canada: MIT Press, 2014. 2672 – 2680.

[11] 王万良,李卓蓉. 生成式对抗网络研究进展[J]. 通信学报,2018,39(2):135 – 148.
WANG Wan-Liang, LI Zhuo-Rong. Advances in generative adversarial network[J]. Journal on Communications, 2018, 39(2): 135 – 148. (in Chinese)

[12] RUSSO P, CARLUCCIi F M, TOMMASI T, et al. From source to target and back: symmetric bi-directional adaptive gan[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City, USA: IEEE, 2018. 8099 – 8108.

[13] VOLPI R, MORERIO P, SAVARESE S, et al. Adversarial feature augmentation for unsupervised domain adaptation[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City, USA: IEEE, 2018. 5495 – 5504.

[14] HOFFMAN J, TZENG E, PARK T, et al. Cycada: cycle-consistent adversarial domain adaptation[EB/OL]. https://arxiv.org/abs/1711.03213, 2017-12-29.

[15] SANKARANARAYANAN S, BALAJI Y, CASTILLO C D, et al. Generate to adapt: aligning domains using generative adversarial networks[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City, USA: IEEE, 2018. 8503 – 8512.

[16] GRETTON A, SEJDINOVIC D, STRATHMANN H, et al. Optimal kernel choice for large-scale two-sample tests[A]. Advances in Neural Information Processing Systems[C]. Nevada, USA: MIT Press, 2012. 1205 – 1213.

[17] LI J, LU K, HUANG Z, et al. Transfer independently together: a generalized framework for domain adaptation[J]. IEEE Transactions on Cybernetics, 2018 (99): 1 – 12.

[18] GRETTON A, FUKUMIZU K, HARCHAOUI Z, et al. A fast, consistent kernel two-sample test[A]. Advances in Neural Information Processing Systems[C]. Vancouver, Canada: MIT Press, 2009. 673 – 681.

[19] DENG L. The mnist database of handwritten digit images for machine learning research [best of the web][J]. IEEE Signal Processing Magazine, 2012, 29(6): 141 – 142.

[20] CHAABAN I, SCHEESSELA M R. Human Performance

on the USPS Database[R]. South Bend:Indiana University South Bend,2007.

[21] NETZER Y,WANG T,COATES A,et al. Reading digits in natural images with unsupervised feature learning[A]. NIPS Workshop on Deep Learning and Unsupervised Feature Learning[C]. Granada,Spain:MIT Press,2011.

[22] BOUSMALIS K,TRIGEORGIS G,SILBERMAN N,et al. Domain separation networks[A]. Advances in Neural Information Processing Systems[C]. Barcelona,Spain:MIT Press,2016. 343 – 351.

[23] MAATEN L,HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research,2008,9(3):2579 – 2605.

作者简介

王格格　女,1995 年生于重庆市. 现为四川师范大学计算机科学学院硕士研究生. 主要研究方向为人工智能与深度学习.
E-mail:347673996@qq.com

郭　涛(通信作者)　女,1967 年生于四川省雅安市. 硕士,现为四川师范大学计算机科学学院教授、硕士生导师. 主要研究方向为数据挖掘与移动学习.
E-mail:tguo@sicnu.edu.cn

余　游　男,1993 年生于四川省遂宁市. 现为四川师范大学计算机科学学院硕士研究生. 主要研究方向为深度学习与数据挖掘.
E-mail:454665275@qq.com

苏　菡　女,1979 年生于四川省阿坝藏族羌族自治州. 博士,现为四川师范大学计算机科学学院教授、硕士生导师. 主要研究方向为智能信息计算及处理、模式识别与图像处理.
E-mail:jkxy_sh@sicnu.edu.cn