

聚类辅助特征对齐的域适应方法

袁 丁 王 茜 邓李维

(重庆大学计算机学院 重庆 400044)

摘 要 有监督深度学习在有大量标记数据的领域可以取得不错的效果,但实际上很多领域只有大量未标记的数据。如何利用大量无标记数据,成为了深度学习发展的一个关键问题,领域自适应就是解决这一问题的一种有效方法。目前,基于对抗训练的域适应方法取得了较好的效果,这类方法利用领域分类损失对源域和目标域的特征分布进行对齐,降低了两个领域特征表示的分布差异,使采用源域数据训练的模型可以应用在目标域数据上。现有的域适应方法是在适配后的特征上进行模型训练的,没有充分利用目标域数据的原始信息,当两个领域差异较大时,会降低目标域特征的域内可鉴别性。针对现有方法的弱点,文中基于对抗判别域适应方法(ADDA),提出了一种基于对目标域数据聚类辅助特征对齐的域适应方法(CAFA-DA)。该方法通过聚类获得目标域数据伪标记,并在域适应阶段约束特征编码器训练,利用目标域数据的原始信息提高目标域特征的可鉴别性;将聚类和域适应两个过程训练的分类器进行集成学习,用高置信度样本进行训练,以提升模型的最终效果。CAFA-DA 可用在任何基于对抗损失的领域自适应方法上。最后,在领域自适应的 4 个标准数据集上将 CAFA-DA 方法与目前几种先进的领域自适应方法进行了对比实验,结果表明:CAFA-DA 方法的实验结果比其他几种方法都好;相对于 ADDA 方法,该方法在 2 个对比实验上的准确率分别提升了 3.2% 和 17.2%。

关键词 域适应,特征对齐,聚类,集成学习,特征表示

中图分类号 TP181 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.03.033

Clustering Assist Feature Alignment for Unsupervised Domain Adaptation

YUAN Ding WANG Qian DENG Li-wei

(College of Computer Science, Chongqing University, Chongqing 400044, China)

Abstract Supervised deep learning can reach good results in the areas with large amounts of labeled data, but the reality is that there are only a lot of unlabeled data in many areas. How to take advantages of large amounts of unlabeled data has become a key issue in the development of deep learning. Domain adaptation is an effective way to solve this problem. At present, domain adaptation methods based on adversarial training have achieved a good effect. This method uses domain classification loss to align the feature distribution of source domain, and target domain and reduce the difference of distribution between the feature representations of two domain, so the classifier trained with source domain data can be applied to target domain data. The existing domain adaptation method trains the model on the features after domain adaptation and does not make full use of the original information of the target domain data. When the differences between two domains are large, the intra-domain discriminability of target domain features will be reduced. In view of the disadvantages of the present methods, this paper proposed a method for clustering target domain data to assist feature alignment(CAFA-DA) based on the adversarial discriminative domain adaptation (ADDA). Pseudo-labels of target domain data are obtained by clustering and the feature encoder training is constrained in the domain adaptation stage, and the original information of the target domain data is used to improve the discriminability of target domain features. Classifiers trained in the two processes of clustering and domain adaptation are used for ensemble learning and high confidence samples are trained to improve the final effect of the model. The CAFA-DA can be applied to any domain adaptation method based on adversarial loss. Finally, this paper compared CAFA-DA with several advanced domain adaption methods on four standard domain adaption data sets. The results show that the accuracy of the CAFA-DA method is better than other methods. Compared with the ADDA method, the results of two comparative experiments are improved by 3.2% and 17.2% respectively.

Keywords Domain adaption, Feature alignment, Clustering, Ensemble learning, Feature representation

1 引言

随着计算能力的增强和数据的积累,深度神经网络凭借强大的学习能力,在图像、语音和语义等领域得到了广泛应用,并取得了很好的效果。在有监督学习下,深度学习在图像识别领域能取得比人类更高的准确率,但前提是训练数据有标记,且测试数据和训练数据必须有相同的分布。当测试数据与训练数据的分布不完全一致时,模型的效果就会变得很差,这种现象被称为数据集偏差或者领域差异^[1]。但实际情况中往往没有足够的有标记数据,如果使用其他领域的标记数据来训练模型,则数据分布问题会导致模型效果比较差。对未标记的数据进行人工标注,不仅代价昂贵,而且模型的效果与标记质量直接相关。域适应方法通过降低有标记的源域数据和无标记的目标域数据的分布差异来解决这个问题。域适应方法一般有3个基本模块:(1)基于领域特征差异学习特征编码器,降低源域和目标域特征表示的差异,即进行域适配,有些文献称其为特征对齐;(2)利用源域数据和特征编码器训练分类器;(3)通过编码器对目标域数据编码,将分类器应用在目标域数据上。

现有的域适应方法大多是在以下几个方面进行优化和改进。

(1)源域和目标域特征表示的相似度,即编码器产生特征的领域不变性,这决定了利用源域数据训练的模型能否用在目标域数据上。根据如何表示两个领域特征表示的差异,可以将现有方法分为两类^[2]:1)直接使用距离度量的方法^[3-5],这类方法一般会把两个领域的最大平均差异(MMD)^[6]降到最小;2)使用对抗的方法^[2,7-8],这类方法会学习一个鉴别器来区分编码器产生的特征来自哪个领域。这类方法的训练方式类似于生成对抗网络(GAN)^[9-10],通过优化对抗目标来优化编码器。当领域数据分布比较复杂时,基于距离度量的方法有时不能很好地表示领域间的差异;而基于对抗的方法因为深度网络的表示能力较强,效果会好于前者。

(2)源域特征的域内可鉴别性,即对源域数据分类有用的特征。文献^[2,7-8]采用深度神经网络作为分类器,通过优化分类器的交叉熵损失,使源域的特征表示具有域内可鉴别性。其中,文献^[8]抽取了源域的私有特征和共享特征,并通过正交损失使两个特征差异最大化。文献^[11]通过结合虚拟对抗训练(VAT)^[12],对源域的特征表示施加了平滑正则约束。

(3)目标域特征的域内可鉴别性,即对最后目标域数据分类有用的特征。早期的方法^[2,7]没有通过损失函数对目标域编码器的训练施加约束,文献^[13-14]通过对目标域样本进行预测,并利用高置信度的样本迭代训练来辅助模型学习目标域特征的可鉴别性。由于特征在对齐过程中会有损失,因此当源域和目标域差距较大时,用域适应学习到的分类器进行迭代训练或者进行集成学习的效果不一定好,而且难以保证两个分类器的多样性。文献^[11,15]通过结合VAT和条件熵^[16]损失,来辅助目标域特征编码器的训练。VAT方法的时间复杂度高,不容易找到最优化的方向,对模型训练的帮助有限。文献^[8]通过图像重建损失的约束来减少在域适配过程中有用特征的损失。但是,任务最终关注的是分类准确率,

样本的可重建特征并不一定对分类任务有用。

本文利用聚类算法获得数据原始的统计信息,帮助进行特征对齐,以提高目标域特征的可鉴别性。具体过程为:(1)利用聚类结果优化特征编码器的训练,通过优化带伪标签目标域的交叉熵损失来提高目标域特征的可鉴别性;(2)将聚类结果训练的分类器和原域适应的分类器进行集成学习,用高置信度的样本对模型进行再次训练。聚类过程中,分类器的训练数据不会有特征损失的问题,而且不同方法训练的两个分类器更具多样性。

本文第2节介绍域适应领域的相关工作;第3节介绍提出的聚类辅助特征对齐的域适应方法;第4节介绍实验的相关细节和结果;最后总结全文。

2 相关工作

因为深度学习模型要达到较好的训练效果,需要大量有标记数据,但实际上很多领域只有大量无标记数据,所以利用现实中大量的无标记数据来训练模型具有重要意义,因此领域自适应方法在最近几年成为了研究热点。领域自适应方法主要是通过降低源域和目标域间特征表示的差异,来利用大量无标记的目标域数据训练模型,通常是将两个领域的样本映射到一个具有域不变特征的空间,以减少领域差异^[1]。

为了减少源域和目标域分布的差异,文献^[17]根据极大似然估计值对源域样本加上不同的权重,最小化加权后的源域数据和目标域数据分布的差异。文献^[18]提出了一种新的分布距离的度量方式,并对样本进行加权表示,当两个领域分布差异较大时,基于样本加权的方法会因为源域中距离目标域较近的样本点很少而效果较差。

文献^[3-5]使用了MMD距离度量。MMD通过核函数将两个领域映射在一个再生核希尔伯特空间(RKHS)中,用映射后源域和目标域分布的均值之差表示两个分布之间的差异。文献^[3]在分类损失的基础上加了一层适配层,通过在适配层上引入MMD距离度量来最小化两个领域的分布差异。文献^[4]在文献^[3]的基础上进行了扩展,通过多层适配和多核MMD来最小化域间差异。文献^[5]在文献^[4]的基础上引入了联合概率分布。

基于对抗训练来降低领域分布差异的方法与对抗生成网络GAN^[9,19]密切相关,其目标是通过迭代优化学习一个特征编码器和一个领域特征鉴别器,这种方法通过对抗训练来最小化特征的领域差异,学习一种既有域间不变性又有域内可鉴别性的特征表示。文献^[20-21]利用二分类问题来处理特征的域间不变性,通过最大化鉴别器的分类误差来最小化领域差异。文献^[8,22]通过样本的重建损失约束特征编码器的训练,在提高特征的域间不变性的同时,保持特征的可鉴别性。文献^[8]分别抽取了两个领域的私有特征和共享特征,并通过正交损失使两个特征差异最大化。文献^[23-25]采用了生成模型,将生成任务与域适应任务结合在一起。

有一些方法基于聚类假设帮助训练特征编码器。聚类假设^[26]指数据的分布应该保持类内稠密而类间稀疏。基于聚类假设的很多方法^[12,16,27]在半监督学习上取得了不错的效果,目前被广泛用于域适应。文献^[11]对源域和目标域的特

征学习均施加了聚类假设约束。文献[15]在文献[11]的基础上,通过训练两个有差异的分类器进行集成学习。文献[13]提出利用两个分类器集成对目标域数据进行标记,并利用标记数据继续训练模型。文献[14]对高置信度的目标域数据打标记,再利用标记数据反过来训练模型,并进行迭代训练。这几种方法都基于高置信度的预测是正确预测的假设^[28]。

文献[2]结合前面的工作提出了一个通用的领域自适应

框架,并实例化了一种 ADDA 方法,并将之应用到图像分类任务上。其结构简单,并在很多有挑战性的域适应任务上达到了较高水平。ADDA 方法的步骤如图 1 所示:(1)使用源域图像训练一个源域编码器和一个源域分类器;(2)通过对抗方式训练一个目标域编码器,以降低鉴别器对不同领域数据特征的区分能力;(3)使用目标域编码器生成目标域图像特征,使用源域分类器对特征进行分类。图中的虚线框表示固定参数的网络,不参与训练。

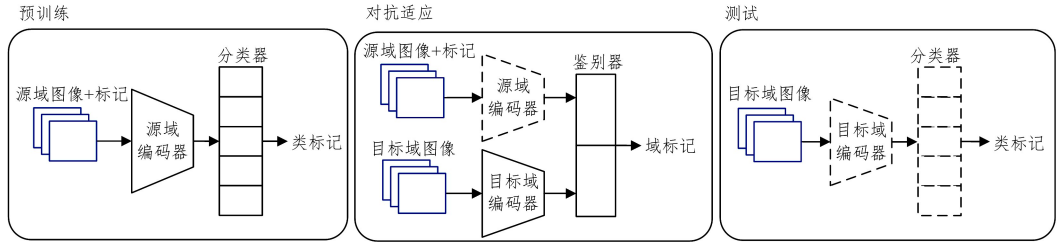


图 1 ADDA 方法

Fig. 1 Adversarial discriminative domain adaptation approach

3 CAFA-DA 方法

为了描述方便,首先约定任务和数据的表示形式。本文方法可以用在任何基于对抗形式的领域自适应方法上,为了进行实验和效果比较,主要关注图像分类任务。源域数据: $X_s = \{(x_i^s, y_i^s)\}, x_i^s \sim D_s\}_{i=0}^{N_s}$, 表示 N_s 个源域的有标记数据,数据分布为 D_s ; 目标域数据: $X_t = \{(x_i^t)\}, x_i^t \sim D_T\}_{i=0}^{N_t}$, 表示 N_t 个目标域的无标记数据,数据分布为 D_T 。两个领域的样本

具有相同的标记空间。 M_s 和 M_t 分别表示源域和目标域编码器, C_i 表示不同阶段训练的分类器。

ADDA 方法在训练目标域编码器时,只关注了目标域数据特征的域间不变性,当源域和目标域差距较大时,会降低目标域特征的可鉴别性,从而影响最终的分类效果。本文方法是对 ADDA 的改进,如图 2 所示,其中虚线框对应 ADDA 方法的相应步骤,实线框为本文对 ADDA 方法的改进。

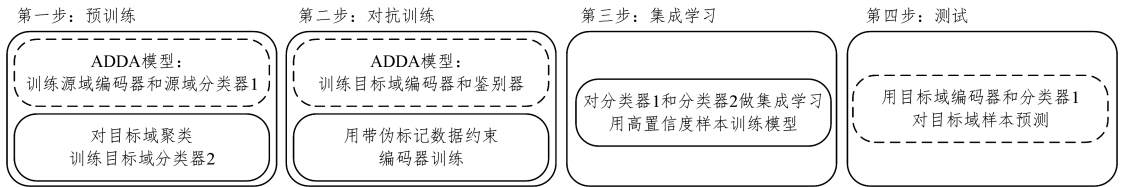


图 2 CAFA-DA 方法

Fig. 2 Clustering assist feature alignment for unsupervised domain adaptation approach

第一步,在 ADDA 方法的基础上增加了对目标域聚类的操作,根据聚类结果对目标域数据打上伪标记,并用带伪标记的数据训练一个目标域分类器;第二步,在对抗训练的过程中利用第一步产生的带伪标记的目标域数据训练一个分类器,通过结合分类器的损失约束目标域特征编码器的训练;第三步,用第一步中训练的两个分类器进行集成学习,选择高置信度的样本再次训练模型;第四步,与 ADDA 方法测试部分相同,即用目标域编码器和源域分类器对目标域数据进行预测。

3.1 目标域聚类

通过对目标域数据聚类获得目标域数据的伪标记,用带伪标记的样本训练目标域分类器 C_2 。这里有两个关键点:

(1) 获得目标域样本的伪标记,即通过聚类结果对目标域样本进行标记,并根据样本点到各个簇心的距离给样本 x_i 赋予一个权重 α_i 。

$$\alpha_i = (1 - \frac{(k-1)d_i^t}{\sum_{k \neq t} d_i^k})^2 \quad (1)$$

其中, d_i^t 表示样本 x_i 到本簇中心 t 的距离, d_i^k 表示样本 x_i 到

其他簇中心的距离。 α_i 的取值范围为 $0 \sim 1$, 样本点越接近簇中心,说明其属于该类的置信度越高, α_i 的值就越大;反之 α_i 的值越小。

(2) 使用带伪标记的目标域样本训练的分类器,将其用于第三步的集成学习。该分类器使用的特征未经过领域适配,可以最大程度地避免因特征对齐造成的对分类任务有用的特征的损失。ADDA 方法中的分类器使用的是经过领域适配的特征,其更多地关注两个域的共有信息。本文方法在第三步利用两个分类器进行集成学习,结合各自的优点选出高置信度的样本。

对目标域数据聚类和训练分类器采用了 DeepCluster^[29] 方法,如图 3 所示。DeepCluster 方法通过深度卷积网络 Convnet 提取样本特征,网络的输出分为两个分支:聚类分支和分类分支。分类算法和聚类算法共享 Convnet 参数。聚类分支对样本进行聚类,并根据聚类结果对目标域数据打上伪标记;分类器用标记的数据进行训练,并优化 Convnet 参数。算法进行迭代训练,不断优化聚类和分类的效果。这种互学习的

方法有利于两种任务的相互促进,最后使聚类 and 分类都达到很好的效果。通过 DeepCluster 方法处理,得到带伪标记的目标域数据和分类器 C_2 。

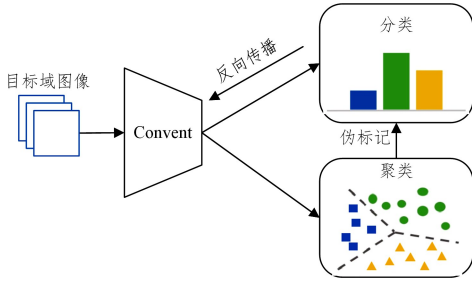


图3 DeepCluster方法

Fig. 3 DeepCluster approach

ADDA方法的第一步是根据源域数据训练一个分类器;本文方法第一步中源域分类器的训练与ADDA方法保持一致。优化目标如下:

$$\min_{M_s, C} l_{cls}(X_s, Y_s) = -E_{(x_s, y_s) \sim (X_s, Y_s)} \sum_{k=1}^K \mathbf{I}(k=y_s) \cdot \log C_1(M_s(x_s)) \quad (2)$$

其中, M_s 表示映射源域特征的编码器, C_1 表示源域分类器。

3.2 域适应训练

根据源域数据和第一步产生的带伪标记的目标域数据训练目标域特征编码器。ADDA方法只用了鉴别器来约束编码器学习,我们增加了带伪标记的目标域数据的分类损失,如图4所示。图中的虚线框表示固定参数的网络,不参与训练。

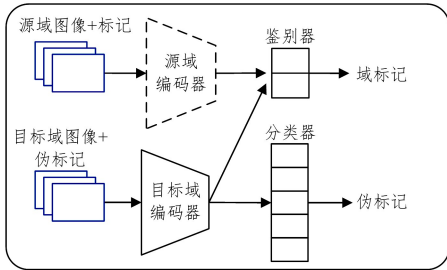


图4 对抗训练

Fig. 4 Adversarial training

ADDA方法的第二步基于对抗的域适应训练类似于GAN,即交替优化领域鉴别器和特征编码器。优化鉴别器可提高其对两个领域特征的识别准确度;优化目标域特征编码器可降低鉴别器的识别准确度。通过交替优化迭代训练,最终特征编码器可以将两个领域的图像映射到一个相似的空间中。这两个步骤的优化目标如下:

$$\min_D l_{advD}(X_s, X_t, M_s, M_t) = -E_{x_s \sim X_s} [\log D(M_s(x_s))] - E_{x_t \sim X_t} [1 - \log D(M_t(x_t))] \quad (3)$$

$$\min_{M_t, M_s, C_3} l_{advM}(X_s, X_t, D) = -E_{x_t \sim X_t} [\log D(M_t(x_t))] \quad (4)$$

ADDA方法在第二步只关注了两个编码器映射的特征是否相似。但是当两个领域的分布差异较大时,模型会因为鉴别器分类损失的约束减小特征的域间差异,降低目标域特征的域内可鉴别性,从而影响最终分类效果。本文方法通过目标域分类器交叉熵损失的约束,来降低因特征对齐

造成的有用特征的损失。在ADDA优化目标的基础上,增加了优化目标(5),其中 C_3 为本步骤训练的分类器。

$$\min_{M_t, C} l_{advM}(X_t, Y_t) = -E_{(x_t, y_t) \sim X_t} \sum_{k=1}^K \mathbf{I}(k=y_t) \cdot \alpha_t \log C_3(M_t(x_t)) \quad (5)$$

3.3 集成学习

本文方法增加了集成学习。在第一步学到两个分类器:(1)根据源域数据训练的分类器,用 C_1 表示;(2)聚类过程中学到的分类器,用 C_2 表示,如图5所示。图中的虚线框表示固定参数的网络,不参与训练。

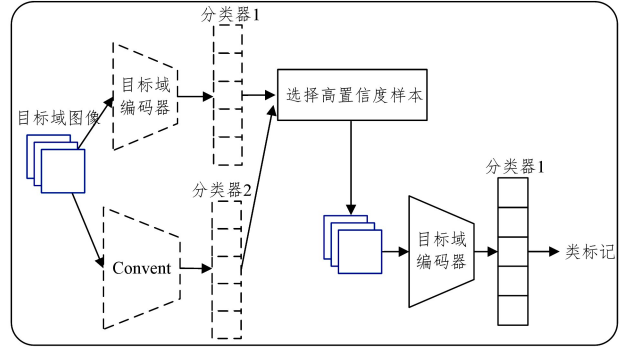


图5 集成学习

Fig. 5 Ensemble learning

本步骤可分为两个阶段。第一阶段对伪标记进行校正,学习到伪标记与真实标记的对应关系。通过 C_1 得到目标域样本的预测标记,并统计同一伪标记样本中预测标记的比例,将比例最高的预测标记作为伪标记对应的真实标记。公式如下:

$$n_i^k = \sum_{x_i \in y_i'} \sum_{k \in y} \mathbf{I}(C_1(x_i) = k) \quad (6)$$

$$y_i' = \max_k(n_i^k) \quad (7)$$

其中, y_i' 表示第 i 类伪标记, y 表示真实类标记的集合, k 表示真实类标记。当 $C_1(x_i) = k$ 时, $\mathbf{I}(C_1(x_i) = k) = 1$; 否则 $\mathbf{I}(C_1(x_i) = k) = 0$ 。式(6)中的 n_i^k 表示目标域样本中伪标记为 y_i' 、 C_1 预测为 k 的数量。式(7)对伪标记 y_i' 与真实标记 k 建立对应关系,其中 k 为目标域中伪标记为 y_i' 的样本预测比例最高的真实标记。

第二阶段结合分类器 C_1 和 C_2 预测目标域数据标记,用标记的高置信度样本再次训练 M_t 和 C_1 ,训练方法与第一步中 M_s 和 C_1 的训练方法相同,分别用原 M_t 和 C_1 的网络参数作为初始值。优化目标如下:

$$\min_{M_t, C_1} l_{cls}(X_t, Y_t) = -E_{(x_t, y_t) \sim (X_t, Y_t)} \sum_{k=1}^K \mathbf{I}(k=y_t) \cdot \log C_1(M_t(x_t)) \quad (8)$$

4 实验

本文提出的利用聚类辅助特征对齐和集成学习的方法可以用在其他基于对抗损失的域适应方法上。为了验证本文方法的效果,在领域自适应常用的数据集上将其与ADDA以及其他先进的方法进行对比。本文实验主要关注图像分类任务。

4.1 实现细节

(1)DeepCluster。为了保证聚类效果及基于伪标签的分

类器的效果,实验采用文献[29]的网络结构;Convent 采用标准的 AlexNet^[30],其包含 5 个卷积层和 3 个全连接层;训练采用 dropout^[31],batch size 为 256;聚类采用文献[32]实现的 k-means 算法。

(2)ADDA。为了对比实验结果,基于 ADDA 部分的实现与原模型保持一致:特征编码器和源域分类器采用 LeNet;领域鉴别器采用 4 层全连接前馈网络,前三层神经元数量分别为 500,500,500,最后一层为 softmax 层;第二步中的目标域分类器结构与源域分类器结构保持相同。

(3)集成学习。第三步集成学习中高置信度样本的选择阈值设置为 0.75,即将两个分类器预测相同且分数均在 0.75 以上的样本作为高置信度样本。在置信度阈值的选择上,过低的置信度会导致选择一些预测错误的样本,过高的置信度会导致可用样本数量较少。综合考虑,实验选取 0.5 和 1 的折中值 0.75 作为阈值。

4.2 实验数据集

(1)MNIST to MNIST-M。将手写数字数据集 MNIST^[33]作为源域,将 MNIST-M 数据集作为目标域。MNIST-M 是由文献[7]提出的用于无监督域自适应的 MNIST 数据集的变体。MNIST 的背景为黑色,MNIST-M 的背景由 BSDS500^[34]数据集随机采样组成。

(2)Synthetic Digits to SVHN。将合成数字数据集作为源域,将 SVHN^[35]作为目标域。Synthetic Digits^[7]数据集由大约 50k 合成数字图像组成,图像中的数字具有不同的位置、倾斜方向、背景、颜色和模糊程度等。SVHN 是一个真实世界的街道门牌数字识别数据集。

(3)SVHN to MNIST。将手写数字数据集 SVHN 作为源域,将 MNIST 数据集作为目标域。SVHN 数据集有着明显的变化(比例、背景、模糊、倾斜、对比度、旋转、顺序等),但图像中数字的形状并没有太多变化。而其(如 MNIST)则不同,其在形状、笔迹宽度变化和数字本身的噪声方面存在很多差异,因此这是一个很好但较为困难的实验场景。

(4)MNIST to USPS。将手写数字数据集 MNIST 作为源域,将 USPS 数据集作为目标域。USPS 是美国邮政服务手写数字识别数据集,包含 9298 张 16 * 16 像素的灰度图像,灰度值已被归一化。

4.3 实验结果

实验有针对性地对比了最近比较流行的领域自适应方法,包括 DANN、基于图像重建损失的 DSN、ADDA,以及通过集成学习^[36]进行标签预测并用标记数据再次训练的 ATT。分别测试了不同方法在上述实验数据集上的分类准确率,结果如表 1 所列。

表 1 分类准确率

Table 1 Classification accuracy

(单位:%)

| Source Target | MNIST MNIST-M | Synthetic Digits SVHN | SVHN MNIST | MNIST USPS |
|------------------|------------------|--------------------------|---------------|---------------|
| DANN | 81.5 | 90.3 | 71.1 | — |
| DSN | 83.2 | 91.2 | 82.7 | — |
| ATT | 94.2 | 92.9 | 86.2 | — |
| ADDA | — | 76.0 | — | 89.4 |
| CAFA-DA | 94.7 | 93.2 | 88.3 | 92.6 |

从表 1 可见,CAFA-DA 的准确率在 2 个对比实验上均超过了 ADDA,其中在 Synthetic Digits to SVHN 数据集上的实验结果提高了 17.2%,原因是:Synthetic Digits 为合成数据,而 SVHN 为真实世界数据,两个数据集分布的差异较大;ADDA 方法在训练中只考虑了特征的域间不变性,导致最终效果较差;CAFA-DA 通过提高目标域特征的可鉴别性,使得准确率有了很大的提升。另外,CAFA-DA 在 4 个实验中的准确率均对比模型有不同程度的提升,即使在比较困难的 SVHN to MNIST 实验中,也取得了两个百分点的提升。

结束语 本文提出了一种聚类辅助特征对齐的域适应方法,其核心点有两个:(1)根据目标域数据的聚类结果,用伪标签标记目标域数据并设置权重,在特征对齐过程中用带伪标签的目标域数据辅助训练目标域编码器,使编码器产生的特征更具有域内鉴别性;(2)通过聚类结果训练的分类器与原域适应的分类器进行集成学习,产生高置信度的目标域标签,并用其继续训练模型。用聚类结果训练的分类器不会出现特征对齐过程中的特征损失,而且聚类与域适应学习的两个分类器更具有多样性,有利于结合其优点进行集成学习。实验结果表明:相较于对比方法,本文所提方法在不同域适应实验中的分类准确率均有不同幅度的提升,并且达到了较高水平。本文的后续工作可以在以下几个方面开展:(1)调整第一步聚类和分类器的学习过程以及样本权重,以产生对域适应过程更好的辅助标记;(2)第二步中对目标域编码器的学习方式可以结合更多无监督学习特征表示的方法,使编码器产生更具有鉴别性的特征,以提升特征对齐后的效果。

参考文献

- [1] BEN-DAVID S, BLITZER J, CRAMMER K, et al. A theory of learning from different domains[J]. Machine learning, 2010, 79 (1-2): 151-175.
- [2] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial discriminative domain adaptation[C]// Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 4.
- [3] TZENG E, HOFFMAN J, ZHANG N, et al. Deep Domain Confusion: Maximizing for Domain Invariance [J]. arXiv: 1412. 3474, 2014.
- [4] LONG M, CAO Y, WANG J, et al. Learning transferable features with deep adaptation networks[J]. arXiv: 1502. 02791, 2015.
- [5] LONG M, ZHU H, WANG J, et al. Deep transfer learning with joint adaptation networks[C]// Proceedings of the 34th International Conference on Machine Learning-Volume 70. 2017: 2208-2217.
- [6] 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.
- [7] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. The Journal of Machine Learning Research, 2016, 17(1): 2096-2030.
- [8] BOUSMALIS K, TRIGEORGIS G, SILBERMAN N, et al. Domain separation networks[C]// Advances in Neural Information Processing Systems. IEEE, 2016: 343-351.

- [9] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. IEEE, 2014; 2672-2680.
- [10] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein gans[C]// Advances in Neural Information Processing Systems. IEEE, 2017; 5767-5777.
- [11] SHU R, BUI H H, NARUI H, et al. A DIRT-T Approach to Unsupervised Domain Adaptation[J]. arXiv:1802.08735, 2018.
- [12] TAKERU M, SHIN-ICHI M, SHIN I, et al. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018; 1.
- [13] SAITO K, USHIKU Y, HARADA T. Asymmetric tri-training for unsupervised domain adaptation [J]. arXiv: 1702.08400, 2017.
- [14] CHEN C, XIE W, XU T, et al. Progressive Feature Alignment for Unsupervised Domain Adaptation [J]. arXiv: 1811.08585, 2018.
- [15] KUMAR A, SATTIGERI P, WADHAWAN K, et al. Co-regularized Alignment for Unsupervised Domain Adaptation [C]// Advances in Neural Information Processing Systems. IEEE, 2018; 9367-9378.
- [16] GRANDVALET Y, BENGIO Y. Semi-supervised learning by entropy minimization [C]// Advances in Neural Information Processing Systems. IEEE, 2005; 529-536.
- [17] SHIMODAIRA H. Improving predictive inference under covariate shift by weighting the log-likelihood function[J]. Journal of statistical planning and inference, 2000, 90(2): 227-244.
- [18] MANSOUR Y, MOHRI M, ROSTAMIZADEH A. Domain adaptation: Learning bounds and algorithms[J]. arXiv:0902.3430, 2009.
- [19] CHEN X, DUAN Y, HOUTHOOFT R, et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets[C]// Advances in Neural Information Processing Systems. IEEE, 2016; 2172-2180.
- [20] TZENG E, HOFFMAN J, DARRELL T, et al. Simultaneous deep transfer across domains and tasks[C]// Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2015; 4068-4076.
- [21] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[J]. arXiv:1409.7495, 2014.
- [22] GHIFARY M, KLEIJN W B, ZHANG M, et al. Deep reconstruction-classification networks for unsupervised domain adaptation[C]// European Conference on Computer Vision. Cham: Springer, 2016; 597-613.
- [23] LIU M Y, TUZEL O. Coupled generative adversarial networks [C]// Advances in Neural Information Processing Systems. IEEE, 2016; 469-477.
- [24] BOUSMALIS K, SILBERMAN N, DOHAN D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks[C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017; 7.
- [25] TAIGMAN Y, POLYAK A, WOLF L. Unsupervised cross-domain image generation[J]. arXiv:1611.02200, 2016.
- [26] CHAPELLE O, ZIEN A. Semi-Supervised Classification by Low Density Separation[C]// AISTATS. 2005; 57-64.
- [27] DAI Z, YANG Z, YANG F, et al. Good semi-supervised learning that requires a bad gan[C]// Advances in Neural Information Processing Systems. IEEE, 2017; 6510-6520.
- [28] ZHOU Z H, LI M. Tri-training: Exploiting unlabeled data using three classifiers[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(11): 1529-1541.
- [29] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep Clustering for Unsupervised Learning of Visual Features[M]// Computer Vision-ECCV 2018. Cham: Springer, 2018; 139-156.
- [30] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[C]// Advances in Neural Information Processing Systems. IEEE, 2012; 1097-1105.
- [31] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [32] JOHNSON J, DOUZE M, JÉGOU H. Billion-scale similarity search with gpus[J]. arXiv:1702.08734, 2017.
- [33] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [34] ARBELAEZ P, MAIRE M, FOWLKES C, et al. Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 898-916.
- [35] NETZER Y, WANG T, COATES A, et al. Reading digits in natural images with unsupervised feature learning [C]// NIPS Workshop on Deep Learning and Unsupervised Feature Learning. 2011; 5.
- [36] WANG Y, WANG Q, LV Q F, et al. An Improved K-means Algorithm Based on Initial Clustering Center Optimization and Weighted Between Dimension[J]. Journal of Chongqing University of Technology (Natural Science), 2013, 27(4): 77-80. (in Chinese)
王越, 王泉, 吕奇峰, 等. 基于初始聚类中心优化和维间加权的改进 K-means 算法[J]. 重庆理工大学学报(自然科学), 2013, 27(4): 77-80.