◆人工智能算法与应用◆

# 基于自适应噪声校正的鲁棒域适应学习*

汪云云[1,2**],桂　旭[1,2],郑潍雯[1,2],薛　晖[3]

(1. 南京邮电大学计算机科学与技术学院,江苏南京　210023;2. 南京邮电大学,江苏省大数据安全与智能处理重点实验室,江苏南京　210023;3. 东南大学计算机科学与工程学院,江苏南京　210023)

**摘要**:域适应(Domain Adaptation,DA)学习旨在利用标签丰富的源域来帮助标签稀缺的目标域学习。DA 方法通常假设源域数据已正确标记,然而现实中通常很难收集到大量带有干净标签的源实例,带有噪声源标签的噪声 DA 学习可能会降低目标学习性能。为此,本文提出基于自适应标签噪声校正的鲁棒 DA 学习方法(Robust DA Method through Adaptive Noise Correction,RoDAC)。RoDAC 包含两个学习阶段,即自适应噪声标签检测(Adaptive Noise Label Detection,ANLD)和自适应噪声标签校正(Adaptive Noise Label Correction,ANLC)。在 ANLD 中,使用自适应噪声检测器识别带有噪声标签的源实例,并进一步在 ANLC 中自适应地校正噪声标签,将其重新投入域适应学习中。与基准数据集进行比较,结果表明 RoDAC 方法在源域标签存在噪声的域适应场景中取得了显著的性能提升。该学习策略可集成至许多现有的 DA 方法中,以提升其在噪声标签场景下的学习性能。

**关键词**:域适应　噪声标签检测　噪声标签校正　鲁棒性　元网络

**中图分类号**:TP391　　**文献标识码**:A　　**文章编号**:1005-9164(2022)04-0660-08

**DOI**:10.13656/j.cnki.gxkx.20220919.006

深度神经网络基于大规模标注训练数据,在众多机器感知任务上取得了优秀的性能,然而大规模人工标注意味着高昂的数据标注代价,限制了深度神经网络在数据标注不足任务上的应用。近年来,备受关注的域适应(Domain Adaptation,DA)学习尝试采用标签丰富的源域来帮助标签稀缺的目标域学习[1],在许多任务上性能得到了显著的提升。

近年来,域适应学习方法相继涌现,可大致分为 3 类,分别为基于差异的方法、基于对抗的方法以及基于重构的方法。基于差异的域适应方法旨在缩小域间的分布差异,常用的差异性度量包括最大均值差异(Maximum Mean Discrepancy,MMD)[2]、KL 散度[3]和 Wasserstein 距离[4]等。基于对抗的方法通过对抗学习方式生成域不变特征。例如,DANN[5]和

收稿日期:2022-03-30

*国家自然科学基金面上项目(61876091)和中国博士后科学基金项目(2019M651918)资助。

【作者简介】

汪云云(1986-),女,博士,副教授,主要从事模式识别、机器学习和神经计算研究,E-mail:wangyunyun@njupt.edu.cn。

【**通信作者】

【引用本文】

汪云云,桂旭,郑潍雯,等. 基于自适应噪声校正的鲁棒域适应学习[J]. 广西科学,2022,29(4):660-667.

WANG Y Y,GUI X,ZHENG W W,et al. Robust Domain Adaptive Learning with Adaptive Noise Correction [J]. Guangxi Sciences,2022,29(4):660-667.

DA2NN[6]等采用特征提取器和域鉴别器间的博弈和对抗,而 MCD[7]则利用两个分类器间的对抗学习产生域不变特征。基于重构的方法通过对抗生成跨域样本实现域迁移,如 cycle-GAN[8]、CyCADA[9]等方法。尽管这些方法可有效提升目标域的学习性能,但都假设存在大量正确标记的源域数据,在许多真实学习任务中很难满足。在含噪声的 DA 学习中,噪声源实例可能会导致负迁移现象[1],限制了 DA 方法的实际应用。有文献研究从理论上证明了源域中的标签噪声会给 DA 学习带来严重的负迁移问题[10]。也有学者研究了更通用的噪声 DA 场景,包括标签噪声、特征噪声以及混合噪声[11]。

DA 方法通常假设存在一个已正确标记的源域,然而在现实任务中,精确的人工标注耗时耗力,通常很难收集大量带有干净标签的源实例。虽然很容易从网络和社交媒体中采集数据,但是此类数据集可能存在标签噪声。带有噪声源标签的噪声 DA 学习可能会损害目标域的学习性能,但针对此问题的研究仍然有限。

带噪声标签的鲁棒学习目前已获得广泛研究。噪声标签学习方法可大致分为 3 类:基于无偏损失或风险最小化[12-15]、基于自助(Bootstrapping)损失[16]以及噪声样本重采样方法[17-20]。基于无偏损失或风险最小化方法旨在定义可有效抑制噪声数据的损失函数。基于自助损失方法通过在训练过程中不断修正数据标签来减轻噪声标签实例对模型学习的影响。噪声样本重采样方法则通过设计从训练损失到样本权重的映射函数,根据损失计算样本权重,最小化分类器更新的加权实例损失。此外,根据是否使用转移矩阵对标签噪声进行显式建模,将带标签噪声的鲁棒学习方法分为基于转移矩阵法和无转移矩阵法,本研究属于第二类,不同的是本研究场景更具挑战性,旨在将含有噪声标签的源域知识迁移至目标域,以提升目标域的学习性能。

针对源域存在噪声标签的学习场景,本文提出基于自适应噪声校正的鲁棒域适应学习方法(Robust DA Method through Adaptive Noise Correction,RoDAC)。首先,通过自适应噪声检测(Adaptive Noisy Label Detection,ANLD)识别噪声源实例、加权实例损失以及减少噪声实例对分类学习的贡献,同时划分数据集以进一步减少其对域适应学习的影响;接着,通过自适应噪声标签校正(Adaptive Noisy Label Correction,ANLC)修正噪声实例,使噪声实例获得更为准确的标签,并重新投入学习。

## 1 自适应噪声校正的鲁棒域适应学习方法

在源域标签存在噪声的域适应学习场景中,主要挑战在于:①如何减少噪声标签对分类学习的影响;②如何减少噪声标签对域间分布对齐的影响。因此,本文提出了一个系统的解决方案,主要包括两个学习阶段:自适应噪声标签检测(ANLD)和自适应噪声标签校正(ANLC)。

### 1.1 符号定义和网络结构

为方便描述,首先给出符号定义,噪声标签源域数据和无标签目标域数据分别表示为 $(X^s,Y^s)=\{(x_i^s,y_i^s)\}_{i=1}^{n_s}$ 和 $X^t=\{(x_i^t)\}_{i=1}^{n_t}$,其中,$n_s$ 和 $n_t$ 分别表示源域和目标域样本数;s 和 t 分别代表源域和目标域。

进一步地,采用数学符号来表述本文的研究目标。将成对标签上的损失函数定义为 $L:Y\times Y\to\mathbb{R}$,多分类任务的决策函数 $f:X\to\mathbb{R}^K$,因而有假设函数 $h_f=x\to\underset{y\in Y}{\operatorname{argmax}}f(x,y)h_f=x\to\underset{y\in Y}{\operatorname{argmax}}f(x,y)$。对于 $X\times Y$ 上的任意分布 $D$,定义假设函数 $h$ 的经验误差为 $\mathrm{err}_D\triangleq E_{(x,y)\sim D}L(h(x),y)$。本文的研究目标是采用源域的有标签噪声样本和目标域的无标签样本习得一个决策函数 $f$,使得其在目标域上的经验误差最小。图 1 展示了标签噪声 DA 学习过程。

RoDAC 利用干净实例的 small-loss[21]基于元网络自适应地区分干净和噪声实例,同时,基于数据的聚簇特性,利用原型分类器对检测出的噪声实例进行自适应校正,并重新投入域适应学习中。RoDAC 的网络结构如图 2 所示。其中,F 为特征提取器,G 为多层感知分类器,源域和目标域共享 F 和 G。M 为基于元网络的自适应噪声检测器,用于自适应地检测噪声实例。P 为原型分类器,用于自适应地校正噪声实例。通过局部最大均值差异(Local Maximum Mean Discrepancy,LMMD)[22]计算源域与目标域间的条件分布差异,并缩小域间的分布距离,实现源域知识迁移。

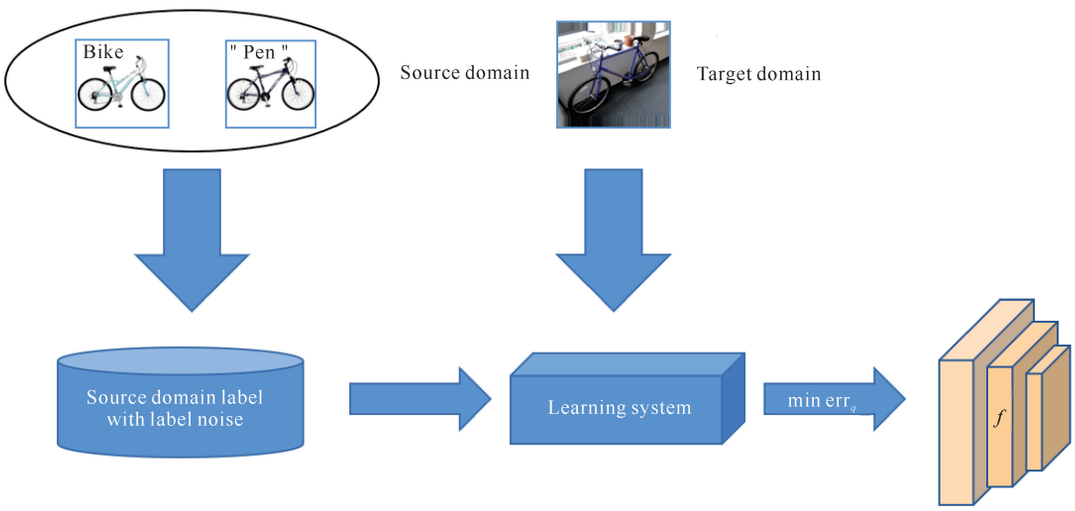


图 1　标签噪声 DA 学习过程

Fig. 1　Label noise DA learning process

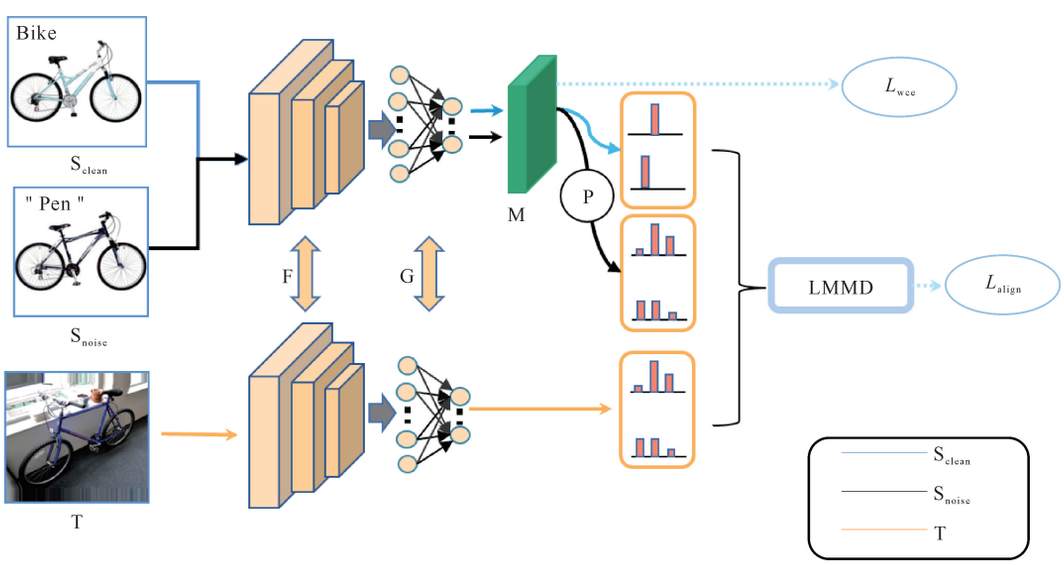


图 2　RoDAC 网络结构图

Fig. 2　Network structure diagram of RoDAC

## 1.2　自适应噪声标签检测

由于源域存在噪声标签,源域学习可能会错误地拟合噪声实例,并将错误知识迁移至目标域。鉴于此,RoDAC 通过基于元网络的自适应噪声标签检测器来识别噪声实例。首先,利用元网络自适应地对源域实例进行加权,权值大小用来描述其属于干净样本的可能性,从而降低噪声标签对分类学习的影响。接着,基于阈值法将源域数据集划分为干净实例集 $X^s_{clean}$ 和噪声实例集 $X^s_{noise}$,并在 ANLC 中校正噪声标签,用于提升域间对齐和源域迁移知识。

### 1.2.1　实例损失加权

若仅考虑源域上的监督分类任务,分类学习损失可表示为

$$\min_f \frac{1}{n_s}\sum_{i=1}^{n_s} J(f(x_i^s;\theta),y_i^s)\ , \tag{1}$$

式中,$J$ 是交叉熵损失函数,$f$ 是由 $\theta$ 参数化的假设函数。为减少噪声标签对分类学习的影响,RoDAC 自适应地对源域样本加权,即基于损失对样本重采样,获得加权的源域分类损失 $L_{wce}$ 描述如下:

$$\min_f \frac{1}{n_s}\sum_{i=1}^{n_s} w(x_i^s,\alpha)J(f(x_i^s;\theta),y_i^s)\ , \tag{2}$$

式中,$w$ 是由 $\alpha \in B^d$ 参数化的权重函数,将实例映射至实例权重。

RoDAC 的学习目标之一是找到一个由 $\alpha$ 参数化的权重函数 $w$,使高置信度的噪声实例获得较低的权重,以减少其对分类学习的影响。从 $\theta$ 优化的角度来看,$w(x_i^s,\alpha)$ 为超参数,因此使用源域中一小组正确标记的验证集 $X_{\text{vali}}^s=\{(x_i^{\text{vali}},y_i^{\text{vali}})\}_{i=1}^V$ 用于超参数优化,其中 $V$ 为验证集样本个数。$w(x_i^s,\alpha)$ 的优化依赖于其在该验证集上的学习性能,即

$$\hat{\alpha}=\underset{\alpha\in B^d}{\operatorname{argmin}}\sum_{i=1}^{V}J_i^{\text{vali}}(f(x_i^{\text{vali}},\hat{\theta}(\alpha),y_i^{\text{vali}})), \tag{3}$$

式中,$J_i^{\text{vali}}$ 是在验证集实例上的交叉熵损失。

$\theta$ 和 $\alpha$ 的求解是一个双线性问题。在域适应学习中,$\theta$ 的更新包括两部分:源域分类损失和域分布对齐损失。

#### 1.2.2 源域数据划分

基于实例损失的元网络输出,可反映源域噪声实例的置信度。基于阈值法,可将噪声实例从干净实例中分离出来,对其自适应地校正并重新投入域适应学习中。设立权重函数 $w$ 的阈值 thres,若 $w(x_i^s)>\text{thres}$,将 $x_i^s$ 标记为干净实例;反之,标记为噪声实例,即

$$x_i^s\in\begin{cases}X_{\text{clean}}^s,\text{若 } w(x_i^s)>\text{thres}\\ X_{\text{noise}}^s,\text{否则}\end{cases}, \tag{4}$$

式中,$X_{\text{clean}}^s$ 和 $X_{\text{noise}}^s$ 分别表示划分后的干净和噪声源数据集。

### 1.3 自适应噪声标签校正

在 ANLD 中,RoDAC 利用实例加权来减少噪声实例对分类学习的影响,同时将源域数据集划分为干净和噪声实例集。在 ANLC 中,对噪声实例集进行标签校正并重新投入域适应学习,而非简单丢弃,以充分利用源域知识。

通过可视化研究发现,即使将干净实例和噪声实例一同训练,干净实例仍会聚簇在所属类别原型的周围。因此,RoDAC 选择原型分类器[22]代替传统的多层感知器进行噪声校正,即使用干净实例集中的样本计算原型中心,并基于该原型中心校正噪声实例集中的样本。

对于每个源域噪声实例 $x_i^s$,利用原型分类器校正其类别标签 $\hat{y}_i^s$,其中隶属于类别 $c$ 的概率预测计算如下:

$$\hat{y}_{i,c}^s=\frac{e^{\Phi_1(x_i^s,\mu_c^s)}}{\sum_{j=1}^{K}e^{\Phi_1(x_i^s,\mu_j^s)}}, \tag{5}$$

式中,$K$ 为类别总数,$\Phi_1(x_i^s,\mu_c^s)$ 用于衡量源域实例与某个特定类原型间的相似性,$\mu_c^s$ 是源域中类别 $c$ 的原型中心,通过干净实例集中样本计算获得。

接下来,对标签校正后的数据进行域间条件分布对齐。对于干净源实例,直接使用真实类标签;而对于噪声源实例,则使用原型分类器的预测结果作为其软标签,即

$$\tilde{y}_i^s=\begin{cases}y_i^s,x_i^s\in X_{\text{clean}}^s\\ \hat{y}_i^s,x_i^s\in X_{\text{noise}}^s\end{cases}。 \tag{6}$$

RoDAC 将校正后的源域实例重新投入域适应学习并采用 LMMD 度量[23]进行条件分布对齐,LMMD 可刻画如下:

$$L_{\text{align}}=\frac{1}{K}\sum_{k=1}^{K}\Big\|\sum_{x_i^s\in X^s}z_{i,k}^s\Phi_2(x_i^s)-\sum_{x_j^t\in X^t}z_{j,k}^t\Phi_2(x_j^t)\Big\|^2, \tag{7}$$

式中,$\Phi_2$ 是从特征空间到再生核希尔伯特空间的映射,与原型网络的度量函数相区分。$z_{i,k}^s$ 和 $z_{j,k}^t$ 分别表示源域和目标域实例在条件域适应中对 $k$ 类的贡献权重,其通用的计算方式如下:

$$z_{i,k}=\frac{\tilde{y}_{i,k}}{\sum_{(x_j,y_j)\in D}\tilde{y}_{j,k}}, \tag{8}$$

式中,$\tilde{y}_{i,k}$ 表示基于校正后源域实例的类标签 $\tilde{y}_i^s$,或目标域实例的预测伪标签 $\tilde{y}_i^t$,实例 $x_i$ 属于第 $k$ 类的概率。

### 1.4 总体学习目标和优化

上述内容已详细阐明了 RoDAC 中的噪声标签检测和噪声标签校正过程,RoDAC 的总目标函数刻画如下:

$$\begin{aligned}&\hat{\alpha}=\underset{\alpha\in B^d}{\operatorname{argmin}}\sum_{i=1}^{V}J_i^{\text{vali}}f(x_i^{\text{vali}},\hat{\theta}(\alpha),y_i^{\text{vali}}))\\ &\hat{\theta}=\underset{\theta\in\Theta}{\operatorname{argmin}}L_{\text{wce}}+L_{\text{align}}。\end{aligned} \tag{9}$$

此类双线性问题的求解需要两层嵌套的优化循环,在几何上降低了求解效率。为进一步加速优化,借鉴 Meta-Weight-Net[24]和 DS3L[25]的在线学习策略,分别通过单个优化循环更新 $\theta$ 和 $\alpha$。为进一步阐述优化过程,将内循环的更新标记为 $L^{\text{inner}}(\theta,\alpha)$,外循环 $\alpha$ 的更新标记为 $L^{\text{outer}}(\theta)$。给定权重函数 $w$ 的参数 $\alpha_t$,$\theta_{t+1}$ 的更新由以下梯度等式得到:

$$\theta_{t+1}=\theta_t-\eta_\theta\nabla_\theta L^{\text{inner}}(\theta_t,\alpha_t), \tag{10}$$

得到参数 $\theta_{t+1}$($\hat{\theta}$ 的近似值)后,通过以下方式更新外部目标函数的参数 $\alpha$:

$$\alpha_{t+1}=\alpha_t-\eta_\alpha\nabla_\alpha L^{\text{outer}}(\theta_{t+1})。\tag{11}$$

具体学习框架如算法 1 所示:

**算法 1** RoDAC 的学习框架

输入:噪声标签源域数据 $(X^s,Y^s)=\{(x_i^s,y_i^s)\}_{i=1}^{n_s}$,无标签目标域数据 $X^t=\{(x_i^t)\}_{i=1}^{n_t}$,迭代次数 $T$,域值 thres;

输出:权重函数参数 $\alpha$ 和模型参数 $\theta$;

①从含噪声标签的源域采样 $n$ 个实例 $\{(x_i^s,y_i^s)\}_{i=1}^{n}$;

②从目标域采样 $m$ 个实例 $\{(x_i^t,y_i^t)\}_{i=1}^{m}$;

③从源域验证集采样 $v$ 个实例 $\{(x_i^{\text{vali}},y_i^{\text{vali}})\}_{i=1}^{v}$;

④基于权重函数 $w(x_i^s,\alpha)$,通过阈值法划分源域数据;

⑤由公式(2)和(7)分别计算加权监督损失 $L_{\text{wce}}$ 和 $L_{\text{align}}$;

⑥计算内循环损失 $L^{\text{inner}}(\theta,\alpha)$;

⑦更新参数 $\theta_{t+1}\leftarrow\theta_t-\eta_\theta\nabla_\theta L^{\text{inner}}(\theta_t,\alpha_t)$;

⑧由公式(3)计算外循环损失 $L^{\text{outer}}(\theta)$;

⑨更新参数 $\alpha_{t+1}\leftarrow\alpha_t-\eta_\alpha\nabla_\alpha L^{\text{outer}}(\theta_{t+1})$;

⑩end for

## 2 结果与分析

在本节中,分别在图像识别和数字分类任务上将 RoDAC 方法与 SOTA 域适应学习方法进行性能比较。

### 2.1 实验设置

数据集:Office-31 是无监督域适应的基准数据集,用于图像识别任务,包含 3 个域,共计 31 个类别,4 110 张图像。其中 Amazon (A)是从 https://www.amazon.com 下载的图像,Webcam (W)和 DSLR (D)分别包含由不同摄影设置的网络相机和数码单反相机拍摄的图像。

Digits 用于数字分类任务,其中 usps 和 mnist 都是灰色手写数字数据集,包括 0-9 类 10 个类别。usps 包含 7 291 张训练图像和 2 007 张测试图像,大小为 16×16。mnist 包含 60 000 张训练图像和 10 000 张测试图像,大小为 28×28。svhn 是彩色手写数字数据集,包含 4 578 张训练图像和 1 627 张测试图像,大小为 32×32。

实验设置:为模拟噪声标签的域适应场景,手动创建噪声标签数据集。遵循无监督域适应(Unsupervised Domain Adaptation,UDA)的实验设置,训练中仅包含噪声标签的源域数据和无标签的目标域数据。所有对比方法都在 pytorch 深度学习框架中复现。为了公平比较,所有方法都设置相同的超参数、预处理和特征提取网络,然后对每个学习任务详细说明实验设置。

对于图像识别任务,将所有图像缩放至 256×256,同时通过随机裁剪进行数据增强,最终将图片尺寸统一为 224×224。使用 resnet-50 作为特征提取器,256 输出单元的瓶颈层和 31 类分类器层作为多层感知分类器。

对于手写数字分类任务,所有图像被缩放为相同尺寸的三通道图片,并归一化每张图像的像素。使用 AlexNet[27] 作为特征提取器,分类器包括一个 2 048 输出单元的瓶颈层和一个 10 类输出的分类器层,元网络是隐藏层(仅含 100 个节点)的 MLP 网络。所有参数使用随机梯度下降(SGD)更新,并使用相同的动态学习率 $lr=0.001\times1/(1+10\times\frac{\text{epoches}-1}{\text{iteration}})^{0.75}$ 和动量值 0.9。

### 2.2 实验结果

在源域标签噪声率为 40%的场景中,将本文方法与经典无噪声的 UDA 方法进行对比,在 Office-31 和 Digits 数据集上的实验结果如表 1、表 2 所示。结果表明,RoDAC 在源域存在噪声标签的场景中有着明显的性能提升,验证了其有效性。

**表 1 Office-31 数据集上 40%噪声比的准确率(%)**

**Table 1 Accuracy of 40% noise ratio on Office-31 dataset (%)**

| 方法 Method | A→W | W→A | A→D | D→A | D→W | W→D | AVG |
|---|---|---|---|---|---|---|---|
| ResNet[26] | 47.2 | 33.0 | 47.1 | 31.0 | 68.0 | 58.8 | 47.5 |
| DAN[2] | 63.2 | 39.0 | 58.0 | 36.7 | 71.6 | 61.6 | 55.0 |
| DANN[5] | 61.2 | 46.2 | 57.4 | 42.4 | 74.5 | 62.0 | 57.3 |
| ADDA[6] | 61.5 | 49.2 | 61.2 | 45.5 | 74.7 | 65.1 | 59.5 |
| RoDAC | 71.8 | 60.7 | 67.3 | 55.0 | 81.8 | 73.8 | 68.4 |

**表 2　Digits 数据集上 40%噪声比的准确率(%)**

**Table 2　Accuracy of 40% noise ratio on Digits dataset (%)**

| 方法 Method | mnist→usps | usps→mnist | svhn→mnist |
|---|---|---|---|
| AlexNet[27] | 48.53 | 22.00 | 24.26 |
| DAN[2] | 60.78 | 45.62 | 53.87 |
| DANN[5] | 58.56 | 43.38 | 49.64 |
| ADDA[6] | 58.80 | 45.79 | 50.70 |
| RoDAC | 71.58 | 50.44 | 57.87 |

RoDAC 包括两个学习阶段,分别为噪声标签检测 $L_{\mathrm{NLD}}$ 和噪声标签校正 $L_{\mathrm{NLC}}$。为进一步验证方法的有效性,在 40%噪声比的 Digits 数据集上进行消融实验,实验结果如表 3 所示。通过自适应噪声检测器检测噪声实例,减少其对分类和域适应学习的负面影响,可有效提升目标域的学习性能;与此同时,采用自适应噪声标签校正器对噪声标签进行校正,从而将噪声实例重新投入训练而并非直接丢弃,可充分利用源域知识,进一步提升学习性能。

**表 3　RoDAC 的消融实验**

**Table 3　Ablation experiment of RoDAC**

| 方法 Method | mnist→usps | usps→mnist | svhn→mnist |
|---|---|---|---|
| $L_{\mathrm{wce}}$ | 48.53 | 22.00 | 24.26 |
| $L_{\mathrm{wce}}+L_{\mathrm{NLD}}$ | 69.89 | 50.01 | 55.93 |
| $L_{\mathrm{wce}}+L_{\mathrm{NLD}}+L_{\mathrm{NLC}}$ | 71.58 | 50.44 | 57.87 |

在 Digits 数据集分类任务上对 3 种不同噪声比下的基准方法和 RoDAC 方法的性能进行比较,结果见图3。在域适应学习中,目标域学习精度受噪声比的影响很大,随着噪声比的增大,性能锐降。RoDAC 的学习性能也随着噪声比的增加而下降,但相较于基准方法下降程度较为平缓,特别是在噪声比为 60%时,RoDAC 的学习性能优于噪声比为 40%时基准方法的学习性能,进一步验证了 RoDAC 的有效性。

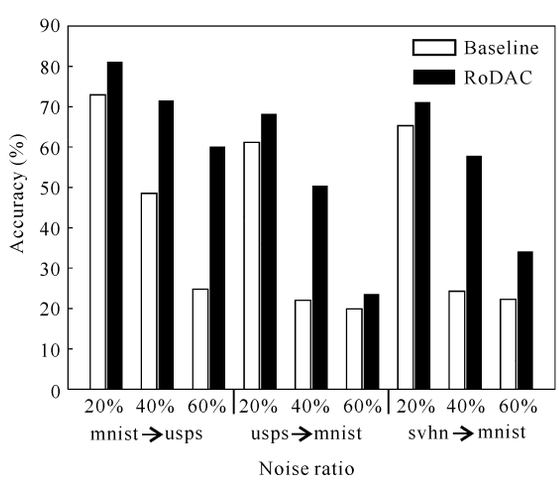


图 3　不同噪声比下的性能比较

Fig. 3　Performance comparison under different noise ratios

## 3　结论

为解决源域数据中噪声标签的域适应问题,本文提出两阶段的鲁棒学习策略 RoDAC,包含自适应噪声标签检测和自适应噪声标签校正。首先,采用基于元网络的自适应噪声检测器识别噪声源实例,以缓和其对训练过程的影响;其次,采用基于原型分类器的自适应噪声标签校正方法,对检测出的噪声实例进行自适应校正,并重新投入学习。实验结果表明,与经典的 UDA 方法相比,在源域数据存在噪声标签的域适应场景中,RoDAC 的学习性能有显著提升。

**参考文献**

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

[2] LONG M S, CAO Y, WANG J M, et al. Learning transferable features with deep adaptation networks [C]// Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR, 2015, 37: 97-105.

[3] KULLBACK S, LEIBLER A R. On information and sufficiency [J]. The Annals of Mathematical Statistics, 1951, 22(1): 79-86.

[4] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks [C]// Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: PMLR, 2017, 70: 214-223.

[5] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks [J]. Journal of Machine Learning Research, 2016, 17(1): 2096-2030.

[6] WANG Y Y, GU J M, WANG C, et al. Discrimination-aware domain adversarial neural network [J]. Journal of Computer Science and Technology, 2020, 35(2): 259-267.

[7] SAITO K, WATANABE K, USHIKU Y, et al. Maximum classifier discrepancy for unsupervised domain adaptation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 3723-3732.

[8] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-

image translation using cycle-consistent adversarial networks [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2223-2232. DOI:10.1109/ICCV.2017.244.

[9] HOFFMAN J, TZENG E, PARK T, et al. CYCADA: Cycle-consistent adversarial domain adaptation [C]// Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018, 80:1989-1998.

[10] YU X Y, LIU T L, GONG M M, et al. Label-noise robust domain adaptation [C]//Proceedings of the 37th International Conference on Machine Learning. Online: PMLR, 2020, 119:10913-10924.

[11] SHU Y, CAO Z J, LONG M S, et al. Transferable curriculum for weakly-supervised domain adaptation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii, Honolulu, USA: AAAI Press, 2019, 33(1):4951-4958.

[12] NATARAJAN N, DHILLON I S, RAVIKUMAR P, et al. Learning with noisy labels [C]//Annual conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA: [s. n.], 2013, 26(2):1198-1206.

[13] XU Y L, CAO P, KONG Y Q, et al. $L_{DMI}$: A novel information-theoretic loss function for training deep nets Robust to label noise [C]//Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver, Canada: [s. n.], 2019: 6222-6233.

[14] SUKHBAATAR S, BRUNA J, PALURI M, et al. Training convolutional networks with noisy labels [C]//International Conference on Learning Representations 2015. San Diego, CA, USA: ICLR, 2015:1-11.

[15] PATRINI G, ROZZA A, MENON A K, et al. Making deep neural networks robust to label noise: A loss correction approach [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017:1944-1952.

[16] ARAZO E, ORTEGO D, ALBERT P, et al. Unsupervised label noise modeling and loss correction [C]// Proceedings of the 36th International Conference on Machine Learning. Long Beach, California, USA: PMLR, 2019, 97:312-321.

[17] JIANG L, ZHOU Z Y, LEUNG T, et al. MentorNet: Learning data-driven curriculum for very deep neural networks on corrupted labels [C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018, 80:2304-2313.

[18] CHEN P F, LIAO B B, CHEN G Y, et al. Understanding and utilizing deep neural networks trained with noisy labels [C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach, California, USA: PMLR, 2019, 97:1062-1070.

[19] THULASIDASAN S, BHATTACHARYA T, BILMES J, et al. Combating label noise in deep learning using abstention [C]//Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, USA: PMLR, 2019, 97:6234-6243.

[20] SONG H, KIM M, PARK D, et al. Robust learning by self-transition for handling noisy labels [C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Virtual Event, Singapore: ACM, 2021:1490-1500.

[21] WANG Y X, KUCUKELBIR A, BLEI D M. Robust probabilistic modeling with Bayesian data reweighting [C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: PMLR, 2017, 70:3646-3655.

[22] ZHU Y C, ZHUANG F Z, WANG J D, et al. Deep subdomain adaptation network for image classification [J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(4):1713-1722.

[23] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: [s. n.], 2017:1-13.

[24] SHU J, XIE Q, YI L X, et al. Meta-weight-net: Learning an explicit mapping for sample weighting [C]// Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver, Canada: [s. n.], 2019:1-23.

[25] GUO L Z, ZHANG Z Y, JIANG Y, et al. Safe deep semi-supervised learning for unseen-class unlabeled data [C]//Proceedings of the 37th International Conference on Machine Learning. Online: JMLR, 2020:3897-3906.

[26] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016:770-778.

[27] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6):84-90.

# Robust Domain Adaptive Learning with Adaptive Noise Correction

WANG Yunyun[1,2],GUI Xu[1,2],ZHENG Weiwen[1,2],XUE Hui[3]

(1. School of Computer Science and Technology, Nanjing University of Posts and Telecommunication, Nanjing, Jiangsu, 210023, China; 2. Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, 210023, China; 3. School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu, 210023, China)

**Abstract**: Domain Adaptation (DA) learning aims to use lable-rich source domains to help the learning of label-scarce target domain. The DA method usually assumes that the source domain data has been correctly labeled. However, in reality, it is usually difficult to collect a large number of source instances with clean labels. Noise DA learning with noise source labels may reduce the target learning performance. Therefore, this article proposes a Robust DA Method through Adaptive Noise Correction (RoDAC). RoDAC consists of two learning stages, Adaptive Noise Label Detection (ANLD) and Adaptive Noise Label Correction (ANLC). In ANLD, an adaptive noise detector is used to identify the source instance with noise labels, and the noise labels are further adaptively corrected in ANLC and reinvested in domain adaptation learning. Compared with the benchmark data set, the results show that the RoDAC method achieves significant performance improvement in the domain adaptation scenario where the source domain label has noise. This learning strategy can be integrated into many existing DA methods to improve its learning performance in noisy label scenarios.

**Key words**: domain adaptation; noise label detection; noise label correction; robustness; meta network

责任编辑:唐淑芬