Semi-supervised Domain Adaptation via Minimax Entropy

Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell and Kate Saenko

摘要

当代的域自适应方法在无需任何目标监督的情况下对齐源域和目标域的特征分布是非常有效的。然而我们发现，即使在目标域中有几个标记的示例。为了解决这种半监督域自适应（SSDA）设置，我们提出了一种新的 Minimax 熵（MME）方法，对自适应小样本模型。我们的基本模型包含一个功能编码网络，然后是分类层计算特征与估计原型的相似度。实现了适应通过相对于分类器交替地最大化未标记目标数据的条件熵和相对于特征编码器最小化它。在 SSDA 方法上，我们通过经验证明了我们的方法优于许多基线，包括传统的要素对齐和小样本学习模型。

关键词：领域自适应；半监督学习

1 引言

在用神经网络模型对数据进行训练和预测时，经常会出现训练好的模型用于实际任务时发生预测效果较差的问题。这些因为采集数据时的环境差异或者时间差异，导致训练数据和预测数据的分布特性不能保证完全一致，所以不能对实时获取数据做出准确的预测。领域自适应学习中源域和目标域的概念就是尝试减小源域和目标域的差距，将源域中学习的知识迁移到目标域中，利用这种学习思想可以让我们的神经网络模型具有更强的泛化能力，在数据上存在差异时也能做出较为准确的预测。半监督式的学习方式能够减少数据标记的开销。

2 相关工作

2.1 领域自适应

领域自适应（DA）的主要挑战是域之间特征分布的差距，这会降低源分类器的性能。最近的工作集中在无监督域自适应（UDA），特别是特征分布对齐。基本方法测量源和目标中特征分布之间的距离，然后训练模型以最小化该距离。许多 UDA 方法使用域分类器来测量距离[1][2]。训练域分类器以区分输入特征是来自源还是目标，而训练特征提取器以诱导域分类器以匹配特征分布。UDA 已应用于各种应用，如图像分类[3]、语义分割[4]和对象检测[5][6]。一些方法最小化任务特定决策边界对目标示例的分歧[7][8]，以将目标特征推离决策边界。在这方面，它们增加了目标特征的类间方差。

2.2 半监督学习（SSL）

生成式[9][10]、模型集成[11]和对抗性方法[12]提高了半监督学习的性能，但没有解决领域转移问题。条件熵最小化（CEM）是 SSL 中广泛使用的方法[13]。然而，我们发现，当源域和目标域之间存在较大的域间隙时，CEM 无法提高性能。MME 可以被视为熵最小化的变体，它克服了 CEM 在域自适应方面的局限性。

2.3 小样本学习（FSL）

小样本学习[14][15][16]旨在通过一些标记的例子和标记的“基础”分类来学习新颖的分类。SSDA 和 FSL 做出了不同的假设：FSL 不使用未标记的示例，旨在获取新类的知识，而 SSDA 旨在适应新领域

中的相同类。然而，这两项任务都旨在从一个新的领域或新的类中提取一些带有标签的特征。

3 本文方法

3.1 本文方法概述

文章提出了一种新的 SSDA 方法[17]，该方法克服了以前方法的局限性，并显著提高了每个类只有几个标签的新域上深度分类器的准确性。这种方法称为最小最大熵（MME）是基于优化未标记数据的条件熵的最小最大损失以及任务损失；这在学习任务的辨别特征的同时减少了分布差距。关键思想是最小化类原型和相邻未标记目标样本之间的距离，从而提取区分特征。问题是如何在没有许多标记的目标示例的情况下估计域不变原型。原型由源域主导，因为绝大多数标记示例来自源。为了估计域不变原型，本方法将权重向量移向目标特征分布。目标实例上的熵表示估计的原型和目标特征之间的相似性。具有高熵的均匀输出分布表明，示例与所有原型权重向量相似。因此，通过在第一个对抗步骤中最大化未标记目标示例的熵，将权重向量移向目标。第二，更新特征提取器以最小化未标记示例的熵，使它们更好地围绕原型聚集。该过程被公式化为权重向量和特征提取器之间的最小-最大博弈，并应用于未标记的目标示例。本文的方法如图 1所示。

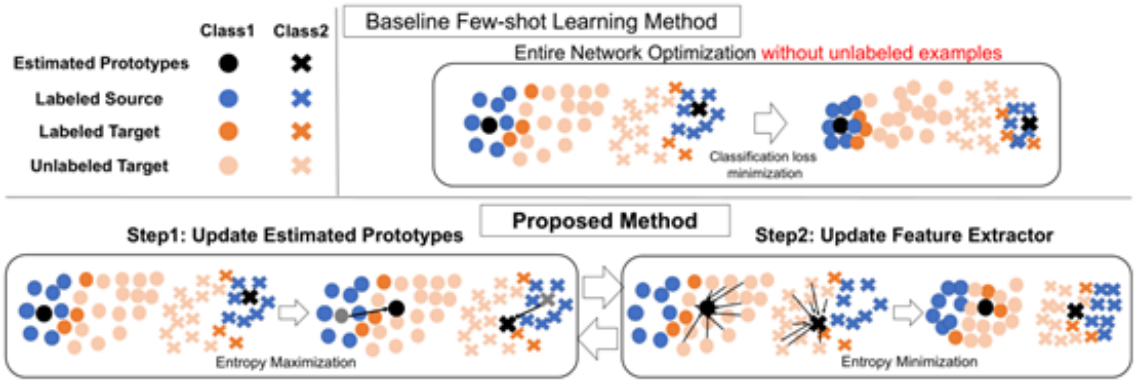

图 1: 方法示意图

3.2 模型架构

本文的基础模型由特征提取器 F 和分类器 C 组成。对于特征提取器，使用深度卷积神经网络，并对网络的输出执行 L2 归一化。然后，归一化特征向量被用作 C 的输入，C 由权重向量 W ＝ [w1，w2，…，wK] 组成，其中 K 表示类的数量和 T 温度乘法。为了正确地对示例进行分类，权重向量的方向必须代表对应类的归一化特征。在这方面，权重向量可以被视为每个类的估计原型。本文模型的架构如图 2所示。

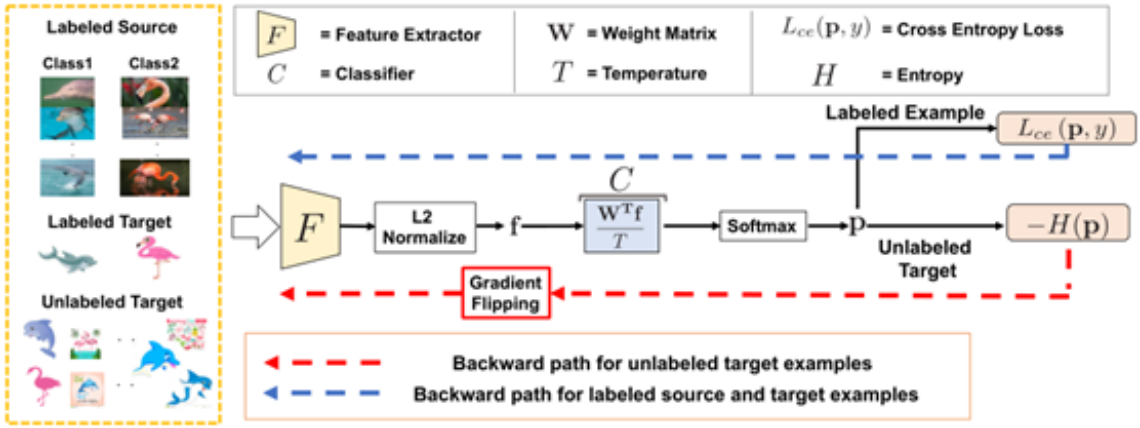

图 2: 网络模型结构

3.3 损失函数定义

在半监督领域自适应中，使用源图像和源域中的相应标签 Ds=（x，y）。在目标域中，还使用了有限数量的标记目标图像 Dt=（x，y），以及未标记目标图像 Du=（x）。使用带标签的示例，并训练整个网络以最小化分类损失。最小化源示例和小部分目标示例的交叉熵损失（公式 1）确保带标签示例的熵非常小。

$$L = E_{(x,y)\in D_s,D_t} L_{ce}(P(x), y) \tag{1}$$

其中，L（公式 1）为有标注的 Ds、Dt 的监督损失，H（公式 2）为 Du 的熵值。对抗的具体实现由梯度反转层 GRL 实现，即是 Du 的熵值 H 在 C 和 F 之间的梯度反转。

$$H = -E_{(x,y)\in D_u} \sum_{i=0}^{K} p(y=i|x) logp(y=i|x) \tag{2}$$

每个类别都存在一个域不变原型，优化分类头 C 最大化 Du 的熵值，将类别的边界划分的更加的宽松，所以导致类别原型从源域数据集中的位置中扩散更接近目标域中该类别的特征，这样原型才为域不变原型。熵最大化可以防止表征的过拟合，可以认为是一种选择可以防止过拟合的原型的过程。优化特征提取器 F 最小化 Du 的熵值，即使固定了类别边界，去优化特征的空间分布，让每个目标域特征更加接近某一个特定的类别。所以对抗训练优化函数为

我们导出了我们提出的学习方法的最小最大目标（公式 3、4）。总之，我们的最大熵过程可以被视为测量域之间的散度，而我们的熵最小化过程可以被看作是最小化散度。

$$\theta_F^- = argmin L_{\theta_F} + \lambda H \tag{3}$$

$$\theta_C^- = argmin L_{\theta_C} - \lambda H \tag{4}$$

4 复现细节

4.1 与已有开源代码对比

复现了开源代码的网络模型，其中包括使用 AlexNet 网络模型构建的特征提取器 F 和带有梯度反转功能的分类器 C，以及计算带标签数据损失的 L 公式和计算目标域未标签数据熵值的 H 公式。

重新使用了一种文本和图像组合的数据集加载方式，可以更加方便的对数据集进行细致的处理，并且重写了分类器 C 中的梯度反转函数，使梯度反转函数更适合当前实验运行环境。

在特征提取器和分类器之间新加入了文章中未提及的分布对齐功能，在原来单纯的减小源域和目标域之间整体差异的基础之上，还减小了源域和目标域之间数据分布的差异，在实验中提高了训练过程中损失函数的收敛速度，也提高了预测的准确度。

4.2 实验环境搭建

数据集：Office[3]包含 3 个域（Amazon、Webcam、DSLR），共 31 个类。Webcam 和 DSLR 是小领域，有些类没有很多示例，而 Amazon 有很多示例。为了用足够多的例子对域进行评估，其中我们将 Amazon 设置为目标域，Webcam 设置为源域。

对每个类别随机选择三个标记的示例作为标记的训练目标示例。选择其他三个标记示例作为目标

域的验证集。验证示例用于提前停止、选择超参数 λ 和训练计划。其他目标示例用于无标签的训练，其标签仅用于评估分类精度。源域的所有示例均用于训练。

所有实验都在 Pytorch 中进行。使用 AlexNet 网络模型在 Office 数据集上预训练。删除了这些网络的最后一个线性层以构建 F，并添加了具有随机初始化的权重矩阵 W 的 K 路线性分类层 C。在所有设置中，温度 T 的值设置为 0.05。每次迭代，我们都准备了两个小批量，一个由标记的实例组成，另一个由未标记的目标实例组成。一半的标记示例来自源，一半来自标记目标。使用两个小批次，计算了公式 3、4 中的目标。为了公式 3、4 中的对抗性学习，使用梯度反转层来翻转熵损失的梯度。在反向传播期间，梯度的符号在 C 和 F 之间翻转，采用动量为 0.9 的 SGD。

4.3 创新点

加入了分布对齐操作，可以在未知目标域数据分布的情况下，以源域数据分布为基准预测目标域数据的分布情况，在已知目标域数据分布的情况下，减小源域数据和目标域数据的分布差异，帮助网络模型能够更好地把源域数据所学知识迁移到目标域当中，提高对目标域数据预测的准确度。

5 实验结果分析

本部分对实验所得结果进行分析，详细对实验内容进行说明，实验结果进行描述并分析。将所有带标记的源域数据和每个类别中 3 个带标记的目标域数据一起在网络模型中进行训练，计算出交叉熵损失 L，未标记目标域数据先在网络模型中的特征提取器中训练得到的结果与带标签数据在特征提取器中获得结果进行分布对齐操作，接着将分布对齐后的未标记目标域数据在分类器中进行进行梯度反转操作，最终进行分类预测获得最后预测结果。训练总体交叉熵损失、未标记目标域损失和预测成功率如下图 3、4所示

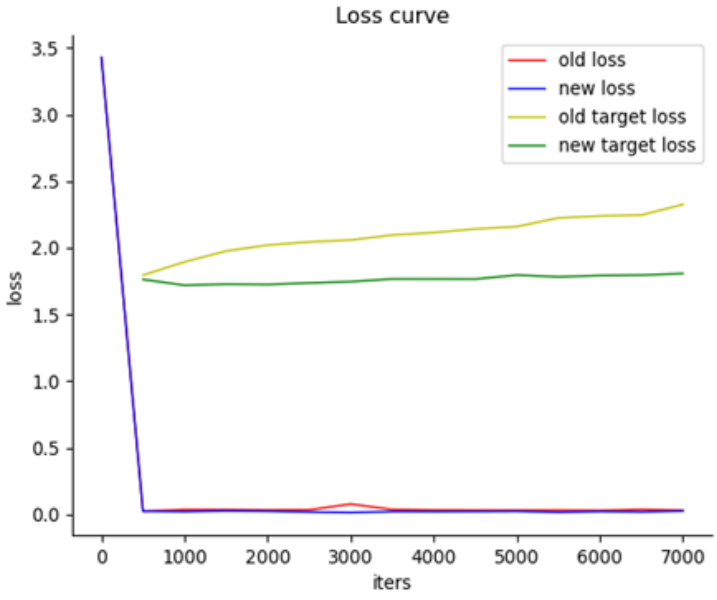

图 3: 损失

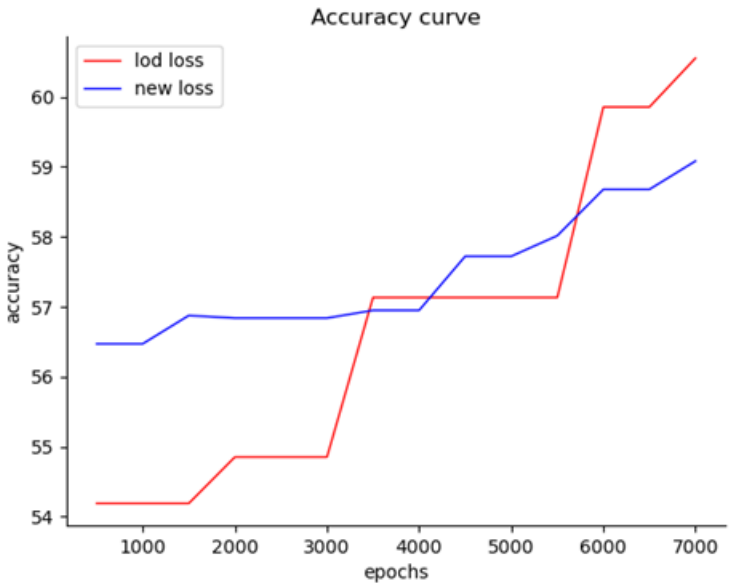

图 4: 预测成功率

6 总结与展望

文章提出了一种新的最小最大熵（MME）方法，该方法对抗性地优化用于半监督域自适应（SSDA）的自适应小样本模型。模型包括一个特征提取网络，然后是一个分类层，该层计算特征与一组估计原型（每个类的代表）的相似度。

自适应是通过交替地相对于分类器最大化未标记目标数据的条件熵和相对于特征提取器最小化它来实现的。通过经验证明该方法确实达到了文章中的预测准确度，该方法为 SSDA 创造了新的技术。

参考文献

[1] LONG M, CAO Z, WANG J, et al. Conditional Adversarial Domain Adaptation[Z]. 2017.

[2] RUI S, BUI H H, NARUI H, et al. A DIRT-T Approach to Unsupervised Domain Adaptation[Z]. 2018.

[3] SAENKO K, KULIS B, FRITZ M, et al. Adapting Visual Category Models to New Domains[J]. European Conference on Computer Vision, 2010.

[4] SANKARANARAYANAN S, BALAJI Y, JAIN A, et al. Learning from Synthetic Data: Addressing Domain Shift for Semantic Segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018.

[5] CHEN Y, LI W, SAKARIDIS C, et al. Domain Adaptive Faster R-CNN for Object Detection in the Wild[Z]. 2018.

[6] SAITO K, USHIKU Y, HARADA T, et al. Strong-Weak Distribution Alignment for Adaptive Object Detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019.

[7] SAITO K, WATANABE K, USHIKU Y, et al. Maximum Classifier Discrepancy for Unsupervised Domain Adaptation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018.

[8] SAITO K, USHIKU Y, HARADA T, et al. Adversarial Dropout Regularization[Z]. 2017.

[9] DAI Z, YANG Z, YANG F, et al. Good Semi-supervised Learning that Requires a Bad GAN[Z]. 2017.

[10] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved Techniques for Training GANs[J]., 2016.

[11] LAINE S, AILA T. Temporal Ensembling for Semi-Supervised Learning[Z]. 2016.

[12] MIYATO T, MAEDA S I, KOYAMA M, et al. Distributional Smoothing with Virtual Adversarial Training[J]. Computer Science, 2015.

[13] ERKAN A, ALTUN Y. Semi-Supervised Learning via Generalized Maximum Entropy[J]. Journal of Machine Learning Research, 2010, 9: 209-216.

[14] SNELL J, SWERSKY K, ZEMEL R S. Prototypical Networks for Few-shot Learning[J]., 2017.

[15] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching Networks for One Shot Learning[Z]. 2016.

[16] LAROCHELLE S. OPTIMIZATION AS A MODEL FOR FEW-SHOT LEARNING[C]//International Conference on Learning Representations. 2017.

[17] SAITO K, KIM D, SCLAROFF S, et al. Semi-supervised Domain Adaptation via Minimax Entropy[J]. ICCV, 2019.