

基于伪标签依赖增强与噪声干扰消减的小样本图像分类

唐芮琪, 肖婷, 迟子秋, 王喆

引用本文

唐芮琪, 肖婷, 迟子秋, 王喆. 基于伪标签依赖增强与噪声干扰消减的小样本图像分类[J]. 计算机科学, 2024, 51(8): 152-159.

TANG Ruiqi, XIAO Ting, CHI Ziqiu, WANG Zhe. [Few-shot Image Classification Based on Pseudo-label Dependence Enhancement and Noise Interference Reduction](#) [J]. Computer Science, 2024, 51(8): 152-159.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于注意力机制的CNN和BiGRU的加密流量分类](#)

Encrypted Traffic Classification of CNN and BiGRU Based on Self-attention
计算机科学, 2024, 51(8): 396-402. <https://doi.org/10.11896/jsjcx.230500032>

[基于原型修正的小样本半监督语义图像翻译算法](#)

Few-shot Semi-supervised Semantic Image Translation Algorithm Based on Prototype Correction
计算机科学, 2024, 51(8): 224-231. <https://doi.org/10.11896/jsjcx.230500038>

[基于双鉴别器和伪视频生成的视频异常检测方法](#)

Video Anomaly Detection Method Based on Dual Discriminators and Pseudo Video Generation
计算机科学, 2024, 51(8): 217-223. <https://doi.org/10.11896/jsjcx.230600148>

[基于多样化标签矩阵的医学影像报告生成](#)

Diversified Label Matrix Based Medical Image Report Generation
计算机科学, 2024, 51(8): 200-208. <https://doi.org/10.11896/jsjcx.230600018>

[嵌入注意力机制的并行多尺度点云上采样方法](#)

Parallel Multi-scale with Attention Mechanism for Point Cloud Upsampling
计算机科学, 2024, 51(8): 183-191. <https://doi.org/10.11896/jsjcx.230500094>

基于伪标签依赖增强与噪声干扰消减的小样本图像分类

唐芮琪 肖婷 迟子秋 王喆

华东理工大学信息科学与工程学院 上海 200237

(18516660531@163.com)

摘要 深度学习在图像分类领域的成功很大程度上依赖于大规模数据,然而在许多应用场景中,收集足够的数据用于模型的训练是比较困难的。因此,旨在利用有限的的数据获得高性能模型的小样本学习成为热点研究方向。在小样本图像分类领域,使用无标签数据来扩充训练数据集是一种常用的方法,但该方法面临两个亟待解决的难题:如何获取无标签数据的伪标签以及如何减轻噪声标签累积的负面影响?首先,为获得高质量的伪标签,需要解决由源域和目标域分布偏移导致的噪声标签问题,因而提出基于希尔伯特-施密特独立准则(Hilbert-Schmidt Independent Criterion, HSIC)的依赖增强方法,通过最大化图像特征表示与标签之间的相关性,从而提高伪标签的预测可靠度。其次,为克服标签预测误差随着时间推移不断累积的问题,提出噪声标签干扰消减(Noise Label Interference Reduction, NLIR)方法,确保具有正确标签的样本的梯度始终主导着训练动态,从而将模型引向最优解。所提方法在小样本图像分类基准数据集 mini-ImageNet 和 tiered-ImageNet 上进行了评估,实验结果表明,该方法能够很好地利用无标签数据提升分类精度,具有良好的分类性能。

关键词:深度学习;图像分类;小样本学习;伪标签;噪声标签;希尔伯特-施密特独立准则

中图分类号 TP391

Few-shot Image Classification Based on Pseudo-label Dependence Enhancement and Noise Interference Reduction

TANG Ruiqi, XIAO Ting, CHI Ziqiu and WANG Zhe

School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

Abstract The success of deep learning in image classification relies heavily on large-scale data. However, in many application scenarios, it is difficult to collect enough data for model training. Therefore, few-shot learning aimed at obtaining high-performance models with limited data becomes a hot research direction. In the field of few-shot image classification, using unlabeled data to augment the training datasets is a common method, but it faces two urgent problems: how to obtain pseudo-labels of unlabeled data and how to mitigate the negative impact of accumulated noise labels? Firstly, in order to obtain high-quality pseudo-labels, it is necessary to solve the problem of noise labels caused by the distribution shift of the source domain and the target domain. A dependence enhancement method based on Hilbert-Schmidt independent criterion is proposed to improve the prediction reliability of pseudo-labels by maximizing the correlation between image feature representation and labels. Secondly, in order to overcome the problem of label prediction error that accumulates over time, a noise label interference reduction method is proposed to ensure that the gradient of samples with correct labels always dominates the training dynamics, so as to lead the model to the optimal solution. The above methods are evaluated on the benchmark datasets for few-shot image classification, namely mini-ImageNet and tiered-ImageNet. The results demonstrate that the proposed approach effectively utilizes unlabeled data to improve classification accuracy and achieves impressive classification performance.

Keywords Deep learning, Image classification, Few-shot learning, Pseudo-labels, Noise-labels, Hilbert-Schmidt independent criterion

1 引言

深度学习作为新兴的人工智能技术,发展十分迅速,基于深度学习的图像分类算法成为国内外学者研究的热点。然而,深度学习需要依靠大规模训练数据的支持来获取具有高性能的

图像分类是计算机视觉领域研究的经典问题。近年来,

到稿日期:2023-05-10 返修日期:2023-09-23

基金项目:上海市科技计划项目(21511100800&20511100600);国家自然科学基金(62076094)

This work was supported by the Shanghai Science and Technology Program(21511100800&20511100600) and National Natural Science Foundation of China(62076094).

通信作者:王喆(wangzhe@ecust.edu.cn)

模型。在实际的应用场景中,由于数据稀缺以及标签注释的成本很高,只有有限的带标签数据可供使用。受到人类能够结合已有经验从少量样本中快速学习新概念的启发,小样本学习应运而生,这种学习方法旨在利用有限的训练数据具有较好泛化效果的网络模型^[1]。大量小样本学习方法集中在元学习范式上,模型基于任务进行训练,任务由支持集和查询集构成,通过插曲式(Episode)训练来模拟人类的快速学习能力。然而,有相关工作^[2]发现,最简单的迁移学习基线——微调,可以达到比元学习更有竞争力的效果。微调的方法首先需要训练一个深度骨干网络,然后在新的小样本任务上添加新的分类器并使用少量的带标签数据进行微调,以适应新任务的特定要求。然而,基于微调的小样本学习不仅数据是有限的,监督也是有限的。

最近的工作开始探索直推式和半监督小样本学习,通过无标签查询集或额外的未标记数据集的信息来提升分类性能。在各种方法中,自训练^[3]是利用未标记数据最直接的方法之一。通常,在支持集上训练得到的模型可用于推断无标签数据的伪标签,然后利用这些伪标签样本和支持集来重新训练模型。由于源域与目标域存在分布偏移,小样本场景下的训练过程又只利用了很少的带标签支持集,模型可能无法正确捕获任务中目标类的数据分布,因此,依赖源模型产生的伪标签不可避免地存在噪声。现有的标签净化方法严重依赖于伪标签本身的质量^[4],通常会受到噪声标签的影响,将不值得信赖的伪标签加入目标类,会导致分类精度下降。

为了克服元学习和传统自训练的局限性,本文提出了一种新的小样本学习方法。该方法不需要引入额外的数据,而是充分利用无标签的查询集来改进模型,以达到数据利用最大化并有效提升图像分类的性能。首先,提出了基于希尔伯特-施密特独立准则的依赖增强方法,该方法最大程度地提高了查询集特征表示与源模型生成的伪标签之间的相关性,有效解决由因分布偏移导致的伪标签质量不佳的问题。其次,进一步提出了噪声标签干扰消减方法,通过梯度纠正的方式,令带有正确标签的样本的梯度始终主导着训练动态,使模型朝着真正的最优路径前进。

本文在 miniImageNet 和 tieredImageNet 两个主流的小样本图像分类基准数据集上开展了对所提方法的相关研究与讨论。实验结果表明,本文所提方法在小样本场景下有良好的分类性能。

2 相关工作

小样本图像分类是一项具有挑战性的机器学习任务,稀缺的训练数据很可能导致过拟合问题。研究人员探索了元学习的思想,通过一种插曲式训练策略将从基类学习到的知识迁移到新类,以提高模型的快速适应能力。元学习方法大致可以分为两类:基于优化的模型^[5]和基于度量的模型。前者的主要思想是学习良好的网络初始化参数,以便分类器快速适应一个未见任务或类。后者试图找到一个合适的距离度量来量化查询图像与给定支持图像之间的关系,如 Matching Networks^[6], ProtoNet^[7], TADAM^[8] 和 Relation Networks^[9],分别在基于余弦相似度、欧几里得距离、度量尺度

缩放和可学习度量的特征嵌入空间中学习一个合适的度量。相比之下,本文方法没有使用复杂的元训练设置,也没有使用任何距离度量。另一类算法是基于生成模型或数据增强^[10-11]的方法,这类方法通过给定的带标签数据合成新的样本或特征来扩展支持集。类似地,本文方法将伪标签技术作用于无标签的查询集,支持集和带伪标签的查询集共同作为训练数据,相当于增加了新样本。下面将介绍与本文所提方法紧密相关的直推式小样本学习以及半监督小样本学习。

直推式小样本学习表明可以访问所有的测试数据,但并不包括其真实标签,这些数据提供了关于新类分布的附加信息。Liu 等^[12]提出了一种标签传播算法,通过图将标签从带标签样本传播到无标签样本。Rodríguez 等^[13]提出了一种特征嵌入传播算法,获得了更平滑的决策边界,提升了模型的抗噪性。Kye 等^[14]提出了元置信度方法,通过元学习一个依赖于数据的缩放函数来规范化每个样本,并迭代更新类中心。Hou 等^[15]提出了基于交叉注意力的方法,通过相关性对齐图像的特征映射,迭代地选择对无标签样本最有信心的预测数据去扩充训练集。Hu 等^[16]提出了一个拉普拉斯正则,以鼓励邻近查询样本具有一致的标签分配。Dhillon 等^[17]提出对经过标准交叉熵损失训练的神经网络进行直推式微调。

在半监督小样本学习方法中,除了支持集和查询集之外,还使用了额外的未标记数据。Li 等^[18]对未标记集应用自标记和软注意力,并对标记和自标记样本进行微调。Ren 等^[19]针对未标记样本提出了基于软赋值的原型细化。Wang 等^[20]引入线性回归假设,选择伪标签样本进行分类器训练。与上述方法不同的是,本文方法所针对的未标记数据指测试集中的查询集,并不引入额外的数据开销。

3 模型与方法

3.1 问题定义

在小样本图像分类任务中,给定基类集合 $D_{\text{base}} = \{(x_i, y_i), x_i \in X_{\text{base}}, y_i \in Y_{\text{base}}\}$,其中 X_{base} 表示基类数据集, Y_{base} 为基类数据集的类别标签。新类集合为 $D_{\text{novel}} = \{(x_i, y_i), x_i \in X_{\text{novel}}, y_i \in Y_{\text{novel}}\}$,其中 X_{novel} 表示新类数据集, Y_{novel} 为新类数据集的类别标签。基类数据集和新类数据集中的类别没有重叠,即 $Y_{\text{base}} \cap Y_{\text{novel}} = \emptyset$ 。针对小样本任务标准的 N -way K -shot 范式,首先在基类集合 D_{base} 上进行训练,然后对于每一个任务 T_i ,从 D_{novel} 中随机抽取 N 个类,对 N 个类中的每一个类抽取 K 个带标签样本用于构造支持集 $D_{T_i}^S$,支持集的大小为 $|D_{T_i}^S| = N \times K$ 。类似地,从与支持集类别相同的 N 个类中抽取 Q 个样本作为每个任务中的查询集 $D_{T_i}^Q$,查询集的大小为 $|D_{T_i}^Q| = N \times Q$ 。小样本学习的目标是利用基类集合训练得到一个初始模型,该模型通过对新类集合进行进一步的学习,能够正确识别出新类中的未见样本。

3.2 总体框架

本文设计了一种基于伪标签依赖增强和噪声干扰消减的小样本学习方法,通过重构训练目标分类器的损失函数,解决伪标签质量不佳以及噪声标签累积的问题。训练目标分类器的总体框架如图 1 所示。模型训练分为两个阶段,在预训练

阶段,利用全部的基类数据来训练特征提取器以及基分类器,该阶段得到的特征提取器将用于提取新类图像的特征。在微调阶段,首先利用基分类器预测支持集的标签,计算得到预测标签与支持集真实标签之间的交叉熵损失 L_{CE} 。然后利用

基分类器对查询集的伪标签进行预测,将预测结果和查询集特征作为基于 HSIC 的依赖增强以及噪声标签干扰消减算法的输入,得到另外两项损失函数 L_{HSIC} 以及 L_{NLIR} ,与支持集上的交叉熵损失联合训练目标分类器。

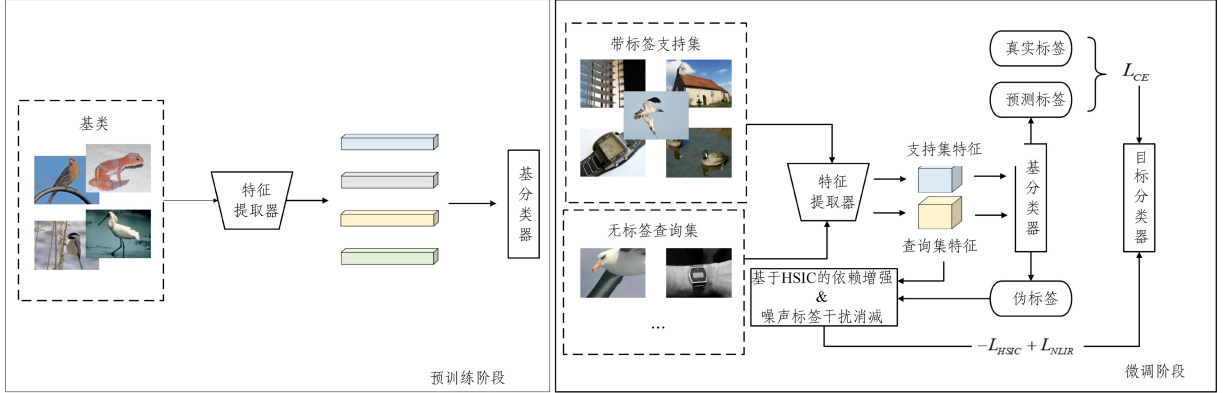


图1 基于重构损失函数的目标分类器训练总体框架

Fig.1 Overall framework of training target classifier based on reconstructed loss function

最后,我们重新设计的用于训练目标分类器的总损失函数为:

$$L = L_{CE} - \lambda_1 L_{HSIC} + \lambda_2 L_{NLIR} \quad (1)$$

其中, λ_1 和 λ_2 为超参数, L_{HSIC} 和 L_{NLIR} 两项损失将分别在 3.3 节及 3.4 节展开说明。

3.3 基于 HSIC 的依赖增强

本文提出的基于伪标签的学习算法首先利用基类带标签样本来训练源模型,再利用该源模型生成新类中查询集的伪标签,最后使用新类中的带标签支持集样本以及带伪标签查询集样本共同训练目标域模型。

如图 2 所示,由于源域和目标域之间存在分布差异,因此依赖源域生成的伪标签准确度并不理想。为了获得更高质量的伪标签,本节充分挖掘图像特征与其标签之间的内在联系,通过强化两者的依赖关系来提升伪标签的质量。

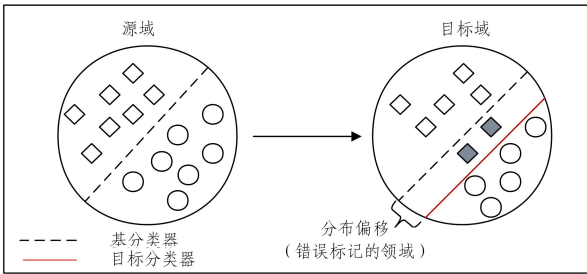


图2 分布偏移产生噪声标签

Fig.2 Noise labels caused by distribution shift

HSIC^[21],即希尔伯特-施密特独立准则,是一种基于核函数的独立性度量统计检验方法。HSIC 利用核函数将查询集特征及其伪标签映射到高维空间中,并度量两个变量之间的相似度,以检测它们是否独立。如果两者相互独立,那么它们在高维空间中的位置是随机的并且相似度为 0,反之,则两者存在一定的依赖关系。我们的目标是增强查询集特征和预测伪标签之间的相关性,具体而言,就是通过最大化两者的 HSIC 经验估计值来实现。

假设 \mathbf{X} 是与查询集特征表示相关的随机向量, \mathbf{Y} 是与

查询集伪标签预测相关的随机向量。令 \mathbf{F} 为 \mathbf{X} 的再生核希尔伯特空间, \mathbf{K} 是特征嵌入空间到再生核希尔伯特空间的核函数,记作 $\mathbf{K}; \mathbf{X} \rightarrow \mathbf{F}$,核函数写为:

$$f(x, x') = \langle \mathbf{K}(x), \mathbf{K}(x') \rangle_{\mathbf{F}}, x, x' \in \mathbf{X} \quad (2)$$

其中, $\langle \cdot, \cdot \rangle_{\mathbf{F}}$ 表示空间 \mathbf{F} 上的内积。

在查询集的特征上选择的核函数为高斯核,即

$$f(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), x_i, x_j \in \mathbf{X} \quad (3)$$

类似地,令 \mathbf{L} 为 \mathbf{Y} 的再生核希尔伯特空间, Ψ 是预测标签向量空间到再生核希尔伯特空间的核映射,记作 $\Psi; \mathbf{Y} \rightarrow \mathbf{L}$ 。核函数写为:

$$l(y, y') = \langle \Psi(y), \Psi(y') \rangle_{\mathbf{L}}, y, y' \in \mathbf{Y} \quad (4)$$

在查询集的预测标签上选择的核函数为线性核,即:

$$l(y_i, y_j) = y_i^T y_j, y_i, y_j \in \mathbf{Y} \quad (5)$$

HSIC 采用的是希尔伯特-施密特互协方差算子,通过对该算子范数的经验估计得到独立性判断准则。为了度量 \mathbf{X} 和 \mathbf{Y} 之间的相关性,定义互协方差算子如下:

$$C_{x,y} = E_{x,y}[(\mathbf{K}(x) - \mu_x) \otimes (\Psi(y) - \mu_y)] \quad (6)$$

其中, μ_x 和 μ_y 为向量均值, \otimes 表示张量积。

为了描述 \mathbf{X} 和 \mathbf{Y} 之间的依赖程度,本文使用 $C_{x,y}$ 的希尔伯特-施密特范数,令 $P_{X,Y}$ 为 \mathbf{X} 和 \mathbf{Y} 联合分布,满足如下公式:

$$HSIC(\mathbf{F}, \mathbf{L}, P_{X,Y}) = \|C_{x,y}\|_{HS}^2 \quad (7)$$

得到样本数据 $\mathbf{Q} = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ 后,可以计算出 HSIC 的经验估计值:

$$HSIC(\mathbf{F}, \mathbf{L}, \mathbf{Q}) = (U-1)^{-1} \text{tr}(\mathbf{F}\mathbf{H}\mathbf{L}\mathbf{H}) \quad (8)$$

其中, U 为查询集的样本个数, \mathbf{H} 是一个中心矩阵,定义为 $\mathbf{I}_U - (1/U)\mathbf{1}_U\mathbf{1}_U^T$,其中 \mathbf{I}_U 为单位矩阵, $\mathbf{1}_U$ 为全 1 的列向量。 $\text{tr}(\cdot)$ 为矩阵迹运算。 \mathbf{F} 和 \mathbf{L} 都定义在查询集上, \mathbf{F} 为特征表示的 Gram 矩阵, \mathbf{L} 为预测伪标签的 Gram 矩阵。

当且仅当 \mathbf{X} 和 \mathbf{Y} 独立时,即查询集特征和预测伪标签是独立的, HSIC 的经验估计值为 0。而我们的目的是增强查询集特征与其伪标签之间的依赖,因此需要最大化 HSIC 的

经验估计值。具体而言,将 HSIC 的经验估计值作为损失函数中的一项 L_{HSIC} ,代表分类器在查询集上的经验风险,这有助于限制分类器的假设空间,理论上能够对给定的无标签样本进行更准确的预测。由于我们的目标是最大化 L_{HSIC} ,因此在将其作为损失函数中的一项时,通过添加负号将最大化问题转化为最小化问题。

3.4 噪声标签干扰消减

采用基于 HSIC 的依赖增强方法可以在一定程度上提高伪标签的质量,但是并不能够保证预测伪标签的完全正确。然而,将不值得信赖的伪标签加入目标类别会导致分类精度下降,背离了我们使用无标签数据来提升精度的目标。

随着训练的进行,受到噪声标签的影响,预测误差将会不断累积,导致模型性能变差。然而,有研究发现一种“早期训练现象”^[22],即在早期学习阶段,正确标记的样本的梯度主导着训练动态,使得前期朝着真正的最优路径前进,但后来错误标记的样本的梯度逐渐占据主导地位,此时分类器只学习拟合噪声标签。基于此,本文提出了噪声标签干扰消减的方法,通过梯度纠正,鼓励模型延续早期对样本的高精度预测。

假设源域 \mathcal{D}_s 和目标域 \mathcal{D}_T 在 $\mathbf{X} \times \mathbf{Y}$ 上是具有相同先验的双分量高斯混合分布^[23-25],定义查询集样本 x 存在一个最高概率属于正确的类 y , x 从分布 $N(y\mu, \sigma)$ 中采样。将不正确的伪标签视为噪声标签,用 \tilde{y} 表示:

$$\tilde{y} = \text{ysign}(\mathbb{I}\{y\mathbf{x}^T\boldsymbol{\mu} > r\} - 0.5) \quad (9)$$

其中, $\text{sign}(\cdot)$ 为符号函数, $\mathbb{I}(\cdot)$ 为指示函数, r 控制错误标记的概率。

下面计算在早期训练时间 T ,对错误标记样本的预测准确率。假设 $\mathbf{Z} = \{\mathbf{X}; \tilde{\mathbf{Y}} \neq \mathbf{Y}\}$ 是一组错误标记的样本,令 $\mathbb{k}(\mathbf{Z}; \theta)$ 是由真实标签 y 和分类器 $f(x; \theta)$ 在错误标记样本上得到的预测标签共同计算出的预测准确率,其中 θ 为分类器参数。 $\mathbb{k}(\mathbf{Z}; \theta)$ 满足定理^[22]:如果最多有一半的样本是错误标记的 ($r < 1$),那么存在一个固有时 T 和常数 $c_0 > 0$,使得对任意 $0 < \sigma < c_0$,有:

$$\mathbb{k}(\mathbf{Z}; \theta_T) \geq 1 - \exp\left\{-\frac{1}{200}g(\sigma^2)\right\} \quad (10)$$

其中, $g(\sigma)$ 表示如下:

$$g(\sigma) = \frac{2}{\sqrt{\pi}} \int_0^{\frac{1-r}{\sqrt{2}\sigma}} e^{-t^2} dt + \frac{\exp\left(-\frac{(r-1)^2}{2\sigma^2}\right)}{\sqrt{2\pi}(1+2\sigma)} \quad (11)$$

对任意 $\sigma > 0$,有 $g(\sigma) > 0$ 。 $g(\sigma)$ 是一个单调递减函数,当 σ 趋近于 0 时, $g(\sigma)$ 趋近于无穷大,所有错误标记样本的预测值等于它们的真实标签。

上述定理证明,在适当的时间 T ,分类器可以为错误标记的样本提供准确的预测。但当分类器继续训练时,它开始逐渐记忆错误标签。为了利用模型早期训练预测标签准确率高的性质,本文提出了噪声标签干扰消减的方法,通过修正梯度,防止模型记忆错误标签,从而减少噪声标签的影响。噪声标签干扰消减项可以用如下公式表示:

$$L_{\text{NLIR}}(\theta_t) = \log(1 - y_t^T f(x; \theta_t)) \quad (12)$$

其中, $f(x; \theta_t)$ 为分类器输出的样本标签概率, $y_t = \beta y_{t-1} + (1-\beta)f(x; \theta_t)$ 是 x 的滑动平均预测 (Moving Average Pre-

diction, MAP), β 是超参数。

为了进一步说明如何防止记忆噪声标签,我们对样本标签概率求导,如式(13)所示:

$$\frac{dL_{\text{NLIR}}(\theta_t)}{df(x; \theta_t)} = -\frac{y_t}{1 - y_t^T f(x; \theta_t)} \quad (13)$$

滑动平均预测很可能是真实标签^[26],但将其直接作为伪标签无法充分利用“早期训练现象”的特性,因为在训练中后期,噪声标签的梯度仍将主导训练。本文方法通过最小化式(13),使 $f(x; \theta_t)$ 与 y_t 接近,即强制模型的预测结果与滑动平均预测保持接近。这是一种梯度纠正方法,该方法能够使模型预测与滑动平均预测对齐的梯度压倒将模型预测与噪声标签对齐的其他损失项的梯度,在学习过程中,隐式地防止模型记忆噪声标签。

4 数据集

本文在两个广泛使用的小样本图像分类基准数据集 mini-ImageNet 和 tiered-ImageNet 上分别进行了实验。mini-ImageNet 是 ImageNet^[27] 的一个子集,由 100 个类组成,这些类被随机分为 64 个基类,16 个验证类和 20 个新类。每个类中有 600 个样本,共计 60 000 张大小为 84×84 的图像。tiered-ImageNet 同样是 ImageNet 的一个子集,但包含了比 mini-ImageNet 更多的类别,有 608 个子类。这些子类来自 34 个超类,其中 20 个(351 个子类)为基类,6 个(97 个子类)为验证类,8 个(160 个子类)为新类,共计 779 165 张大小为 84×84 的图像。

5 实验

5.1 实验细节

本实验利用 Pytorch 深度学习框架对所提出的模型进行训练和测试。本文方法从头开始训练,采用 ResNet-12 作为特征提取主干网络。为了训练 softmax 分类器,使用学习率为 0.0001 的 Adam 优化器。在预训练阶段,我们进行了 400 轮训练,批大小为 16。在微调阶段,将整个数据集划分成多个小的训练集合,其中 5-way 1-shot 任务的训练集合大小为 60 000,5-way 5-shot 任务的训练集合大小为 40 000。使用验证集来选择分类精度最高的一次训练结果并保存。在预训练和微调阶段都应用了标准的数据增强,包括随机裁剪、随机翻转、色彩抖动等。此外,还对数据进行了自注意力、交叉注意力特征表示的高阶处理。最后,我们在一个训练集合中为每个类测试了 15 个查询样本,并得到了在 95% 置信区间,随机抽样 2000 个测试集的平均分类精度。一些重要的超参数设置为:基于 HSIC 的依赖增强损失项系数 $\lambda_1 = 0.01$,噪声标签干扰消减损失项系数 $\lambda_2 = 3$,高斯核带宽 $\sigma = 0.5$,滑动平均预测的平滑系数 $\beta = 0.5$ 。

5.2 对比实验

针对小样本图像分类任务,通常选择分类精度(准确度)来评估模型的性能和分类效果。将本文方法与其他小样本图像分类方法进行比较,并在 mini-ImageNet 和 tiered-ImageNet 数据集以及常用的 5way-1shot 以及 5way-5shot 设置下检测它们的分类精度,具体结果如表 1 所列。

表 1 与其他方法的精度比较

Table 1 Accuracy comparison of the proposed method with other methods

Setting	Method	Backbone	Accuracy (%)			
			mini-ImageNet		tiered-ImageNet	
			5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
Inductive	MAML ^[5]	ConvNet-4	48.70±1.84	63.11±0.92	51.67±1.81	70.30±1.75
	ProtoNet ^[7]	ResNet-12	62.11±0.44	80.77±0.30	68.31±0.51	83.85±0.36
	APP2S ^[28]	ResNet-12	66.25±0.20	83.42±0.15	72.00±0.22	86.23±0.15
	DAM ^[29]	ResNet-12	60.39±0.21	73.84±0.16	64.09±0.23	78.39±0.18
Transductive	CAN+T ^[15]	ResNet-12	67.19±0.55	80.64±0.35	73.21±0.58	84.93±0.38
	DPGN ^[30]	ResNet-12	67.77±0.32	84.60±0.43	72.45±0.51	87.24±0.39
	DSN-MR ^[31]	ResNet-12	64.60±0.72	79.51±0.50	67.39±0.82	82.85±0.56
	EASE+SIAMESE ^[32]	ResNet-12	70.47±0.30	80.73±0.16	84.54±0.27	89.63±0.15
Semi-Supervised	TPN ^[12]	ResNet-12	62.70±1.37	74.20±0.31	72.10±1.49	83.30±0.33
	LST ^[18]	ResNet-12	70.01±1.90	78.70±0.80	77.70±1.60	85.20±0.80
	TransMatch ^[33]	WRN-28-10	60.02±1.02	79.30±0.59	72.19±1.27	82.12±0.92
	ICI ^[20]	ResNet-12	69.66±1.13	80.11±0.72	84.01±1.03	89.00±0.67
Transductive	HSIC+MAP(ours)	ResNet-12	70.27±0.45	84.62±0.31	84.37±0.51	89.64±0.35
	HSIC+NLIR(ours)	ResNet-12	70.53±0.45	84.73±0.31	84.74±0.51	89.99±0.35

对比方法包括三大类:归纳式(Inductive)学习、直推式(Transductive)学习以及半监督(Semi-Supervised)学习。归纳式学习在训练过程中不访问测试数据,仅使用训练集中的数据信息。MAML和ProtoNet分别是基于优化、基于度量的元学习方法;APP2S^[28]利用庞加莱嵌入建模类之间的几何关系,根据样本分布自适应调整距离度量;DAM^[29]学习了一个通用度量网络,能够生成适用于多个小样本分类任务的度量函数。直推式学习在训练过程中访问了测试数据,但不包括测试数据的标签,本文方法也属于直推式学习。CAN+T引入了交叉注意力机制,更好地利用了支持集和查询集之间的信息交互;DPGN^[30]通过建立一个分布传播图网络来表达样本之间的关系;DSN-MR^[31]通过学习每个类别的子空间来提升性能;EASE+SIAMESE^[32]结合了无监督判别子空间学习和均值偏移聚类。半监督学习包括基于标签传播的TPN、基于自训练和动态权重更新机制的LST、基于迁移学习的TransMatch^[33]以及基于实例置信度的ICI方法。此外,我们

还将基于HSIC的依赖增强与3.4节提到的滑动平均预测方法相结合,作为一种比较方法。

由表1可知,本文所提的基于伪标签依赖增强与噪声干扰消减的小样本图像分类方法的表现优于上述对比方法,在两个数据集上均取得了最优性能。实验结果验证了本文方法的有效性和优越性,这得益于本文方法能充分利用已有的数据信息和模型的“早期训练现象”,并针对因数据分布偏移和噪声标签而产生的问题进行了有效处理。

本文方法的t-SNE降维可视化^[34]结果如图3所示。从mini-ImageNet数据集中抽取来自5个不同类别的共计500张图像,将其映射到二维特征空间。其中,不同的形状代表不同的类别,叉号表示错误的预测。图3(a)显示了这500张图像的分布,并且带有真实标签;图3(b)直接利用源模型估计图像的伪标签;图3(c)引入了基于HSIC的依赖增强方法估计伪标签;图3(d)同时使用了基于HSIC的依赖增强和噪声标签干扰消减。

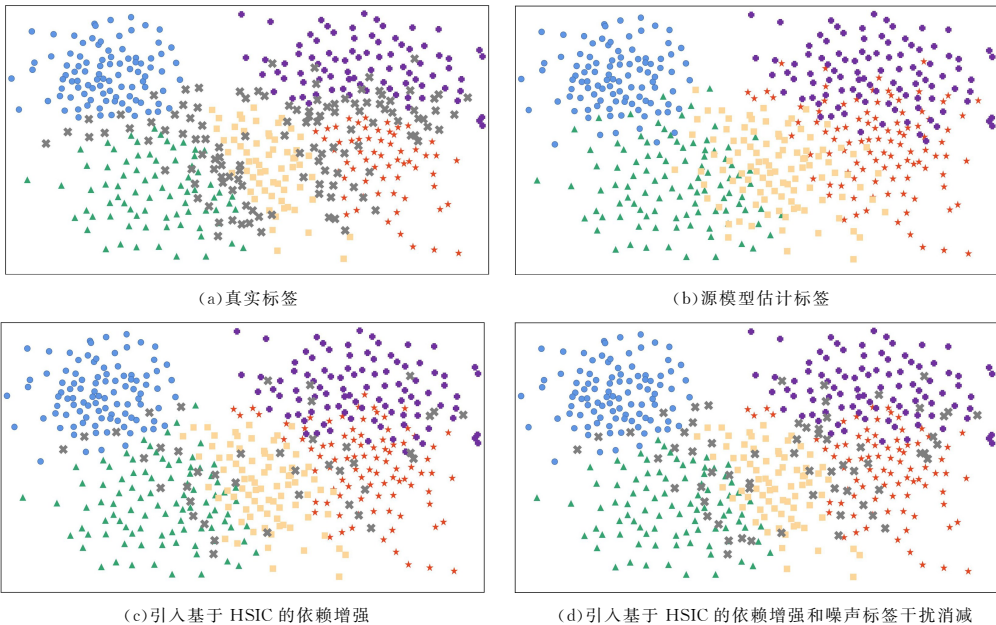


图 3 t-SNE 标签预测可视化结果

Fig. 3 Visualization results of label prediction based on t-SNE

观察图 3 可以发现,直接利用源模型生成的伪标签中存在着较多的错误标记。然而,当我们引入基于 HSIC 的依赖增强方法后,通过特征-标签依赖最大化,极大地提升了伪标签的质量,错误标记的标签数量大大减少;当再次引入噪声标签干扰消减方法后,通过延续“早期训练现象”,进一步减少了错误标记的标签数量。

5.3 消融实验

为了深入探究基于 HSIC 的依赖增强和噪声标签干扰消减对实验结果的贡献,我们进行了消融实验,并分别在 mini-ImageNet 和 tiered-ImageNet 数据集上测试了各方法的性能。具体地,我们比较了不使用任何本文方法、仅使用基于 HSIC 的依赖增强方法,以及叠加噪声标签干扰消减方法的分类效果,结果如表 2 和表 3 所列。

表 2 在 mini-ImageNet 上的消融实验
Table 2 Ablation study on mini-ImageNet (%)

HSIC	NLIR	mini-ImageNet	
		5-way 1-shot	5-way 5-shot
×	×	67.31±0.24	82.45±0.21
✓	×	70.21±0.45	84.59±0.31
✓	✓	70.53±0.45	84.73±0.31

在 mini-ImageNet 数据集上,引入基于 HSIC 的依赖增强方法在 5way-1shot 设置下分类精度提升约 2.90%,在 5way-5shot 设置下分类精度提升约 2.14%。同时,叠加引入噪声标签干扰消减方法可以进一步提升分类性能,在 5way-1shot 设置下分类精度提升约 3.22%,在 5way-5shot 设置下分类精度提升约 2.28%。

表 3 在 tiered-ImageNet 上的消融实验
Table 3 Ablation study on tiered-ImageNet (%)

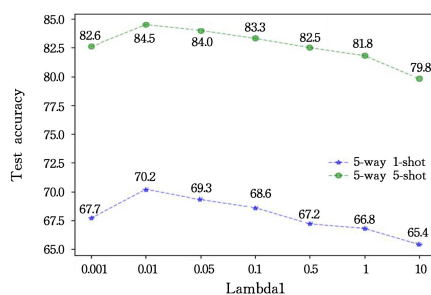
HSIC	NLIR	tiered-ImageNet	
		5-way 1-shot	5-way 5-shot
×	×	82.14±0.49	88.11±0.24
✓	×	84.36±0.45	89.62±0.35
✓	✓	84.74±0.51	89.99±0.35

在 tiered-ImageNet 数据集上,引入基于 HSIC 的依赖增强方法在 5way-1shot 设置下分类精度提升约 2.22%,在 5way-5shot 设置下分类精度提升约 1.51%。同时,叠加引入噪声标签干扰消减方法可以进一步提升分类性能,在 5way-1shot 设置下分类精度提升约 2.60%,在 5way-5shot 设置下分类精度提升约 1.88%。根据实验结果可以得出结论:基于 HSIC 的依赖增强方法显著提高了分类精度,而噪声标签干扰消减方法则在一定程度上进一步提高了分类精度,两者结合对实现高效的小样本图像分类具有重要作用。

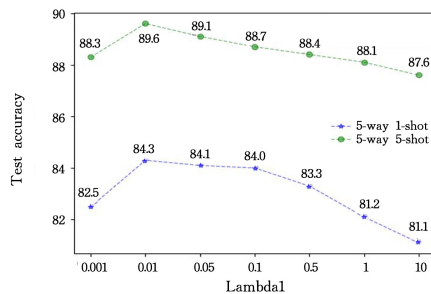
5.4 损失函数超参数

损失函数超参数的存在是为了平衡本文所提出的两种方法对基于交叉熵损失分类算法的影响。我们在两个小样本数据集上分别对总损失函数里的超参数 λ_1 和 λ_2 进行分析讨论。 λ_1 的值从 {0.001, 0.01, 0.05, 0.1, 0.5, 1, 10} 中选择, λ_1 对实验结果的影响如图 4 所示。从图中可以观察到,无论是在 mini-ImageNet 数据集还是 tiered-ImageNet 数据集上,当

λ_1 取值为 0.01 时,模型均达到了最优性能。



(a) mini-ImageNet

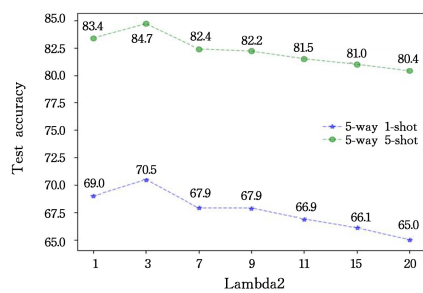


(b) tiered-ImageNet

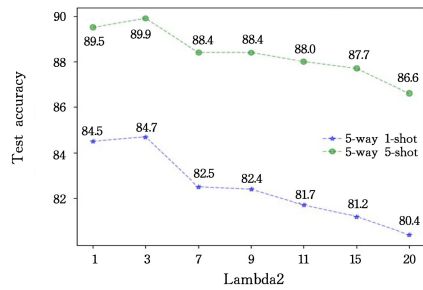
图 4 λ_1 对分类精度的影响

Fig. 4 Influence of λ_1 on classification accuracy

在对 λ_2 进行分析时,首先将超参数 λ_1 固定为最优值 0.01。 λ_2 的值从 {1, 3, 7, 9, 11, 15, 20} 中选择, λ_2 对实验结果的影响如图 5 所示。可以观察到,当 λ_2 取值为 3 时,模型在两个数据集上均达到最优性能。



(a) mini-ImageNet



(b) tiered-ImageNet

图 5 λ_2 对分类精度的影响

Fig. 5 Influence of λ_2 on classification accuracy

结束语 本文提出了一种基于伪标签依赖增强与噪声标签干扰消减的小样本图像分类方法。该方法利用无标签的查询集来提升分类器的训练效果,有效解决了基于伪标签的小样本学习所面临的两大难题:域偏移导致的伪标签质量低下以及标签预测误差随着训练不断累积。本文方法通过重构

分类损失函数, 强化图像特征与伪标签之间的依赖关系, 并采用纠正梯度的方式使模型保持高准确性的预测。实验结果表明, 该方法在小样本场景下具有优异的性能。在未来的工作中, 我们将探索如何获取更多高质量的伪标签, 并采用更加有效的方式来充分利用这些伪标签, 进一步提升图像分类的性能。本文提出的方法为小样本图像分类问题提供了一种有效的解决方案, 并为将来的研究提供了参考。

参 考 文 献

- [1] ZHAO K L, JIN X L, WANG Y Z. A Review of Few-shot Learning[J]. *Journal of Software*, 2021, 32(2): 349-369.
- [2] CHEN W Y, LIU Y C, KIRA Z, et al. A Closer Look at Few-shot Classification[J]. arXiv:1904.04232, 2019.
- [3] RAINA R, BATTLE A, LEE H, et al. Self-taught Learning: Transfer Learning from Unlabeled Data[C]// *Proceedings of the 24th International Conference on Machine Learning*. Corvallis, USA: ACM, 2007: 759-766.
- [4] LIANG J, HU D, FENG J. Do We Really Need to Access the Source Data? Source Hypothesis Transfer for Unsupervised Domain Adaptation [C] // *Proceedings of the 37th International Conference on Machine Learning*. Online: PMLR, 2020: 6028-6039.
- [5] FINN C, ABBEEL P, LEVINE S. Model-agnostic Meta-learning for Fast Adaptation of Deep Networks[C]// *Proceedings of the 34th International Conference on Machine Learning*. Sydney, Australia: PMLR, 2017: 1126-1135.
- [6] VINALS O, BLUNDELL C, LILICRAP T, et al. Matching Networks for One Shot Learning[C]// *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Barcelona, Spain: MIT Press, 2016: 3630-3638.
- [7] SNELL J, SWERSKY K, ZEMEL R. Prototypical Networks for Few-shot Learning[C]// *Proceedings of the 31th International Conference on Neural Information Processing Systems*. Long Beach, USA: ACM, 2017: 4080-4090.
- [8] ORESHKIN B, RODRÍGUEZ LÓPEZ P, LACOSTE A. Tadam: Task Dependent Adaptive Metric for Improved Few-shot Learning[C]// *Proceedings of the 32th International Conference on Neural Information Processing Systems*. Montréal, Canada: MIT Press, 2018: 721-731.
- [9] SUNG F, YANG Y, ZHANG L, et al. Learning to Compare: Relation Network for Few-shot Learning[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE Press, 2018: 1199-1208.
- [10] ZHANG R, CHE T, GHAHRAMAN Z, et al. Metagan: An Adversarial Approach to Few-shot Learning[C]// *Proceedings of the 32th International Conference on Neural Information Processing Systems*. Montréal, Canada: MIT Press, 2018: 2365-2374.
- [11] LI K, ZHANG Y L, LI K, et al. Adversarial Feature Hallucination Networks for Few-shot Learning[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York, USA: IEEE Press, 2020: 13470-13479.
- [12] LIU Y, LEE J, PARK M, et al. Learning to Propagate Labels: Transductive Propagation Network for Few-shot Learning[J]. arXiv:1805.10002, 2018.
- [13] RODRÍGUEZ P, LARADJI I, DROUIN A, et al. Embedding Propagation: Smoother Manifold for Few-shot Classification [C]// *European Conference on Computer Vision*. Graz, Germany: Springer, 2020: 121-138.
- [14] KYE S M, LEE H B, KIM H, et al. Meta-learned Confidence for Few-shot Learning[J]. arXiv:2002.12017, 2020.
- [15] HOU R, CHANG H, MA B, et al. Cross Attention Network for Few-shot Classification[C]// *Proceedings of the 33th International Conference on Neural Information Processing Systems*. Vancouver Canada: MIT Press, 2019: 4005-4016.
- [16] HU S X, MORENO P G, XIAO Y, et al. Empirical Bayes Transductive Meta-learning with Synthetic Gradients[C]// *Proceedings of the 8th International Conference on Learning Representations*. Addis Ababa, Ethiopia: OpenReview. net, 2020.
- [17] DHILLON G S, CHAUDHARI P, RAVICHANDRAN A, et al. A Baseline for Few-shot Image Classification[C]// *Proceedings of the 8th International Conference on Learning Representations*. Addis Ababa, Ethiopia: OpenReview. net, 2020.
- [18] LI X, SUN Q, LIU Y, et al. Learning to Self-Train for Semi-supervised Few-Shot Classification[C]// *Proceedings of the 33th International Conference on Neural Information Processing Systems*. Online: MIT Press, 2019: 10276-10286.
- [19] REN M, TRIANTAFILLOU E, RAVI S, et al. Meta-learning for Semi-supervised Few-shot Classification [J]. arXiv: 1803.00676, 2018.
- [20] WANG Y, XU C, LIU C, et al. Instance Credibility Inference for Few-shot Learning[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York, USA: IEEE Press, 2020: 12836-12845.
- [21] GREENFELD D, SHALIT U. Robust Learning with the Hilbert-schmidt Independence Criterion[C]// *Proceedings of the 37th International Conference on Machine Learning*. Online: PMLR, 2020: 3759-3768.
- [22] LIU S, NILES-WEED J, RAZAVIAN N, et al. Early-learning Regularization Prevents Memorization of Noisy Labels[C]// *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada: MIT Press, 2020: 20331-20342.
- [23] BOUVEYRON C, GIRARD S, SCHMID C. High-dimensional Data Clustering[J]. *Computational Statistics & Data Analysis*, 2007, 52(1): 502-519.
- [24] FRALEY C, RAFTERY A E. Model-based Clustering, Discriminant Analysis, and Density Estimation[J]. *Journal of the American Statistical Association*, 2002, 97(458): 611-631.
- [25] MCNICHOLAS P D, MURPHY T B. Model-based Clustering of Microarray Expression Data via Latent Gaussian Mixture Models[J]. *Bioinformatics*, 2010, 26(21): 2705-2712.
- [26] LAINE S, AILA T. Temporal Ensembling for Semi-supervised Learning[J]. arXiv:1610.02242, 2016.
- [27] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet Large

- Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [28] MA R, FANG P, DRUMMOND T, et al. Adaptive Poincaré Point to Set Distance for Few-shot Classification[C]// Proceedings of the 36th AAAI Conference on Artificial Intelligence. Los Angeles, USA: AAAI, 2022: 1926-1934.
- [29] ZHOU F, ZHANG L, WEI W. Meta-generating Deep Attentive Metric for Few-shot Classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6863-6873.
- [30] YANG L, LI L, ZHANG Z, et al. DPGN: Distribution Propagation Graph Network for Few-shot Learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2020: 13390-13399.
- [31] SIMON C, KONIUSZ P, NOCK R, et al. Adaptive Subspaces for Few-shot Learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE Press, 2020: 4136-4145.
- [32] ZHU H, KONIUSZ P. EASE: Unsupervised Discriminant Subspace Learning for Transductive Few-shot Learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2022: 9078-9088.
- [33] YU Z, CHEN L, CHENG Z, et al. Transmatch: A Transfer-learning Scheme for Semi-supervised Few-shot Learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2020: 12856-12864.
- [34] KOBAK D, LINDERMAN G C. Initialization is Critical for Preserving Global Data Structure in Both t-SNE and UMAP[J]. Nature Biotechnology, 2021, 39(2): 156-157.



TANG Ruiqi, born in 1999, postgraduate. Her main research interests include few-shot learning and image processing.



WANG Zhe, born in 1981, Ph.D, professor, is a member of CCF(No. 16666M). His main research interests include pattern recognition and image processing.

(责任编辑:何杨)