

# 面向小样本学习的动态分布校正方法

赵金伟<sup>1,2</sup>, 任文静<sup>1,2</sup>, 周锦绣<sup>1,2</sup>, 黑新宏<sup>1,2\*</sup>

(1. 西安理工大学 计算机科学与工程学院, 陕西 西安 710048;

2. 网络计算与安全技术陕西省重点实验室, 陕西 西安 710048)

**摘要:**近年来,机器学习在多领域取得巨大成功。在现实世界的真实场景中,由于采集环境复杂或标注难,使得可用于训练的标准样本很少,导致机器学习模型往往出现过学习或欠学习的问题。所以,小样本学习是具有挑战性的机器学习问题。近期人们提出分布校正方法,他们假设特征表示中每个维度都服从高斯分布,利用基类的特征分布来校正新类的特征分布。然而该方法易引入负迁移现象,并且易于淹没新类本身的特征分布。为此,该文提出一种动态分布校正方法解决分布校正过程中的负迁移问题。首先,基于阈值动态地选择近邻基类和远域基类;其次,新类样本的特征经过幂变换的操作后,增加特征标准化处理来消除不同量纲之间的差异;最后,该方法引入参数调节迁移分布与新类特征原分布之间的比例关系来实现新类的特征分布校准。通过在常规数据集 miniImageNet 和 CUB 上与最新算法和传统算法的大量对比实验表明,该方法可以有效提升小样本分类任务的性能。

**关键词:**小样本学习;负迁移;分布校正;阈值;动态分布校正

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2023)06-0173-08

doi:10.3969/j.issn.1673-629X.2023.06.026

## Method of Dynamic Distribution Calibration for Few-shot Learning

ZHAO Jin-wei<sup>1,2</sup>, REN Wen-jing<sup>1,2</sup>, ZHOU Jin-xiu<sup>1,2</sup>, HEI Xin-hong<sup>1,2\*</sup>

(1. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China;

2. Shaanxi Key Laboratory of Network Computing and Security Technology, Xi'an 710048, China)

**Abstract:** In recent years, machine learning has achieved great success in many fields. In the real scene of the real world, due to the complex collection environment or difficult annotation, there are few standard samples available for training, resulting in the problem of over learning or under learning of machine learning models. Therefore, few-shot learning is a challenging machine learning problem. Recently, people have proposed a method of distribution calibration, which assume that each dimension in the feature representation obeys a Gaussian distribution, and use the feature distribution of the base classes to calibrate the feature distribution of the novel classes. However, this method is easy to introduce the phenomenon of negative migration, and it is easy to submerge the characteristic distribution of the novel classes itself. Therefore, we propose the dynamic distribution calibration to solve the problem of negative migration in the distribution correction method. Firstly, the base classes of the nearest neighbor and the base classes of the far domain based on the threshold value are selected dynamically. Secondly, the standardization processing is added to the sample characteristics of the novel classes after power transformation to eliminate the differences between different dimensions. Finally, the method introduces parameters to adjust the proportional relationship between the migration distribution and the original distribution of the features of the novel classes to achieve the calibration of the feature distribution of the novel classes. A large number of comparative experiments with the latest algorithm and the traditional algorithm on the conventional data sets miniImageNet and CUB show that the proposed method can effectively improve the performance of few-shot classification tasks.

**Key words:** few-shot learning; negative transfer; distribution calibration; threshold; dynamic distribution calibration

## 0 引言

人类显著的一个标志是能够仅通过一个或几个示例就快速建立对新概念的认知能力,可以将先验知识

快速推广到仅包含少量具有监督信息样本的学习任务中。例如,儿童只需观察几张有关斑马与犀牛的图片就能轻松通过以前的认知经验识别出“斑马”与“犀

收稿日期:2023-01-07

修回日期:2023-04-12

基金项目:国家自然科学基金项目(62176210,61672027,U20B2050);陕西省教育厅重点实验室资助(18JS076)

作者简介:赵金伟(1974-),男,博士,副教授,CCF会员(32027M),研究方向为可解释的深度神经网络、智慧水利;通信作者:黑新宏(1976-),男,博士,教授,研究方向为轨道交通。

牛”。相比之下,大多数机器学习算法需要大量数据。在实际应用中,由于采集得到的数据往往难以标注或标注成本很高,所以只能得到大量未标记数据。这时模型容易陷入过拟合或者欠拟合的困境。由此可见,目前人工智能与人类智能仍存在巨大差距。

受到人类学习观点的启发,小样本学习(few-shot learning)的概念被提出,近几年得到广泛关注,在图像分类任务上已取得重大发展。如 Li 等人<sup>[1]</sup>使用贝叶斯学习框架求解单样本学习问题。人们又提出了许多基于元学习的解决方案,其核心思想是将从其它学习任务中得到的元模型迁移到新学习任务中。但是训练得到的元模型的结构往往过于复杂,如网络层数太深或者网络参数过多,模型训练过程十分困难。在这类方法中通常假设目标小样本学习任务的训练样本集的分布特征与辅助训练样本集的分布特征一致。

除了图像分类任务,在其他领域中也存在着小样本问题。在计算机视觉领域,利用小样本学习进行人脸识别<sup>[2]</sup>、手写字体识别<sup>[3]</sup>等。在自然语言处理领域,使用小样本方法实现对话系统<sup>[4]</sup>、口语理解<sup>[5]</sup>等。在生物与医学领域,可以应用于疾病诊断<sup>[6]</sup>、水量分析<sup>[7]</sup>等。

最近的工作试图利用有限数量的样本来评估样本的真实分布。因为样本特征空间的分布更具有可分性,所以这类方法尝试对小样本特征空间的分布进行校准,将有偏差的小样本分布向真实分布逼近<sup>[8]</sup>。他们假设特征向量的每个维度服从高斯分布,基于“相似的类别具有相似的分布和相似的属性值域”的思想提出将基类特征分布的统计量迁移到相似的小样本类中以校正其特征分布<sup>[9]</sup>。基于这种校准分布方法即可得到更精确的小样本特征分布,再通过随机采样即可增广出更多的样本。然而,该类方法直接迁移距离小样本均值中心最近的前  $k$  个基类的特征分布,且只考虑最近邻类分布特征在校正中的正作用,未考虑远域类分布特征的负作用,另外计算两类之间的相似度时未考虑样本特征的量纲差异,在分布校正阶段,更多地强调了基类的分布特征,忽略了新类本身的分布特征。这些问题极易引起知识的负迁移现象。

为此,该文提出了动态分布校正方法来解决上述问题,贡献如下:

(1)为了减弱基类特征分布统计量造成的负迁移现象,提出通过主动或自适应地基于阈值来选择近邻基类或远域基类的正向校正或负向校正,并主要研究基于阈值动态选择近邻基类。通过实验证明提出的动态筛选基类的方法比静态筛选方法更有利于实现分布特征的正迁移。

(2)因为嵌入特征量纲不同易导致嵌入特征分布

各向异性,从而造成对新类分布特征的负迁移。故该文提出对嵌入特征进行幂变换后,再对其进行标准化处理,以消除嵌入特征量纲之间的差异。并通过实验证明加入标准化操作后的分类精度要优于不加标准化的分类精度。

(3)引入惩罚参数来校正基类对新类嵌入特征分布的影响,并通过实验证明加入惩罚参数有利于实现分布特征的正迁移。

在常规的数据集 miniImageNet 和 CUB 上与其它最新方法的对比实验验证了提出的动态分布校正方法可以有效地提升小样本分类任务的性能。

## 1 相关工作

小样本学习是一个具有挑战性的机器学习问题。研究人员提出基于模型微调的方法解决小样本问题。该方法是小样本学习中较为传统的方法,通常在大规模数据集上预训练神经网络模型,在目标小样本数据集上对模型的全连接层或者顶端几层进行参数微调,得到微调后的模型。目前许多工作已经证明,与基线模型相比,加入微调可以有效提高 5-way-1-shot 任务的准确率。目前基于模型微调方法的研究取得了较好效果。Cai 等人<sup>[10]</sup>通过利用基于一阶 MAML 的元学习算法将微调集成到训练过程中并在微调期间对支持集进行数据增强。基于模型微调的方法较为简单,适用于目标域与源域数据分布较为类似的情况,但是在现实场景中该条件难以保证,采用模型微调的方法会导致在目标任务上出现欠拟合或过拟合现象。

基于数据增强方法是指借助辅助数据或辅助信息,对原有的小样本数据集进行数据扩充或特征增强。对于数据集数据增强主要是对输入数据的转换,旨在通过对数据进行少量修改来扩展现有数据来实现模型输入的多样性。Zhang 等人<sup>[11]</sup>提出当源类和目标类都只有有限数量的样本时,实现从互联网资源中抓取数据,并自动去除无关噪声,以实现可控的数据扩充。对于特征级数据增强,主要将像素信息映射到高维潜在空间,它比原始像素携带更多有效信息。如 Chu 等人<sup>[12]</sup>提出一种基于最大熵图块采样算法的强化学习模型来解决小样本学习问题。Zhang 等人<sup>[13]</sup>提出两种前景背景融合策略,将图像分割成前景与背景,选择原始图像、前景图像和背景图像重新缝合为新的视觉特征。Li 等人<sup>[14]</sup>提出可微分自动数据增强,可大幅度降低成本。Kang 等人<sup>[15]</sup>针对特殊数据缺失、连续长期数据缺失等问题,提出一种图矢量化算法与条件 Wasserstein 生成对抗网(CWGAN)相结合的方法实现数据增强。Schwartz 等人<sup>[16]</sup>基于一个经过修改的自动编码器,只需通过几个示例,就可以学习为一个看不

见的类别合成新样本。然而基于数据增强方法有可能引入许多噪声,对小样本分类的效果造成负面影响。

基于迁移学习的方法是当前较为前沿的方法,是指将旧知识迁移到一个新的领域中。基于迁移学习的方法又可分为基于度量学习、基于元学习的方法。度量学习也称为相似度学习,是指通过给定的距离函数计算两个样本之间的距离,从而度量它们的相似度。将度量学习的框架应用到小样本学习,通过计算待分类样本和已知分类样本之间的距离,找到邻近类别来确定待分类样本的分类结果。Koch 等人<sup>[17]</sup>最先提出使用孪生神经网络进行单样本图像识别。通过有监督的方式训练孪生网络,并将网络所提取的特征进行单样本或小样本学习。Vinyals 等人<sup>[18]</sup>继续就单样本学习问题进行深入探讨,提出了匹配网络,该网络为支持集和查询集构建不同的编码器,最终分类器的输出为支持集样本与查询集之间预测值的加权求和。为进一步深入解决小样本问题,Snell 等人<sup>[19]</sup>提出原型网络,该网络使用深度神经网络将图像映射成向量,对同属一个类别的样本,求得这一类样本向量的均值作为原型,并比较查询集与每个类别原型的欧氏距离。基于度量学习的方法最简单、容易操作,只需通过距离来衡量样本之间的相似度。但是学到的知识较少,并且在样本数量较少的情况下,简单地通过距离衡量相似度的方法准确率会有所降低。基于元学习比基于度量学习的方法学习能力更强,可以学到更多的知识。元学习是目前解决小样本学习的主流方法。元学习也叫做学会学习<sup>[20]</sup>,是机器学习领域一个前沿的研究框架,针对于解决模型如何学习的问题。元学习的目的是让模型获得一种学习能力,这种学习能力可以让模型自动学习到一些元知识。例如,UMTRA<sup>[21]</sup>、MAML<sup>[22]</sup>、Meta-SGD<sup>[23]</sup>、爬行动物<sup>[24]</sup>和 LEO<sup>[25]</sup>优化了模型,使其能够快速适应新任务。然而,基于元学习解决小样

本问题时时空复杂度较高,而且对于目标域与源域数据分布要求较高。

## 2 动态分布校正

### 2.1 问题设置

框架主要包括两个阶段。在第一阶段,利用预训练的模型提取样本特征,计算基类的统计信息。在第二阶段,为使特征分布更像高斯分布,使用幂阶变换操作保证目标任务中的支持集和查询集的特征接近于高斯分布,并加入特征向量标准化处理来消除特征量纲影响。为了避免“负迁移”,通过距离阈值动态挑选近邻基类与远域基类,并引入校正参数,校正新类的样本特征分布。最后利用校正后的分布生成特征与小样本任务的原始支持集特征一起用作特定任务分类器的训练数据。

遵循典型的小样本分类问题设置,给定由数据和标签对组成的样本集合  $D = \{(x_i, y_i)\}$ , 其中  $x_i \in R^d$  为样本的特征向量,其维度为  $d$ ,  $y_i \in C$  为样本的类别,  $C$  为类集合,类集合分为基类集  $C_b$  和新类集  $C_n$ , 且  $C_b \cap C_n = \emptyset$ ,  $C_b \cup C_n = C$ 。为了评估模型的泛化能力,每个新类只有少量可用的标记样本。该文构建  $N$ -way- $K$ -shot 任务,该任务有  $N$  个新类,每个新类有  $K$  个已标记样本构成支持集  $S = \{(x_i, y_i)\}_{i=1}^{N \times K}$ , 另外每个新类还有  $q$  个测试样本形成查询集  $Q = \{(x_i, y_i)\}_{i=1}^{N \times q}$ , 用于评估模型的性能。

整个过程如图 1 所示。新类样本分布的校正过程分为两个阶段。第一阶段为计算基类分布的统计量。第二阶段是动态分布校正阶段,计算基类与新类之间的嵌入空间的距离,基于阈值选择近邻基类与远域基类的分布特征,将其迁移到新类中对其分布进行校正,基于校正后的分布采样足够数量的样本来增广新类的训练样本集。

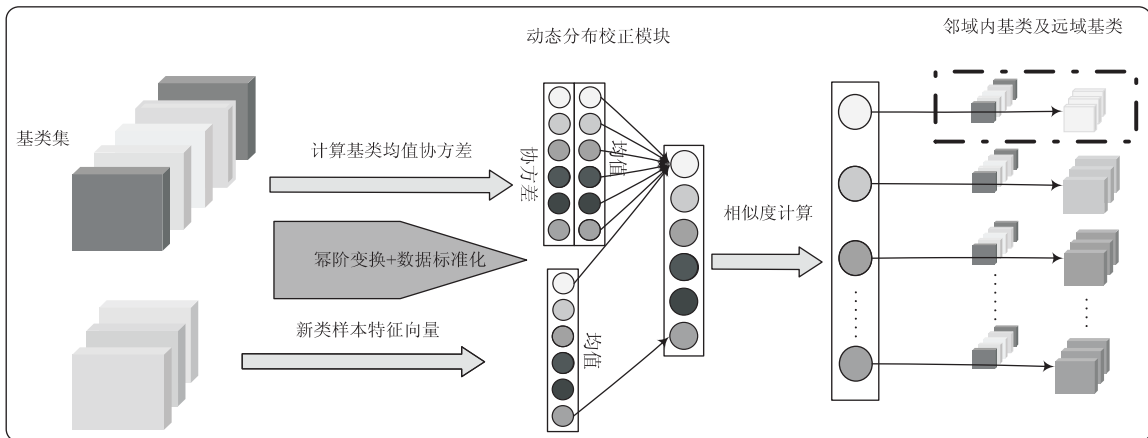


图 1 动态分布校正方法

### 2.2 基类的统计信息提取

假设基类的特征分布为高斯分布。基类  $i$  的特征

向量均值为该类所有数据的特征向量在每个维度上的均值。



$$\mu_i = \frac{\sum_{j=1}^{n_i} x_i^j}{n_i} \quad (1)$$

其中,  $x_i^j$  是基类  $i$  第  $j$  个样本的特征向量,  $n_i$  是基类  $i$  的样本总数。

由于特征向量  $x_i^j$  是多维向量, 所以通常使用协方差表示特征向量分布的分散程度。第  $i$  类样本特征的协方差矩阵  $\gamma_i$  为:

$$\gamma_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_i^j - \mu_i)(x_i^j - \mu_i)^T \quad (2)$$

## 2.3 校正新类分布

### 2.3.1 幂变换

为了实现嵌入空间不同类分布特征的正迁移, 除了将每个类的样本从原始样本空间映射到嵌入空间, 还需将嵌入空间的特征向量进行幂阶变换, 将每个类的支持集和查询集样本的嵌入空间分布变换为类高斯分布。该文基于单位方差投影给出类似于 Tukey 的幂阶变换 (Power Transform, PT), 其公式如下:

$$x' = \begin{cases} \frac{(x + \varepsilon)^\beta}{\|(x + \varepsilon)^\beta\|_2} & \text{if } \beta \neq 0 \\ \frac{\log(x + \varepsilon)}{\|\log(x + \varepsilon)\|_2} & \text{if } \beta = 0 \end{cases} \quad (3)$$

为了避免除数为零, 其中,  $\varepsilon = 1e - 6$ 。  $\beta$  为指数, 通过调整  $\beta$  来减少幂变换结果分布的扁平程度, 其中  $\beta = 1$  时为线性变换。单位方差投影将特征向量缩放到单位空间, 以进一步缩小相似类间的分布差异。

将经过幂变换的特征向量  $x'$  的每一维进行数据标准化处理来实现特征向量的各向同性。

$$\tilde{x} = \frac{x' - \bar{x}}{\sigma} \quad (4)$$

其中,  $\bar{x}$  和  $\sigma$  分别为特征向量  $x'$  的均值向量与方差向量。

### 2.3.2 基于分布特征迁移的新类分布校正

将2.2节提取出来的基类分布特征正迁移到新类, 以校准新类分布特征。首先计算基类  $i$  的嵌入特征中心与新类中第  $j$  个样本  $\tilde{x}_j$  之间的欧氏距离来估算它们之间的相似性。

$$s_j^i = \|\mu_i - \tilde{x}_j\|^2, i \in C_b \quad (5)$$

基于阈值 threshold 在以新类第  $j$  个样本的邻域中选择与该样本相似度最高的基类, 形成备选基类集合  $C_j^m$ , 其中  $m$  为  $C_j^m$  中的基类样本个数。

$$C_j^m = \{i | S_j^i < \text{threshold}\} \quad (6)$$

若  $\min(S_j^i) > \text{threshold}$ , 则基于公式选择距离新类最近的前  $k$  个基类形成备选基类集合  $C_j^k$ 。

$$C_j^k = \{i | i \in \text{topk}(C_b)\} \quad (7)$$

其中,  $\text{topk}(\cdot)$  是依据距离  $S_j^i$  对基类集  $C_b$  中的基类进

行升序排序, 然后选择前  $k$  个元素的函数。综上, 利用备选基类集合中的基类分布特征来校正新类分布, 如公式(8)(9)所示。

$$\mu_j' = \tilde{x}_j + \varphi \left( \frac{\sum_{i \in C_j^m} \mu_i}{m} - \tilde{x}_j \right) \text{ or } \tilde{x}_j + \varphi \left( \frac{\sum_{i \in C_j^k} \mu_i}{k} - \tilde{x}_j \right) \quad (8)$$

$$\gamma_j' = \frac{\sum_{i \in C_j^m} \gamma_i}{m} + a \text{ or } \frac{\sum_{i \in C_j^k} \gamma_i}{k} + a \quad (9)$$

其中,  $\varphi$  为对新类中心点进行校准的权值参数以表征邻域内基类的正迁移作用强度,  $a$  为新类分布中样本嵌入特征分散程度的偏置。

该文主要考虑基于阈值动态选择近邻基类, 利用近邻基类的统计信息校正新类的特征分布。但是考虑到远域基类对新类的特征分布的校正可能有重要的正迁移作用强度, 因此, 根据阈值 threshold 动态选择近邻基类的集合  $C_j^m$  的同时, 根据阈值  $t$  动态选择远域基类的集合  $C_j^t$  对新类样本特征进行校正。公式如下所示:

$$\mu_j' = \tilde{x}_j + b \times \left( \frac{\sum_{i \in C_j^m} \mu_i}{m} - \tilde{x}_j \right) + c \times \left( \frac{\sum_{i \in C_j^t} \mu_i}{t} - \tilde{x}_j \right) \quad (10)$$

$$\gamma_j' = n \times \frac{\sum_{i \in C_j^m} \gamma_i}{m} + v \times \frac{\sum_{i \in C_j^t} \gamma_i}{t} + e \quad (11)$$

其中,  $b, c \in [-1, 1]$ ,  $n, v \in [-1, 1]$  为对新类中心点进行校准的权值参数,  $b$  和  $n$  为正迁移作用的强度,  $c$  和  $v$  为负迁移作用强度,  $e \in [0, 1]$  为新类分布中样本特征分散程度的偏置。

### 2.3.3 数据增强

为了避免一个特定样本所提供的偏差并获得更多样化和更准确的新类分布特征估计, 基于新类支持集中的每个样本的嵌入特征向量都要进行动态分布校正。因此, 对于每个新类进行动态分布校正后, 会得到一组分布特征, 将校正后的分布表示为一组统计数据。对于类别  $y \in C_n$ , 将新类特征分布的统计数据表示为:

$$U_y = \{(\mu_1', \gamma_1'), (\mu_2', \gamma_2'), \dots, (\mu_k', \gamma_k')\}$$

其中,  $U_y$  是新类  $y$  分布校正后的均值和协方差集合,  $(\mu_i', \gamma_i')$  表示类别为  $y$  的支持集中对第  $j$  个样本嵌入特征, 再经过动态分布校正之后的嵌入特征分布均值和协方差,  $U_y$  中的  $k$  值代表类别为  $y$  的支持集中的样本个数。

对于类别  $y$  的目标任务, 利用其校正分布集合中的所有分布组合全部采样, 生成一组标签为  $D_y$  的特征向量:

$$D_y = \{(x, y) \mid x \sim N(\mu, \gamma), \forall (\mu, \gamma) \in U_y\}$$

其中,将每个类别生成的特征总数设置为超参数。将生成的样本特征与原始支持集特征一起作为任务分类器的训练数据,并通过最小化支持集特征和生成特征上的分类交叉熵损失来训练分类器。

### 3 实验

#### 3.1 实验设置

该文使用 miniImageNet、CUB 两种数据集,并利用查询集的准确率作为评估指标来衡量方法的性能。在 5-way-1-shot 和 5-way-5-shot 分类设置上进行评估实验,并且实验结果是 10 000 个任务的平均分类精度。

本章节使用文献[26]训练的 WideResNet 作为特征提取器,并分别使用 SVM 与 LR 两种分类器来验证所提方法的有效性。其中 Logistic Regression/SVM with DC (ours) 表示分类器为 Logistic Regression 或 SVM 的基于阈值选择近邻基类的方法,利用粒子群算法对动态分布校正算法的超参数进行寻优,两种数据集下最终得到的最优超参数为: miniImageNet: threshold = 3.3,  $\beta = 0.7$ ,  $\varphi = 0.2$ ,  $a = 0.1$ , 生成样本的个数为 850; CUB: threshold = 3.3,  $\beta = 0.8$ ,  $\varphi = 0.5$ ,  $a = 0.1$ , 生成样本的个数为 850。Logistic Regression/SVM with RDC (ours) 表示在分类器为 Logistic

Regression 或 SVM 的基于阈值同时选择近邻基类与远域基类的方法。同样采用粒子群算法获得最优超参数。miniImageNet: threshold = 2.5,  $t = 2$ ,  $\beta = 0.7$ ,  $b = 0.15$ ,  $c = 0.05$ ,  $n = 0.4$ ,  $v = 0.05$ ,  $e = 0.3$ , 生成样本的个数为 850; CUB: threshold = 2.5,  $t = 2$ ,  $\beta = 0.8$ ,  $b = 0.6$ ,  $c = 0.0$ ,  $n = 0.65$ ,  $v = 0.05$ ,  $e = 0.3$ , 生成样本的个数为 850。3.2 节中其他先进技术的具体实验细节设置引用于他们的来源论文。

#### 3.2 与最先进技术的性能对比

表 1 给出了动态分布校正算法与目前最先进的小样本学习技术的分类性能对比。从表中的结果发现,所提方法的分布校准性能远高于最先进的小样本分类方法,这证明所提方法能够更好地处理极低的样本的分类任务。从表 1 发现,所有分布校正方法在四个任务上的分类精度明显高于其它方法。说明分布校正方法在解决小样本分类问题上性能是很优秀的。表中将提出的两种动态分布校正算法与分布校正算法进行对比。两种数据集下的两种动态分布校正方法相比分布校正方法的分类精度都平均提升 2% 左右,并且所提方法的分类精度的方差更小。这些结果说明,提出的两种方法能进一步提升分布校正方法的性能,并且相对于分布校正算法,进一步说明所提动态筛选基类的方法比静态筛选方法更有利于实现分布特征的正迁移。

表 1 miniImagenet 和 CUB 数据集上的分类准确率 %

方法	miniImagenet		CUB	
	5-way-1-shot	5-way-5-shot	5-way-1-shot	5-way-5-shot
MAML <sup>[22]</sup>	48.70±1.84	63.10±0.92	50.45±0.97	59.60±0.84
Meta-SGD <sup>[23]</sup>	50.47±1.87	64.03±0.94	53.34±0.97	67.59±0.82
LEO <sup>[25]</sup>	61.76±0.08	77.59±0.12	—	—
Matching Net <sup>[18]</sup>	43.56±0.84	55.31±0.73	56.53±0.99	63.54±0.85
Prototypical Net <sup>[19]</sup>	54.16±0.82	73.68±0.65	72.99±0.88	86.64±0.51
Relation Net <sup>[27]</sup>	52.48±0.86	69.83±0.66	67.59±1.02	82.70±0.58
MetaGAN <sup>[23]</sup>	52.71±0.64	68.63±0.67	—	—
Delta-Encoder <sup>[16]</sup>	59.9	69.7	69.8	82.6
Logistic Regression with DC <sup>[8]</sup>	68.57±0.55	82.88±0.42	79.56±0.87	90.67±0.35
SVM with DC <sup>[8]</sup>	67.31±0.83	82.30±0.34	79.49±0.33	90.26±0.98
Logistic Regression with DC (ours)	69.70±0.46	85.46±0.18	83.15±0.02	92.32±0.05
SVM with DC (ours)	69.38±0.37	69.38±0.37	83.12±0.02	92.25±0.16
Logistic Regression with RDC (ours)	70.03±0.01	85.62±0.01	83.19±0.01	92.27±0.06
SVM with RDC (ours)	69.98±0.03	85.08±0.21	83.05±0.07	90.98±1.09

将 Logistic Regression/SVM with RDC 与 Logistic Regression/SVM with DC 相比较。对于两种数据集的 5-way-1-shot 分类任务,从表 1 中可以观测到 Logistic Regression/SVM with RDC 算法的分类精度比 Logistic Regression/SVM with DC 算法平均提升了 1%

左右。在两种数据集的 5-way-5-shot 分类任务上, Logistic Regression/SVM with RDC 算法的分类精度比 Logistic Regression/SVM with DC 算法平均降低了 1% 左右。说明当小样本分类任务的原始样本较多,样本分布多样性较强时,远离新类的基类分布特征会带来

负迁移。相反,当小样本分类任务的原始样本较少,样本分布多样性较弱时,远离新类的基类分布特征会增强原始样本分布的多样性,从而带来正迁移。

### 3.3 超参数对算法性能的影响

将基于阈值动态选择近邻基类(Logistic Regression/SVM with DC(ours))作为主要的研究方法。在两个数据集的 5-way-1-shot 分类设置且分类器选为 LR 的情况下,研究超参数的设置对该方法分类性能的影响。

图 2 显示了  $k$  值固定的情况下阈值 threshold 的变化对分类准确率的影响。从图 2 中的结果发现,当  $k \in \{0,1,2,3\}$  时,随着阈值 threshold 的不断增大,准确率呈不断增大的趋势。且对于两种数据集,在阈值 threshold = 3.3 时准确率都达到最高值。其原因为,随

着阈值的不断增大,用于校正新类特征分布的基类个数不断增多,当阈值达到 3.3 时,由该阈值确定的邻域内的基类统计信息迁移到新类中能较好地校正新类的特征分布。随之阈值的不断增大,邻域内会引入一些对新类分布产生负迁移现象的基类,从而造成测试准确率下降的状态。当  $k \in \{0,1,2,3\}$  时,曲线处于重合状态,说明不考虑  $k$  时仅通过 threshold 选择基类可以获得较好的小样本分类性能。

从图 3 中的结果观测到,对于 miniImagenet 数据集(左图)和 CUB 数据集(右图),随着  $\varphi$  值不断增大,平均分类准确率(Test Accuracy)呈现先稳定再下降的趋势。曲线变化的主要原因在于,随着  $\varphi$  不断增大,新类的特征分布受基类影响程度也不断增大,但当  $\varphi$  取值太大,又会产生负迁移,造成准确率下降。

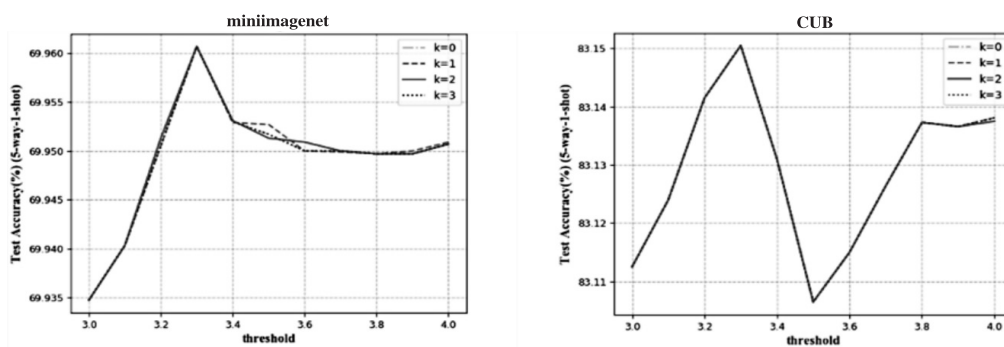


图 2 threshold 的取值对实验结果的影响

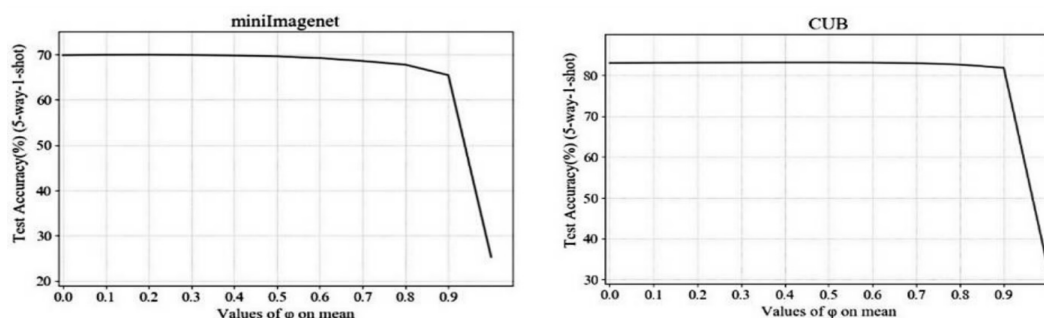


图 3  $\varphi$  的取值对实验结果的影响

公式(9)中的  $a$  确定从校正后的分布中样本嵌入特征分散的程度。图 4 显示了不同的  $a$  取值对测试准确率的影响。随着  $a$  的增大,测试精度快速上升达到最大值后几乎保持稳定。其主要原因在于,当  $a$  的

等于 0.1 的时候,校正的新类协方差是逼近它的真实样本嵌入特征分布。 $a$  随后继续增大,校正的新类协方差又偏离了真实样本嵌入特征分布,所以预测精度会有所降低。

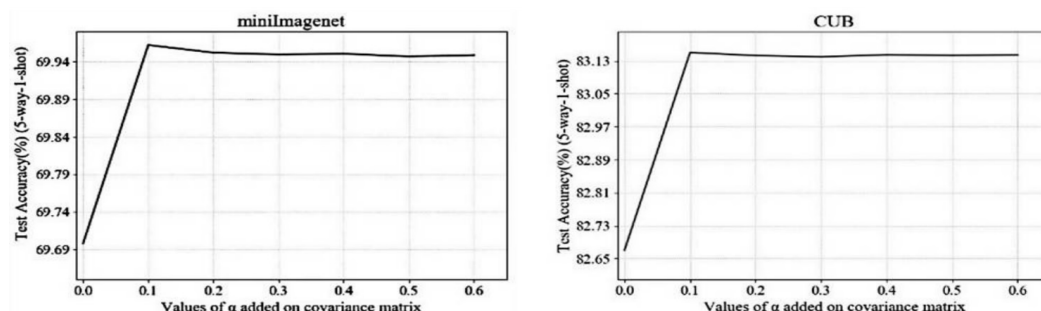


图 4  $a$  的取值对实验结果的影响

### 3.4 消融实验

为研究提出的动态分布校正方法(Logistic Regression/SVM with DC(ours))中每一计算过程对小样本分类性能的影响,进行了消融实验。表2展示了在训练模型时未基于阈值选择近邻基类、在训练模型时未进行幂变换(Power Transform, TP)与数据标准化(Data Standardization, DS)而是使用Tukey的幂阶变换以及在训练模型时未引入校正参数处理时的分类结果。由表2知,通过TP与DS操作使小样本的分类精度提高接近1%,引入校正参数后的分类精度提高2%左右,基于阈值选择近邻基类对于分类精度提高0.1%。由此可知,三步过程对解决分布校正算法负迁移问题起着积极作用。

表2 在 miniImagenet 数据集上的分类准确率 %

threshold	TP&DS	$\varphi$	miniImagenet	
			5-way-1-shot	5-way-5-shot
×	×	×	68.57±0.55	82.88±0.42
×	✓	✓	69.69±0.03	85.41±0.04
✓	×	✓	68.59±0.03	84.08±0.03
✓	✓	×	67.96±0.97	84.21±0.72
✓	✓	✓	69.70±0.46	85.46±0.18

为了研究幂变换之后的数据标准化处理对小样本分类性能的影响,在 miniImageNet 数据集上进行了消融实验。实验的结果为10个任务的测试集平均分类精度。由表3中的实验结果发现,在幂变换之后加入数据标准化处理,分类准确率提高接近1%。实验证明数据标准化处理对于分类性能同样能够起着积极作用。

表3 在 miniImagenet 数据集上的分类准确率 %

TP	DS	miniImagenet
		5-way-1-shot
×	×	70.53
✓	×	71.06
×	✓	72.53
✓	✓	73.20

## 4 结束语

基于分布校正方法,提出了一种动态分布校正方法来解决分布校正过程中的负迁移问题。具体来说,首先,通过幂变换使新类的嵌入特征分布近似于高斯分布,再对新类嵌入特征进行数据标准化处理来消除由特征量纲不同造成的负迁移现象;其次,基于阈值提取了近邻基类与远域基类嵌入特征的分布统计量后,迁移到新类中来校正新类的嵌入特征分布;最后,基于校正后分布扩增新类的嵌入特征数据来提高预测模型的性能。在 CUB 和 miniImageNet 两个数据集上

评估了所提方法,实验证明所提方法得到的预测模型分类性能要高于目前最先进的小样本学习方法。在未来的工作中,会探索更高效的小样本学习方法。

### 参考文献:

- [1] LI F F, FERGUS R, PERONA P. A bayesian approach to unsupervised one-shot learning of object categories [C]//Proceedings ninth IEEE international conference on computer vision. Nice: IEEE, 2003: 1134-1141.
- [2] 黄振文, 谢凯, 文畅, 等. 迁移学习模型下的小样本人脸识别算法[J]. 长江大学学报, 2019, 16(7): 88-94.
- [3] 周添一, 赵磊. 基于神经网络的小样本手写汉字识别[J]. 山东理工大学学报, 2019, 33(3): 69-74.
- [4] 刘继明, 孟亚磊, 万晓榆. 基于小样本机器学习的跨任务对话系统[J]. 重庆邮电大学学报, 2019, 31(3): 299-304.
- [5] UPADHYAY S, FARUQUI M, TUR G, et al. (Almost) Zero-shot cross-lingual spoken language understanding [C]//2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). Calgary: IEEE, 2018: 6034-6038.
- [6] 赵媛. 基于卷积神经网络的小样本颈动脉超声斑块图像识别方法研究[D]. 武汉: 华中科技大学, 2018.
- [7] 王雪, 马铁民, 杨涛, 等. 基于近红外光谱的灌浆期玉米籽粒水分小样本定量分析[J]. 农业工程学报, 2018, 34(13): 203-210.
- [8] YANG S, LIU L, XU M. Free lunch for few-shot learning: distribution calibration [C]//International conference on learning representations. [s. l.]: ICLR, 2020: 1-13.
- [9] SALAKHUTDINOV R, TENENBAUM J, TORRALBA A. One-shot learning with a hierarchical nonparametric Bayesian model [C]//UTLW'11 proceedings of the 2011 international conference on unsupervised and transfer learning workshop. Washington: JMLR.org, 2011: 195-207.
- [10] CAI J, SHEN S M. Cross-domain few-shot learning with meta fine-tuning [C]//CVPR 2020 workshop on visual learning with Limited Labels (VL3). Seattle: IEEE, 2020: 1-4.
- [11] ZHANG J, ZHANG M, LU Z, et al. AdarGCN: adaptive aggregation GCN for few-shot learning [C]//2021 IEEE winter conference on applications of computer vision (WACV). Waikoloa: IEEE, 2021: 3482-3491.
- [12] CHU W H, LI Y J, CHANG J C, et al. Spot and learn: a maximum-entropy patch sampler for few-shot image classification [C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach: IEEE, 2019: 6251-6260.
- [13] ZHANG H, ZHANG J, KONIUSZ P. Few-shot learning via saliency-guided hallucination of samples [C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach: IEEE, 2019: 2770-2779.



- [14] LI Y, HU G, HOSPEDALES T, et al. DADA: differentiable automatic data augmentation [C]//Computer vision - ECCV 2020; 16th European conference. Berlin: Springer-Verlag, 2020; 580-595.
- [15] KANG M, YANG Y, CHEN D, et al. CWGAN: a graph vector based traffic missing data adversarial generation approach[C]//2020 Chinese automation congress (CAC). Shanghai: IEEE, 2020; 6234-6238.
- [16] SCHWARTZ E, KARLINSKY L, SHTOK J, et al. Delta-encoder: an effective sample synthesis method for few-shot object recognition [C]//NIPS '18: proceedings of the 32nd international conference on neural information processing systems. Red Hook: Curran Associates Inc., 2018; 2850-2860.
- [17] KOCH G, ZEMEL R, SALAKHUTDINOV R. Siamese neural networks for one-shot image recognition [D]. Toronto: University of Toronto, 2015.
- [18] VINYALS O, BLUNDELL C, LILICRAP T, et al. Matching networks for one shot learning [C]//NIPS '16: proceedings of the 30th international conference on neural information processing systems. Red Hook: Curran Associates Inc., 2016; 3637-3645.
- [19] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning [C]//Proceedings of the 31st international conference on neural information processing systems. Red Hook: Curran Associates Inc., 2017; 4080-4090.
- [20] THRUN S, PRATT L. Learning to learn: introduction and overview[M]. [s. l.]: Kluwer Academic Publishers, 1998.
- [21] KHODADADEH S, BÖLÖNI L, SHAH M. Unsupervised meta-learning for few-shot image classification [C]//Proceedings of the 33rd international conference on neural information processing systems. [s. l.]: Elsevier Science Inc., 2019; 10132-10142.
- [22] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C]//ICML '17: proceedings of the 34th international conference on machine learning - volume 70. Sydney: PMLR, 2017; 1126-1135.
- [23] LI Z, ZHOU F, FEI C, et al. Meta-SGD: learning to learn quickly for few-shot learning [J]. arXiv: 1707.09835v2, 2017.
- [24] NICHOL A, ACHIAM J, SCHULMAN J. On first-order meta-learning algorithms[J]. arXiv: 1803.02999v2, 2018.
- [25] RUSU A A, RAO D, SYGNOWSKI J, et al. Meta-learning with latent embedding optimization [C]//International conference on learning representations. [s. l.]: ICLR, 2018; 1-15.
- [26] MANGLA P, KUMARI N, SINHA A, et al. Charting the right manifold: manifold mixup for few-shot learning [C]//2020 IEEE winter conference on applications of computer vision (WACV). Snowmass: IEEE, 2020; 2218-2227.
- [27] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: relation network for few-shot learning [C]//2018 IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018; 1199-1208.

(上接第172页)

- [2] ADAMS S, CODY T, BELING P A. A survey of inverse reinforcement learning [J]. Artificial Intelligence Review, 2022, 55; 4307-4346.
- [3] SUTTON R S, PRECUP D, SINGH S P. Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning [J]. Artificial Intelligence, 1999, 112 (1/2); 181-211.
- [4] PARR R. Hierarchical control and learning for Markov decision processes [D]. Berkeley: University of California, 1998.
- [5] HENGST B. Discovering hierarchical reinforcement learning [D]. Sydney: University of New South Wales, 2003.
- [6] DIETTERICH T G. Hierarchical reinforcement learning with the MAXQ value function decomposition [J]. Journal of Artificial Intelligence Research, 2000, 13; 227-303.
- [7] HASHEMZADEH M, HOSSEINI R, AHMADABADI M N. Clustering subspace generalization to obtain faster reinforcement learning [J]. Evolving Systems, 2020, 11; 89-103.
- [8] DAYAN P, HINTON G E. Feudal reinforcement learning [J]. Advances in Neural Information Processing Systems, 1993, 5; 271-278.
- [9] SHI W, FENG Y H, HUANG H L, et al. Efficient hierarchical policy network with fuzzy rules [J]. International Journal of Machine Learning and Cybernetics, 2022, 13; 447-459.
- [10] CAI Q, HANG W, MIRHOSEINI A, et al. Reinforcement learning driven heuristic optimization [J]. arXiv: 1906.06639, 2019.
- [11] 赖俊, 魏竞毅, 陈希亮. 分层强化学习综述 [J]. 计算机工程与应用, 2021, 57(3); 72-79.
- [12] NEFTCI E O, AVERBECK B B. Reinforcement learning in artificial and biological systems [J]. Nature Machine Intelligence, 2019, 1; 133-143.
- [13] ASCIONE G, CUOMO S. A sojourn-based approach to semi-markov reinforcement learning [J]. Journal of Scientific Computing, 2022, 92(2); 36.
- [14] GRZES M, KUDENKO D. Online learning of shaping rewards in reinforcement learning [J]. Neural Networks, 2010, 23(4); 541-550.
- [15] NG A Y, HARADA D, RUSSELL S. Policy invariance under reward transformations: theory and application to reward shaping [C]//Proceedings of the sixteenth international conference on machine learning. Bled: ACM, 1999; 278-287.
- [16] 高文斌, 王睿, 王田丰, 等. 基于深度强化学习的 QoS 感知 Web 服务组合 [J]. 计算机技术与发展, 2022, 32(6); 92-98.