



# 计算机科学

COMPUTER SCIENCE

## 基于引领森林的多粒度广义长尾分类

杨金业, 徐计, 王国胤

### 引用本文

杨金业, 徐计, 王国胤. 基于引领森林的多粒度广义长尾分类[J]. 计算机科学, 2024, 51(11): 229-238.

YANG Jinye, XU Ji, WANG Guoyin. [Multi-granular and Generalized Long-tailed Classification Based on Leading Forest](#) [J]. Computer Science, 2024, 51(11): 229-238.

---

### 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)**

#### [基于不变图卷积神经网络的文本分类](#)

Text Classification Based on Invariant Graph Convolutional Neural Networks

计算机科学, 2024, 51(6A): 230900018-5. <https://doi.org/10.11896/jsjcx.230900018>

#### [代价敏感的多粒度邻域粗糙模糊集的近似表示](#)

Cost-sensitive Multigranulation Approximation of Neighborhood Rough Fuzzy Sets

计算机科学, 2023, 50(5): 137-145. <https://doi.org/10.11896/jsjcx.220500268>

#### [结合多粒度特征融合的自然场景文本检测方法](#)

Natural Scene Text Detection Algorithm Combining Multi-granularity Feature Fusion

计算机科学, 2021, 48(12): 243-248. <https://doi.org/10.11896/jsjcx.201000154>

#### [交叉融合的“人工智能+”学科建设探索与实践](#)

Practical Exploration of Discipline Construction of Artificial Intelligence+

计算机科学, 2020, 47(4): 1-5. <https://doi.org/10.11896/jsjcx.200300144>

#### [基于边界域的邻域知识距离度量模型](#)

Neighborhood Knowledge Distance Measure Model Based on Boundary Regions

计算机科学, 2020, 47(3): 61-66. <https://doi.org/10.11896/jsjcx.190500174>

# 基于引领森林的多粒度广义长尾分类

杨金业<sup>1</sup> 徐计<sup>1</sup> 王国胤<sup>2</sup>

1 贵州大学省部共建公共大数据国家重点实验室 贵阳 550025

2 重庆邮电大学计算智能重庆市重点实验室 重庆 400065

(gs.jinyeyang21@gzu.edu.cn)

**摘要** 长尾分类在现实世界中是一项不可避免且充满挑战的任务。传统方法通常只专注于类间的不平衡分布,然而近期的研究开始重视类内的长尾分布,即同一类别内,具有头部属性的样本远多于尾部属性的样本。由于属性的隐含性和其组合的复杂性,类内不平衡问题更加难以处理。为此,文中提出一种基于引领森林并使用多中心损失的广义长尾分类框架(Cognisance),旨在通过不变性特征学习的范式建立长尾分类问题的多粒度联合求解模型。首先,该框架通过无监督学习构建粗粒度引领森林(Coarse-Grained Leading Forest, CLF),以更好地表征类内关于不同属性的样本分布,进而在不变风险最小化的过程中构建不同的环境。其次,设计了一种新的度量学习损失,即多中心损失(Multi-Center Loss, MCL),可在特征学习过程中逐步消除混淆属性。同时, Cognisance 不依赖于特定模型结构,可作为独立组件与其他长尾分类方法集成。在 ImageNet-GLT 和 MSCOCO-GLT 数据集上的实验结果显示,所提框架取得了最佳性能,现有方法通过与本框架集成,在 Top1-Accuracy 指标上均获得 2%~8% 的提升。

**关键词:** 长尾分类; 不平衡学习; 不变性特征学习; 多粒度联合求解

**中图分类号** TP391

## Multi-granular and Generalized Long-tailed Classification Based on Leading Forest

YANG Jinye<sup>1</sup>, XU Ji<sup>1</sup> and WANG Guoyin<sup>2</sup>

1 State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China

2 Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

**Abstract** Long-tailed classification is an inevitable and challenging task in the real world. Traditional methods usually focus only on inter-class imbalanced distributions, however, recent studies have begun to emphasize intra-class long-tailed distributions, i. e., within the same class, there are far more samples with head attributes than tail ones. Due to the implicitness of the attributes and the complexity of their combinations, the intra-class imbalance problem is even more difficult to deal with. For this purpose, a generalized long-tailed classification framework (Cognisance) is proposed in the paper, aiming to build a multi-granularity joint solution model for the long-tailed classification problem through the invariant feature learning. Firstly, the framework constructs coarse-grained leading forest (CLF) through unsupervised learning to better characterize the distribution of samples about different attributes within the class, and thus constructs different environments in the process of invariant risk minimization. Secondly, the framework designs a new metric learning loss, multi-center loss (MCL), to gradually eliminate confusing attributes during the feature learning process. Additionally, the framework does not depend on a specific model structure and can be integrated with other long-tailed classification methods as an independent component. Experimental results on datasets ImageNet-GLT and MSCOCO-GLT show that, the proposed method achieves the best performance, and existing methods all gain an improvement of 2%~8% in Top1-Accuracy metric by integrating with this framework.

**Keywords** Long-tailed classification, Imbalance learning, Invariant feature learning, Multi-granularity joint problem solving

## 1 引言

在现实世界的应用中,训练样本通常表现出长尾分布,特别是大规模数据集<sup>[1-2]</sup>。长尾分布意味着少数头部类别包含

大量样本,而大多数尾部类别的样本数量相对有限。这种不平衡可能导致传统的分类模型产生偏差,即算法更倾向于处理头部类别,而在处理尾部类别时表现不佳。解决长尾分类问题至关重要,因为尾部类别可能包含重要信息,如罕见

到稿日期:2023-11-19 返修日期:2024-05-09

基金项目:国家自然科学基金(61966005,62221005,62366008)

This work was supported by the National Natural Science Foundation of China(61966005,62221005,62366008).

通信作者:徐计(jixu@gzu.edu.cn)

疾病、关键事件或少数群体的特征<sup>[3-6]</sup>。为了有效应对这一挑战,研究人员提出了各种方法,但是通过调研我们发现,真正困扰工业界的长尾分布问题并不全是目前研究最多的类间长尾问题,类内长尾问题也很重要,例如,当谈到无人车训练数据中的长尾分布时,天气分布、昼夜分布<sup>[7]</sup>等都不是模型预测的目标。因此,这里的长尾不是在类别之间,而是在属性之间。其中,属性表示引起类内变化的所有因素,包括对象级属性(如特定车型、品牌、颜色等)和图像级属性(如照明、气候条件等)。这种分类问题被命名为广义长尾分类(Generalized

Long-Tailed Classification, GLT)<sup>[8]</sup>。

从图1中可以很明显地看到,不同类别之间的样本数量存在显著差异,尤其是“海岸”等头部类别,其样本数量远多于“自动扶梯”等尾部类别。然而,即使在一个类别中,不同属性对应的样本数量也存在显著差异。例如,在“海岸”类别中,白天的样本数量远多于夜间的样本数量;晴天的样本数量也远多于阴天的样本数量。类似地,即使在尾部类别“自动扶梯”中,“阶梯式扶梯”中的样本数量也多于“坡道式扶梯”中的样本数量。

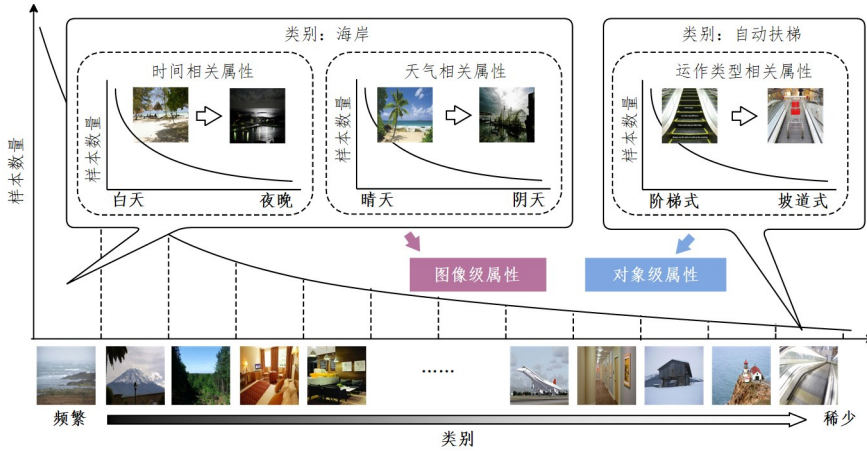


图1 类间长尾分布和类内长尾分布

Fig. 1 Inter-class and intra-class long-tailed distributions

与类别间不平衡相比,相同类别内属性之间的样本数量的不平衡从根本上更难避免,也更难以处理,因为属性对于大多数数据集来说都是隐含的,并且属性组合也较为复杂。如图2所示,这种属性间的不平衡从两个方面破坏了模型的性能<sup>[8]</sup>。首先,它削弱了传统算法在具有尾部属性的图像上的准确性。例如在“天鹅”类别中,“黑天鹅”比“白天鹅”更容易被错误分类,尽管它们都属于同一类别。此外,这会导致一些属性被错误地与某些类别关联在一起,例如“白色”可能与“天鹅”错误地联系在一起。因此,当“白色”这种混淆属性出现在其他鸟类的图像(例如“公鸡”)中时,样本被错误归类为“天鹅”的风险就很高。

数据不平衡不影响特征的学习,因此通过解耦特征学习和分类器学习来将模型的训练分为两个阶段<sup>[13-14]</sup>。然而,这种调整只是准确率和精确率之间的权衡,模型所学习到的特征空间中伪属性所对应的混淆区域并没有发生改变<sup>[8,15]</sup>。如前文所述,正是一个类别中属性之间的长尾导致了某些类别的头部伪属性与该类别之间产生了虚假相关性。本文称这些属性为混淆属性,这意味着这些属性对应于类别的虚假特征(即特征空间中的混淆区域)。本文提出的方法旨在基于不变性特征学习来消除混淆区域,最终达到同时提高模型的准确率和精确率的目的<sup>[16-18]</sup>。



图2 “白色”属性与“天鹅”类别的虚假相关性

Fig. 2 Spurious correlation between attribute “white” and class “swan”

目前的研究主要集中在解决类间长尾的问题上,在处理不平衡数据时,通常使用重采样<sup>[9-10]</sup>或损失重加权<sup>[11-12]</sup>方法,旨在重新平衡训练过程。然而,这些方法大多以牺牲头部类别性能为代价来提高尾部类别的性能,这就像玩性能跷跷板,很难从根本上提高所有类的分类性能。此外,一些方法认为

在本文提出的框架中,通过构建不同的环境来对类别间和类别内两个粒度级别上的长尾问题进行联合求解。其中,我们无监督地构建一种新颖的粗粒度引领森林 CLF,可以对类内的属性分布进行刻画,并可以根据属性的分布,在不同的环境中灵活地构建不同的采样策略。如本文在实验设置中构建了两个环境:一个是没有经过特殊处理的原始环境,即在训练过程中对训练样本进行独立同分布采样;另一个环境的采样策略是类别和属性分布均趋于平衡。最后,为了逐步消除训练过程中易产生混淆的伪特征,我们设计了一种新的度量学习损失,称为多中心损失(Multi-Center Loss, MCL)。这个想法受到了文献<sup>[8]</sup>和文献<sup>[19]</sup>的启发,将中心损失扩展到其不变风险最小化(Invariant Risk Minimization, IRM)版本,相较于前两者,进一步提高了算法的鲁棒性,使模型具备了学习不变性特征的能力。此外,本框架不与特定的主干模型或损失函数耦合,可以无缝地集成到其他长尾分类方法中,在原有方法的基础上进一步提高了性能。本文的贡献可以总结如下:

1)设计了一种新的基于无监督学习的采样方案来指导IRM过程中不同环境的采样,以此来同时处理类间长尾问题以及传统方法中易被忽视的类内长尾问题。

2)结合不变性特征学习的思想,设计了一种新的度量学习损失,可以使模型在训练过程中逐渐消除伪特征,进一步提高模型的鲁棒性,同时提高模型的准确率和精确率。

3)在ImageNet-GLT和MSCOCO-GLT这两个现有基准数据集上进行了广泛的实验,验证了该框架的有效性。同时通过与本框架的集成,当前的SOTA方法的Top1-Accuracy指标在原有基础上得到了2%~8%的提升。

## 2 相关工作

### 2.1 长尾分类

长尾分类的关键挑战是如何有效地处理数据分布的不平衡,以确保分类器在头部和尾部之间都能获得优异的性能。目前对长尾分类的处理方法大致可分为3类<sup>[1]</sup>。

1)重平衡。这是长尾学习的主流范式,其目标是通过在模型训练过程中重新采样<sup>[10,20-22]</sup>、重新加权<sup>[23-27]</sup>或调整logit值<sup>[28]</sup>来增强尾部样本对模型的影响。其中,也有一些方法认为不平衡的数据集不影响特征的学习,因此将特征学习和分类器学习分为两个阶段<sup>[13-14]</sup>,仅在分类器学习阶段执行重采样等操作。

2)信息增强。基于信息增强的方法试图在模型训练中引入额外的信息,以提高长尾学习中的模型性能。这种方法有两个主要类型:迁移学习<sup>[29-31]</sup>和数据扩充<sup>[32-34]</sup>。迁移学习旨在利用从其他任务学到的知识,帮助解决长尾分类问题;数据扩充通过合成新的训练样本,来增加模型的训练数据,从而改善性能。

3)模块改进。除了重平衡和信息增强外,研究人员还研究了改进长尾学习中的网络模块的方法。例如,RIDE<sup>[35]</sup>和TADE<sup>[36]</sup>这两种模型采用了一种集成学习方法,通过在网络中引入多个专家模型来处理独立于测试分布的长尾识别问题。

此外,最近的一项研究提出了广义长尾分类GLT的概念<sup>[8]</sup>。该概念首次关注了类内属性的长尾问题,指出传统的长尾分类方法将分类模型表示为 $p(Y|X)$ ,即从输入图像 $X$ 预测标签 $Y$ 。这可以进一步分解为 $p(Y|X) \propto p(X|Y) \cdot p(Y)$ ,其将类别不平衡归因于 $p(Y)$ 。然而,在实际应用中 $p(X|Y)$ 的分布在不同的领域中也可能会发生变化。因此,在这项研究中,基于贝叶斯定理,他们将分类模型进行扩展,如式(1)所示,从而更全面地研究长尾分类问题。

$$p(Y=k|z_c, z_a) = \frac{p(z_c|Y=k)}{p(z_c)} \cdot \underbrace{\frac{p(z_a|Y=k, z_c)}{p(z_a|z_c)}}_{\text{attribute bias}} \cdot \underbrace{p(Y=k)}_{\text{class bias}} \quad (1)$$

其中, $z_c$ 是该类别中不变属性所对应的表征,而 $z_a$ 为与伪属性相关的表征,即不变部分 $z_c$ 仅依赖于 $Y$ ,而 $z_a$ 会随领域的变化而变化。以图2中的“天鹅”为例,“天鹅”的“颜色”属性属于 $z_a$ ,而“天鹅”诸如“形状”等属性属于 $z_c$ 。值得注意的是,在实际应用中,该公式并未强加表征可分离的假设,也就

是说,它并没有假设我们能够获得完美的特征向量 $z = [z_c; z_a]$ ,即并没有要求 $z_c$ 和 $z_a$ 完全分离。

### 2.2 不变风险最小化

不变风险最小化IRM<sup>[17]</sup>由Arjovsky等于2019年提出,其主要目标是构建在不同数据分布上具有相同性能的鲁棒学习模型。在机器学习中,我们通常希望经过训练的模型能够在未来的数据上表现良好,这就是风险最小化。然而,在实际应用中,训练数据和测试数据之间可能存在分布差异,即域偏移,这导致模型在新域上表现不佳。IRM的核心思想是通过鼓励模型学习那些在不同数据领域中保持不变的特征来解决领域自适应问题。这意味着模型应该专注于所有数据领域中共享的特征,而不是过度拟合特定的数据分布。IRM的优化目标为:

$$\begin{aligned} \min_{\Phi, X \rightarrow H, \omega; H \rightarrow Y} \sum_{e \in \epsilon} R^e(\omega \circ \Phi) \\ \text{s. t. } \omega \in \arg \min_{\omega: H \rightarrow Y} R^e(\bar{\omega} \circ \Phi), \forall e \in \epsilon \end{aligned} \quad (2)$$

其中, $\epsilon$ 表示所有环境,不同环境有着不同的数据分布; $X, H$ 和 $Y$ 分别表示输入、特征表示和预测结果; $\Phi$ 和 $\omega$ 分别是特征学习器和分类器; $R^e$ 表示环境 $e$ 下的风险。IRM的目标是找到一个在所有环境中都能稳定运行的通用解决方案,从而提高模型的泛化能力。

### 2.3 最优引领森林

本文提出的CLF是基于最优引领森林(Optimal Leading Forest, OLeaF)<sup>[37]</sup>进行设计的,因此在这里简要介绍它的思想和算法。引领森林的概念源于一种基于密度峰值的聚类方法<sup>[38]</sup>,而构建引领森林最关键的两个因素是数据点的密度和数据点与密度更高的近邻的距离。设 $I = \{1, 2, \dots, N\}$ 是数据集 $X$ 的索引集; $d_{i,j}$ 表示数据点 $x_i$ 和 $x_j$ 之间的距离(可以使用任意距离度量); $\rho_i$ 为数据点 $x_i$ 的局部密度; $d_c$ 为截断距离。如果存在 $\xi_i = \arg \min_j \{d_{i,j} | \rho_j > \rho_i\}$ ,则 $x_{\xi_i}$ 被称为数据点 $x_i$ 的引领节点。基于此,可以建立偏序关系,即,如果存在 $\xi_i = \eta(x_i)$ ,则 $x_i < x_{\xi_i}$ 。连接每一个 $x_i$ 和 $x_{\xi_i}$ (如果 $r = \arg \max_{1 \leq i \leq N} \{\rho_i\}$ ,则 $x_r$ 是根节点,无须寻找引领节点),可以建立一个树结构,称其为引领树。然后令 $\delta_i = d_{\xi_i, i}$ , $\gamma_i = \rho_i \times \delta_i$ ,则较大的 $\gamma_i$ 表示数据点 $x_i$ 被选为簇中心的潜力较大。直观地,如果某数据点 $x_i$ 有一个大的 $\rho_i$ ,这意味着它有很多近邻。而 $\delta_i$ 很大,说明它离另一个具有更大 $\rho$ 值的数据点很远,所以 $x_i$ 有较大机会成为簇的中心。然后可以根据 $\gamma_i$ 的排序来选择根节点。最后,将一棵引领树以这种方式进行分割,由此产生的多棵引领树被称为最优引领森林。

## 3 基于CLF和MCL的Cognisance框架

之前的GLT<sup>[8]</sup>引入了属性偏差和类偏差,传统的长尾分类方法往往只关注后者;基于此设计重采样、重加权等方法也往往只关注类的分布。事实上,正如前面给出的例子一样,属性偏差不仅会降低模型在类内尾部属性样本上的性能,而且正是某些头部伪属性与相应类之间的虚假相关性导致了模型在类之间的性能不平衡。然而,当前的GLT研究虽然已经注意到了属性的偏差问题,但尚未采用有效方法来充分描述属性的长尾分布。因此,在后续处理阶段,难以以可控的方式

对具有这些属性的样本进行重采样或其他相关操作。为了提高模型对不同类别分布和不同属性分布的数据的泛化能力,本文提出了一个基于不变性特征学习的框架 Cognisance。该框架首先使用粗粒度引领森林来构建不同属性分布的不同环境;然后在多中心损失的优化下,赋予模型学习每个数据域中的不变性特征的能力,而不是过度拟合某个分布,以此来解决多域自适应问题。

### 3.1 粗粒度引领森林

在 IRM<sup>[17]</sup> 的概念中,构建不同的环境是训练的先决条件,而本文的挑战是如何可控地构建具有不同属性分布的环境。首先,具有不同类别分布的环境很容易构建,因为类别的标签在数据集中是显式的,而大多数数据集的属性是隐式的,即使类别不平衡被完全消除,其属性失衡仍然存在。同时,由于属性可以连续叠加和组合,属性的边界也很复杂,因此本文设计了一种基于无监督学习的采样方法。该方法可以刻画同一类别内属性的分布,并可以根据超参数的设置控制属性刻画的精细度。

本文的动机是基于一个合理的假设,即同一类别内样本之间的差异是属性逐渐分离和演化的结果。这在某种程度上类似于生物进化树,根节点与叶节点的差距并不是一蹴而就的,而是不断进化和分支的结果。演化不仅反映在粗粒度的类间层面上,即使在同一类别中,也存在着更加细粒度的过渡。如图 3 所示,即使在“人类”类别内部,不同年龄之间的样本可以形成更加细粒度的聚簇,并且聚簇之间的变化是流型过渡的。

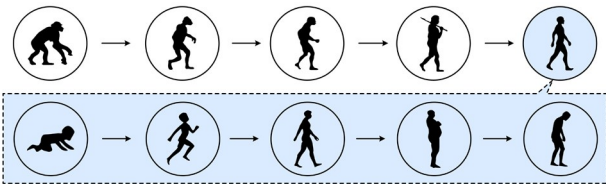


图 3 类内样本之间同样存在聚簇和过渡

Fig. 3 Clustering and transitions exist among intra-class samples

当然,在人工收集的数据集中,可能一个类内的所有变化不能完全缓和地渐进,如果一个类中更细粒度聚簇之间的差异太大,那么这个类就不能用一棵树来描述。换句话说,当有着共同的祖先节点时,一个类别内部的样本从整体上可以被看作是具有流式变换的整体;但当去除祖先节点时,两个或多个子分支之间的差距就会过大,实际上应该用两棵或更多棵树来描绘,而不是强制连接在一起。基于以上的分析,结合 OLeaf<sup>[37]</sup> 的构建,本文设计了一种新的聚类算法:粗粒度引领森林。首先,使用任意距离度量计算数据集  $X$  中样本点之间的距离矩阵;然后计算每个样本点的密度,其中样本点  $i$  的密度计算如式(3)所示:

$$\rho_i = \sum_{j \in \mathbf{I} \setminus \{i\} \setminus \mathbf{O}_i} \exp(-(d_{i,j}/d_{\max})^2) \quad (3)$$

其中,  $\mathbf{I} = \{1, 2, \dots, N\}$  是数据集  $X$  的索引集;  $d_{\max}$  是截断距离;  $d_{i,j}$  是节点  $i$  与节点  $j$  之间的距离;  $\mathbf{O}_i$  是与节点  $i$  的距离超过  $d_{\max}$  的节点的集合。接下来根据密度值对  $\mathbf{I}$  进行降序排序,记为  $\mathbf{S}$ ,即  $S_i$  是密度值第  $i$  大的数据点的索引。

下一步依次对  $\mathbf{S}$  中的点进行处理,如式(4)所示。如果

数据点  $S_i$  尚未合并到粗粒度节点 (CoarseNode) 中,则数据点  $S_i$  和与其相距  $d_{\min}$  距离内的点合并为粗粒度节点,即粗粒度节点的半径为  $d_{\min}$ 。

$$C_{\text{mem}} = \{S_i\} \cup \mathbf{K} \setminus \mathbf{A}, S_i \notin \mathbf{A} \quad (4)$$

其中,  $C_{\text{mem}}$  是新生成的粗粒度节点的所有成员;  $S_i$  会作为该粗粒度节点的代理节点;  $\mathbf{K}$  是距离  $S_i$  在  $d_{\min}$  范围内的节点集合,即  $\mathbf{K} = \{j | j \in \mathbf{I}, d_{S_i, j} < d_{\min}\}$ ;  $\mathbf{A}$  是已访问过节点的集合,即已经合并的粗粒度节点的集合。请注意,如果  $S_i$  本身已经在  $\mathbf{A}$  中,则会跳过新的粗粒度节点的创建,继续处理  $\mathbf{S}$  中的下一个节点。

下一步是寻找新构建的粗粒度节点的引领节点。这里将问题转化为寻找该粗粒度节点的代理节点  $S_i$  的引领节点,记为  $l_i$ 。然后将  $l_i$  所在的粗粒度节点作为当前粗粒度节点中的引领节点。寻找  $S_i$  的前导节点  $l_i$  的过程可以参考式(5):

$$l_{S_i} = \arg \min_j \{d_{S_i, j} | \rho_j > \rho_{S_i}\} \quad (5)$$

$$j \in \mathbf{I} \setminus \{S_i\} \setminus \mathbf{O}_{S_i}$$

其中,  $\mathbf{O}_{S_i}$  是距离  $S_i$  超出  $d_{\max}$  的节点的集合。需要注意的是,  $l_{S_i}$  可能不存在,当找不到  $l_{S_i}$  时,  $S_i$  所在的粗粒度节点就会自动成为粗粒度引领森林中的一个根节点。此外,由于  $\mathbf{S}$  中节点按密度降序排列,因此如果能找到  $l_{S_i}$ , 它一定已经被合并到某一个粗粒度节点中,因为  $l_{S_i}$  密度更大。

### 3.2 基于 CLF 的类内属性重采样

通过 CLF 的构造,可以模拟类别内属性分布的刻画,并且在这个过程中可以通过调整超参数  $d_{\min}$  和  $d_{\max}$  来控制刻画的精细度。接下来,基于 CLF 可以构造不同属性分布的环境,如图 4 所示。CLF 中每一棵树的每个分支都可以代表属性进化的一个新的方向,可将根节点到叶节点的每条路径上的所有数据点归为某个属性分支的成员。

获得了每一个样本的属性标签之后,就可以参考类间重采样<sup>[13]</sup> 的思想,如式(6)所示:

$$p_j = \frac{n_j^q}{\sum_{i=1}^C n_i^q} \quad (6)$$

当进行类间采样时,  $j$  表示类别索引;  $n_j$  表示类别  $j$  中的样本数量;  $C$  表示类别的总数。  $q \in [0, 1]$ , 当  $q=0$  时即进行类间平衡采样。

当对一个类别内的不同属性进行采样时,其思想与类间采样相同,只需要将  $j$  视为属性的索引,  $n_j$  视为属性  $j$  中的样本数,  $C$  视为总属性数。此外,需要注意的是,对于类内采样,与类间采样不同的是,同一样本可能具有多个属性(例如,根节点出现在同一棵树的所有分支中,因此它具有该树可以表示的所有属性),因此需要在一定程度上惩罚这种样本的采样权重。本文惩罚采样权重的方法是将样本的采样权重除以该样本在不同属性中出现的总次数。同时,由于算法中有着粗粒度节点的概念,如图 4 所示,粗粒度节点中成员之间的方差极小,即样本的代表性降低,所提供信息的价值降低,因此,也应该适当减小粗粒度节点中成员的采样权重。本文通过粗粒度节点中的所有成员平分该粗粒度节点的权重来实现。

以图 4 中所构造的 CLF 中的主干树为例(出于演示目的,忽略了孤立样本),下面将详细介绍属性的平衡采样。

首先,可以基于 CLF 来分离出不同的属性,如图 4(b)所示,每条路径中从根节点到叶节点的所有节点都被归划到某一属性,这意味着如果要均匀地对每个属性的样本进行采样,只需要为每个属性分配相等的采样权重即可。其中,对于在多个路径中重复的节点,只需将其采样权重除以重复次数作为惩罚。以计算根节点的权重为例,首先,因为有 3 个属性组,所以  $weight=1/3$ ;然后,由于根节点在 3 个属性群中的权重分别为  $1/5, 1/4$  和  $1/5$ ,因此根节点的全局采样权重分别为  $1/15, 1/12$

和  $1/15$ ,将这 3 个权重相加得到  $weight=13/60$ ;最后对采样权重做出惩罚,即  $weight=weight/repetition=13/180$ 。此外,对于包含多个样本的粗粒度节点,惩罚是类似的,将每个样本的采样权重设置为  $weight=CoarseNode.weight/CoarseNode.length$ 。以属性 2 中的第二个粗粒度节点为例,该节点在属性 2 和属性 3 中重复出现,故采样权重为  $weight=((1/3)\times(1/4)+(1/3)\times(1/5))/2=3/40$ ,所以该粗粒度节点中的每个样本的采样权重为  $weight=(3/40)/2=3/80$ 。

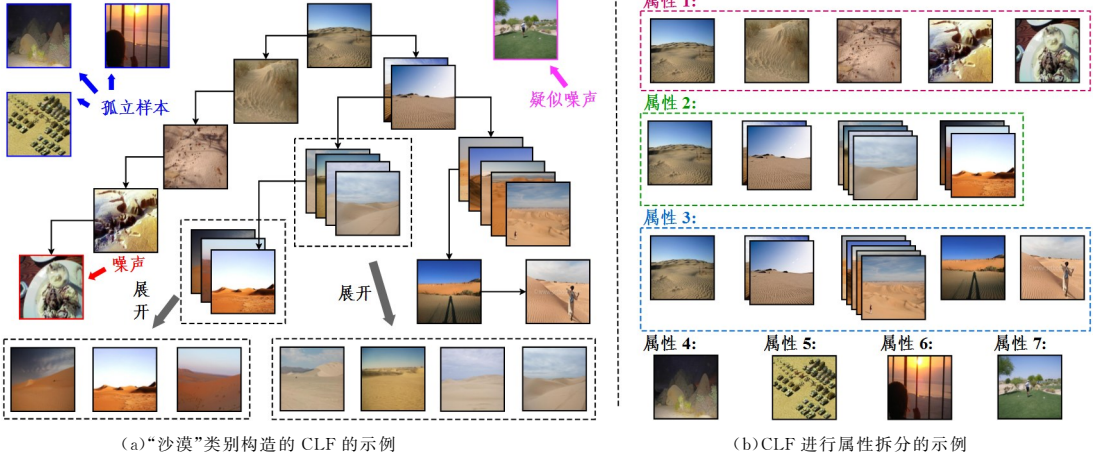


图 4 为“沙漠”类别构建 CLF,并以此进行属性拆分(电子版为彩图)

Fig. 4 Build a CLF for “Desert” category and use it for attribute splitting

### 3.3 多中心损失

在构建了用于不变性特征学习的多个训练环境的基础上,下一步是使用 IRM<sup>[17]</sup>中的损失函数进行训练。然而,由于原始 IRM 损失在现实数据集中存在收敛问题,我们基于 IRM 的思想和 IFL<sup>[8]</sup>中的中心损失设计了一个新的损失函数:多中心损失(MCL)。其可描述为以下优化问题。

$$\begin{aligned} \min_{\theta, \omega} \sum_{e \in \epsilon} \sum_{i \in e} L_{cls}(g(f(x_i^e; \theta); \omega), y_i^e) \\ \text{s. t. } \theta \in \arg \min_{\theta} \sum_{e \in \epsilon} \sum_{i \in e} \|f(x_i^e; \theta) - C(x_i^e)\|_2 \end{aligned} \quad (7)$$

其中,  $\theta$  和  $\omega$  分别是主干网络(特征学习器)和分类器的可学习参数,  $x_i^e$  和  $y_i^e$  分别是环境  $e$  中的第  $i$  个实例及其标签,  $\epsilon$  表示所有的训练环境,  $f(x_i^e; \theta)$  是主干网络从  $x_i^e$  中提取出的特征,  $g(f(x_i^e; \theta); \omega)$  是样本  $x_i^e$  的最终分类结果,  $L_{cls}(g(f(x_i^e; \theta); \omega), y_i^e)$  是环境  $e$  下的分类损失(任意损失函数),  $C(x_i^e)$  是  $x_i^e$  在所有环境中所归属的中心。注意,每个类别中的中心数  $n_{c_{s_i}} \geq 1$ ,  $x_i^e$  所归属的中心取决于  $x_i^e$  位于该类别的 CLF 中的哪棵子树,即  $n_{c_{s_i}} = n_{t_{s_i}}$ , 其中  $n_{t_{s_i}}$  是根据该类别的所有样本构建的 CLF 中的树的总数,  $C(x_i^e)$  的初始值即为  $x_i^e$  所在的树的根节点的代理节点的值。因此,需要注意的是,CLF 与 MCL 两个组件之间是耦合的。该优化问题的实际应用版本如式(8)所示:

$$\begin{aligned} \min_{\theta, \omega} \sum_{e \in \epsilon} \sum_{i \in e} L_{mc} = \min_{\theta, \omega} \sum_{e \in \epsilon} \sum_{i \in e} L_{cls} + \alpha \cdot L_{IFL} \\ = \min_{\theta, \omega} \sum_{e \in \epsilon} \sum_{i \in e} L_{cls}(g(f(x_i^e; \theta); \omega), y_i^e) + \alpha \cdot \|f(x_i^e; \theta) - C(x_i^e)\|_2 \end{aligned} \quad (8)$$

其中,  $L_{IFL} = \|f(x_i^e; \theta) - C(x_i^e)\|_2$  是不变性特征学习的约束损失,  $\alpha$  是权衡参数。该损失函数是 Center Loss 的 IRM

版本,相对于原始版本,该损失函数进一步增加了算法的鲁棒性。正如之前在 CLF 中的介绍,对于一些人工数据集,即使在同一类别中,也可能存在样本之间的差距非常大的情况,即 CLF 中的树的数量大于 1。如图 5 所示,在“车库”类别中,它实际上可以分为 3 个子类,即“车库外”“车库内(有车停放)”和“车库内(无车停放)”,这 3 个子类别的特征差异很大。如果只使用一个中心,那么在训练过程中使每个类别的特征逐渐接近一个中心实际上会损害特征的学习,这就是使用多中心损失的出发点。

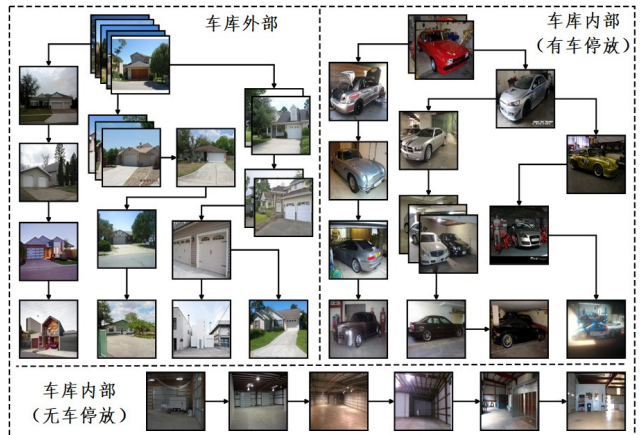


图 5 “车库”类别的 CLF 中存在多棵树(这代表着样本之间存在巨大差异)

Fig. 5 “Garage” category has multiple trees in its CLF(this represents the huge differences among the samples)

### 3.4 整体框架

本方案的总体框架如图 6 所示。每个环境都有一对平衡

因子( $q_{cls}, q_{attr}$ ),  $q_{cls}$  是进行类间采样的平衡因子,  $q_{attr}$  是进行类内采样的平衡因子。可以通过不同的 ( $q_{cls}, q_{attr}$ ) 对来构建不同的环境, 然后在 MCL 的约束下对多环境共享模型参数

进行训练。在采样过程中, 属性分布的刻画是通过 CLF 实现的。在 MCL 中, 中心的个数也是由相应类别的 CLF 中树的数量所决定的。

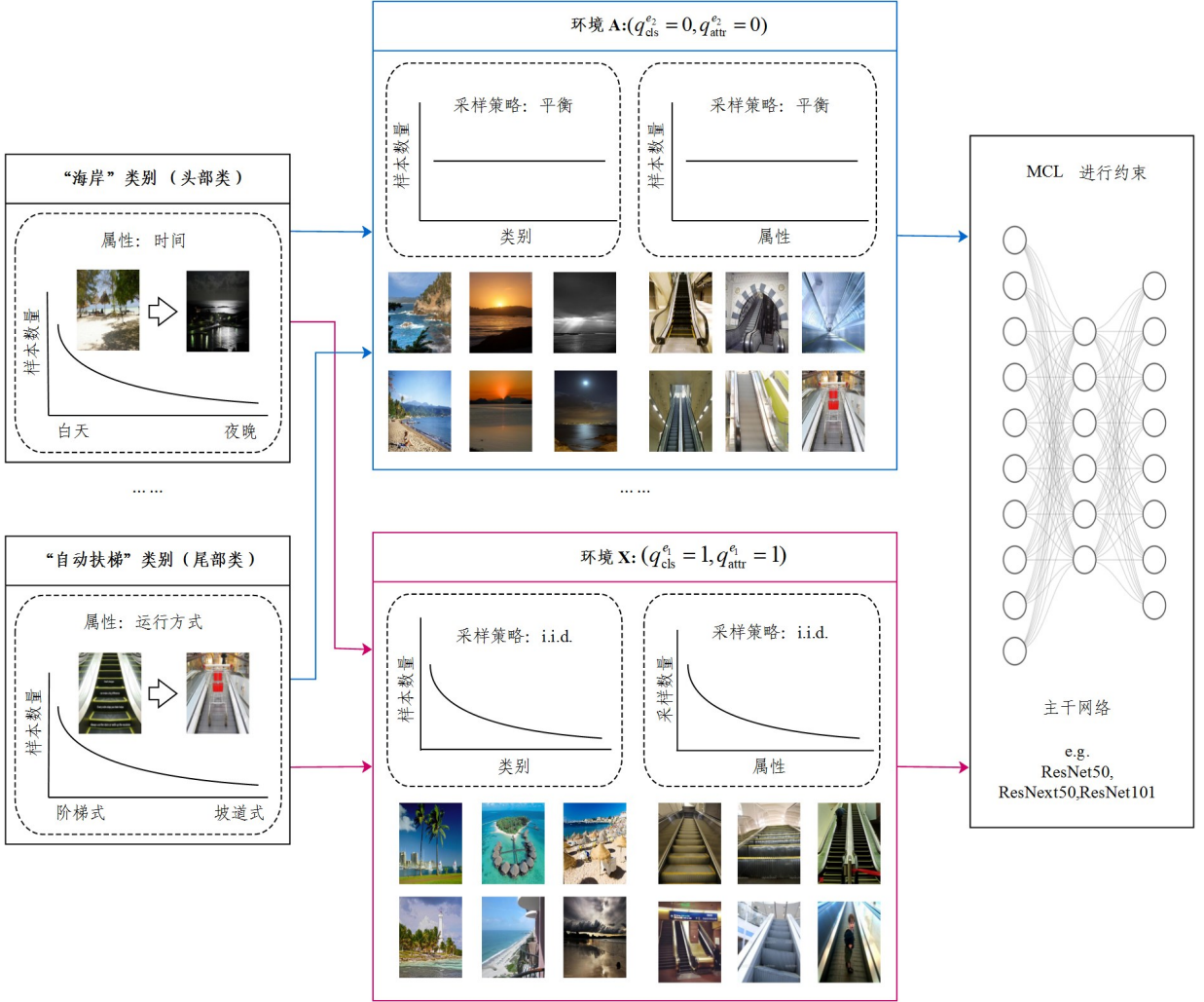


图 6 总体框架图

Fig. 6 Overall framework of the proposed scheme

如算法 1 所示, 整体框架的详细过程可分为两个阶段: 首先, 由于在构建 CLF 时需要使用样本的初始特征进行聚类, 因此需要进行  $M$  轮正常采样的训练, 以获得预测不完美的初始模型。然后, 使用初始特征来构建 CLF, 并通过 CLF 和不同的平衡因子对来构建不同的环境。例如, 在本文的实验设置中构造了两个环境, 分别采用了 ( $q_{cls}^e=1, q_{attr}^e=1$ ) 和 ( $q_{cls}^e=0, q_{attr}^e=0$ ) 作为平衡因子对。其中, 第一个环境是正常进行独立同分布采样的环境; 而第二个环境的采样策略是类别和属性都趋于平衡地采样。然后, 特征学习器被不断更新, 相应地每个类别的 CLF 以及 MCL 中的中心也在不断更新。其中, 在第二阶段中执行更新的  $epoch$  的步长可以调整, 不需要固定为每个  $epoch$  都进行一次更新。

#### 算法 1 整体训练流程

输入: 训练集  $\{(x, y)\}$  和各环境的平衡因子对  $\{(q_{cls}^e, q_{attr}^e)\}$

输出: 特征学习器  $f(\cdot; \theta)$  和分类器  $g(\cdot; w)$

1. 初始化:  $f(\cdot; \theta)$  和  $g(\cdot; w)$  的参数  $\theta$  和  $w$
2. /\*  $M$  轮预热阶段, 尚未引入 MCL \*/

3. for  $k \leftarrow 1$  to  $M$  do
4.  $\hat{\theta}, \hat{w} \leftarrow \hat{\theta}, \hat{w} \in \arg \min_{\theta, w} L_{cls}(g(f(x; \theta); w), y)$
5. end
6. /\* 构造各类别的 CLF \*/
7.  $\{F_y\} = \text{construct\_clf}(\{(x, y)\}, \theta)$
8. /\* 构造不同采样策略的环境 \*/
9.  $\{e_n\} = \text{construct\_env}(\{(q_{cls}^e, q_{attr}^e)\}, \{F_y\})$
10. for  $k \leftarrow 1$  to  $N$  do
11.  $\{(x^e, y^e)\} = \text{data\_load}(\{e_n\})$
12. /\* 根据 CLF 获取各样本的中心值 \*/
13.  $\{C_y\} = \text{read\_centers}(\{F_y\})$
14. /\* 使用 MCL 更新参数 \*/
15.  $\hat{\theta}, \hat{w} \leftarrow \hat{\theta}, \hat{w} \in \arg \min_{\theta, w} L_{cls}(g(f(x; \theta); w), y) + \alpha \cdot L_{mc}(f(x; \theta), \{C_y\})$
16. /\* 不断更新 CLF 和各环境 \*/
17.  $\{F_y\} = \text{construct\_clf}(\{(x, y)\}, \theta)$
18.  $\{e_n\} = \text{construct\_env}(\{(q_{cls}^e, q_{attr}^e)\}, \{F_y\})$
19. end

## 4 实验与分析

### 4.1 估协议

在进行实验分析之前,首先引入两种新的评估协议:CLT(Class Wise Long Tail)协议和 GLT(Generalized Long Tail)协议。这两种协议都是在 GLT<sup>[8]</sup>的 第一个基线中提出的。

#### 1)CLT 协议

训练集中的样本遵循长尾分布,即训练集可以直接从长尾数据集中独立同分布采样获得;而测试集中的样本是在类别层次平衡采样获取的。注意,CLT 中没有考虑属性分布问题。因为 CLT 的训练集和测试集具有相同的属性分布和不同的类别分布,因此可以评估类别层次长尾分类的有效性。

#### 2)GLT 协议

与 CLT 相比,GLT 考虑了属性分布的差异,即引入了式(1)中的属性偏差。GLT 中的训练集与 CLT 中的相同,符合长尾分布;而测试集中的属性分布趋于平衡。由于 GLT 中的训练集和测试集具有不同的属性分布和类分布,因此可以评估模型处理类间长尾分类和类内长尾分类的能力。

### 4.2 基准数据集和评估指标

#### 1)数据集

本文在 ImageNet-GLT 和 MSCOCO-GLT 这两个基准数据集上对本文方法以及当前流行的长尾分类进行了评估。这两个基准都是在 GLT 的第一个基线<sup>[8]</sup>中提出的。

(1)ImageNet-GLT 是 ImageNet<sup>[39]</sup>的长尾子集。对于测试集,根据以下类别频率将其划分为 3 个子集:头部类子集为  $N_c > 100$  的类别,中部类子集为  $100 \geq N_c \geq 20$  的类别,尾部类子集为  $N_c < 20$  的类别,其中  $N_c$  是相应类别内的样本数量。在该数据集中构建用于属性平衡的测试集时,使用 K-Means 将每个类别中的图像简单地聚类为 6 组,然后对每个类别中每个组分别采样 10 张图像。

(2)MSCOCO-GLT 是 MSCOCO-Attribute<sup>[40]</sup>的长尾子集,该数据集明确标记了 196 个不同的属性,其中具有多个标签的每个对象都被裁剪为单独的图像。对于测试集,根据

以下类别将其分为 3 个子集: $Index_c \leq 10$  的类别为头部类子集, $10 < Index_c \leq 22$  的类别为中部类子集, $Index_c > 22$  的类别为尾部类子集,其中  $Index_c$  是按样本数量升序排列的类别的索引。

#### 2)评估指标

在本文的实验中,使用 3 种评估指标来评估各方法的性能。(1)准确率(Accuracy): $\sum_{i=1}^N TP_i / \sum_{i=1}^N (TP_i + FN_i)$ ,该指标也是传统长尾方法中使用的 Top1-Accuracy 或微平均召回率。(2)精确率(Precision): $(1/N) \sum_{i=1}^N (TP_i / (TP_i + FP_i))$ 。(3)准与精确率的平均调和(F1-Score<sup>\*</sup>): $2 \times (Accuracy \times Precision) / (Accuracy + Precision)$ 。引入此度量,是为了更好地揭示传统类间长尾分类方法中未被关注的准确率与精确率失衡的问题<sup>[8]</sup>。上述计算式中, $N$  表示类别的总数, $TP_i$ , $FP_i$  和  $FN_i$  分别表示第  $i$  个样本的真正例(True Positives)、假正例(False Positives)和假负例(False Negatives)的数量。

### 4.3 分类性能对比

所提方案不与特定的模型结构耦合,可以与其他长尾分类方法无缝结合。在下面的比较实验中,遵循当前长尾研究<sup>[1,8]</sup>中的分类,将当前的长尾方法分为 3 类,即重平衡、信息增强和模块改进,并从这 3 类中各取几种主流的方法进行比较和提升。其中,在重平衡类别中选择了 3 种方法:BBN<sup>[14]</sup>,BLSoftmax<sup>[23]</sup>和 Logit-Adj<sup>[6]</sup>。对于信息增强,选择了 Mixup<sup>[34]</sup>和 RandAug<sup>[32]</sup>两种方法;对于模块改进类别,选择了 RIDE<sup>[35]</sup>和 TADE<sup>[36]</sup>两种方法。它们采用了集成学习的思想,都是当前的 SOTA 方法。此外,还加入了 GLT 领域中的第一个强基线 GLTv1<sup>[8]</sup>作为比较,如表 1 和表 2 所列。其中,带星号的方法(\* <方法名>+Mine)是集成了我们的组件之后的方法,粗体数字表示该方法类别中的最佳结果。可以看出,本文提出的方法在所有分类中都得到了最好的结果,特别是与 RandAug 方法或 RIDE 方法相结合时,在两个基准的几乎所有评估指标中都得到了最好的结果。

表 1 在 ImageNet-GLT 基准上各方法的性能对比

Table 1 Performance comparison of various methods on ImageNet-GLT benchmark

(%)

协议 各子集	CLT 协议(结果格式:(准确率 精确率))				GLT 协议(结果格式:(准确率 精确率))					
	头部类	中部类	尾部类	全集	头部类	中部类	尾部类	全集		
重平衡	Baseline	58.39 38.35	36000 52.15	13.98 55.34	41.65 47.11	50.70 32.50	27.80 43.99	10.18 47.70	34.32 39.95	
	BBN <sup>[14]</sup>	62.66 43.45	44.36 55.44	14.66 57.57	47.22 50.96	53.37 36.45	34.97 46.81	10.73 46.08	38.70 42.56	
	BLSoftmax <sup>[23]</sup>	54.03 47.07	41.65 46.83	28.37 37.58	44.61 45.54	46.45 40.45	32.67 38.59	21.38 29.57	36.49 37.98	
	Logit-Adj <sup>[6]</sup>	53.30 49.05	43.49 44.12	31.86 35.00	45.67 44.72	45.43 41.75	34.2 35.90	24.58 27.78	37.25 37.02	
	GLTv1 <sup>[8]</sup>	62.51 42.32	38.69 57.23	17.41  <b>65.00</b>	45.03 52.43	54.40 36.23	30.18 49.61	12.64 55.47	37.24 45.14	
	* Baseline+Ours	<b>62.84</b>  42.82	39.62 57.25	18.61 61.55	45.76 52.12	<b>54.53</b>  36.47	31.30 49.99	13.80  <b>58.27</b>	37.97 45.82	
信息增强	* BLSoftmax+Ours	58.37 53.88	<b>44.98</b>  51.97	33.82 39.05	<b>48.66</b>  50.80	49.92 46.89	<b>36.11</b>  44.31	26.14 30.03	<b>40.14</b>  43.20	
	* Logit-Adj+Ours	43.30  <b>75.13</b>	41.19  <b>59.08</b>	<b>45.43</b>  27.11	42.92  <b>60.70</b>	36.01  <b>70.18</b>	32.6  <b>52.48</b>	<b>36.26</b>  21.76	34.51  <b>54.95</b>	
	模块改进	Mixup <sup>[34]</sup>	60.14 38.02	31.46 56.67	7.59 32.82	39.35 45.63	51.68 32.21	23.87 48.25	5.47 28.27	32.23 38.84
		RandAug <sup>[32]</sup>	64.14 42.23	40.10 58.27	14.96 59.51	45.94 52.04	55.70 35.87	31.61 50.15	10.20 47.29	38.03 44.01
		* Mixup+Ours	67.86 47.90	45.50 62.76	24.98  <b>67.56</b>	51.37 57.54	59.28 40.95	36.16 54.09	17.63 57.72	42.63 49.38
		* RandAug+Ours	<b>69.12</b>   <b>49.31</b>	<b>47.64</b>   <b>63.08</b>	<b>26.97</b>  67.01	<b>53.13</b>   <b>58.16</b>	<b>60.85</b>   <b>42.30</b>	<b>38.46</b>   <b>55.21</b>	<b>19.90</b>   <b>60.11</b>	<b>44.63</b>   <b>50.78</b>
模块改进	TADE <sup>[35]</sup>	57.30  <b>55.22</b>	46.85 50.29	<b>34.69</b>  37.93	49.21 50.41	49.61  <b>48.19</b>	37.55 42.59	<b>27.52</b>  32.21	40.87 43.28	
	RIDE <sup>[36]</sup>	63.18 51.44	47.67 52.55	29.91 47.38	51.21 51.33	54.83 44.02	38.25 44.20	22.77 38.12	42.56 43.21	
	* TADE+Ours	60.69 55.15	48.15  <b>52.22</b>	33.38 42.12	50.95 51.88	52.77 48.15	39.09 44.06	26.51 33.32	42.68 44.08	
* RIDE+Ours	<b>64.83</b>  54.60	<b>50.95</b>   <b>56.21</b>	33.28  <b>50.15</b>	<b>53.85</b>   <b>54.66</b>	<b>57.10</b>  47.63	<b>42.00</b>   <b>48.42</b>	25.50  <b>41.23</b>	<b>45.57</b>   <b>47.03</b>		

表 2 在 MSCOCO-GLT 基准上各方法的性能对比

Table 2 Performance comparison of various methods on MSCOCO-GLT benchmark

(%)

协议 各子集	CLT 协议(结果格式:(准确率 精确率))				GLT 协议(结果格式:(准确率 精确率))				
	头部类	中部类	尾部类	全集	头部类	中部类	尾部类	全集	
	重平衡	Baseline	81.27 71.08	74.13 76.06	50.17 85.61	71.88 76.15	74.59 64.82	66.25 69.08	35.75 77.22
	BBN <sup>[14]</sup>	83.59 70.21	76.13 76.3	47.25  <b>90.98</b>	72.98 77.02	75.36 61.97	68.29 68.61	31.75  <b>81.38</b>	63.41 68.73
	BLSoftmax <sup>[23]</sup>	80.77 71.87	75.67 69.94	45.25 90.08	71.31 74.84	73.23 65.13	68.63 62.64	32.08 79.6	62.81 67.09
	Logit-Adj <sup>[6]</sup>	81.55 73.95	76.00 75.55	<b>60.83</b>  82.04	74.97 76.28	73.64 66.98	68.71 67.61	<b>46.58</b>  70.69	66.00 68.01
	GLTv1 <sup>[8]</sup>	82.45 73.09	76.42 79.53	55.58 86.87	74.40 78.60	76.41 67.07	67.33 70.54	39.75 81.02	65.07 71.39
	* Baseline+Ours	<b>83.00</b>  73.95	77.88  <b>80.48</b>	55.92 90.08	75.28  <b>79.99</b>	<b>76.91</b>  68.24	67.50  <b>71.41</b>	39.67 80.33	65.31  <b>72.05</b>
	* BLSoftmax+Ours	83.00 74.17	78.37 76.24	53.92 86.51	75.07 77.58	75.86 67.98	70.46 68.14	40.08 80.26	66.22 70.59
	* Logit-Adj+Ours	82.68  <b>77.16</b>	<b>80.17</b>  75.45	59.17 83.54	<b>76.78</b>  77.77	75.14  <b>70.56</b>	<b>71.54</b>  66.34	45.83 73.90	<b>67.59</b>  69.50
信息增强	Mixup <sup>[34]</sup>	82.41 72.95	73.12 79.53	54.33 87.47	72.76 78.68	74.64 66.43	65.33 71.00	36.00 78.11	62.79 70.74
	RandAug <sup>[32]</sup>	84.23 74.33	77.42 79.28	56.33 87.07	75.64 79.01	77.91 68.56	69.67 70.95	39.58 78.43	66.57 71.59
	* Mixup+Ours	85.00 76.96	81.42  <b>83.19</b>	<b>63.33</b>  88.89	79.03 82.01	79.23  <b>70.89</b>	72.25  <b>74.84</b>	<b>46.50</b>  80.62	<b>69.57</b>   <b>74.54</b>
	* RandAug+Ours	<b>86.55</b>   <b>78.31</b>	<b>81.50</b>  82.82	62.42  <b>90.29</b>	<b>79.47</b>   <b>82.65</b>	<b>79.05</b>  69.93	<b>72.92</b>  74.51	44.33  <b>81.82</b>	69.33 74.28
模块改进	TADE <sup>[35]</sup>	83.55 76.47	80.79 73.89	50.50 89.34	75.57 78.06	77.18 68.62	71.58 65.63	33.50 82.32	65.83 70.22
	RIDE <sup>[36]</sup>	84.23 77.56	82.29 77.79	58.42 89.69	78.09 80.16	77.45 71.20	75.17 68.67	41.25 83.33	69.02 72.66
	* TADE+Ours	85.59 77.48	82.25 77.85	55.08 90.40	77.90 80.31	79.00 70.54	73.96 68.94	40.67  <b>83.65</b>	68.98 72.59
	* RIDE+Ours	<b>86.27</b>   <b>79.64</b>	<b>83.13</b>   <b>79.89</b>	<b>63.50</b>   <b>90.41</b>	<b>80.26</b>   <b>81.98</b>	<b>80.09</b>   <b>72.77</b>	<b>75.17</b>   <b>71.45</b>	<b>47.83</b>  83.03	<b>71.38</b>   <b>74.34</b>

此外,本文方法旨在同时处理类内长尾和类间长尾,尽管该方法的出发点是解决类内属性长尾的问题,但它也通过消除属性长尾引起的虚假相关性来应对了类间长尾的挑战。从图 7 中可以清楚地看到 Cognisance 框架在两个基准的两个协议上对现有长尾分类方法实现的改进,而从 CLT 到 GLT 的所有方法的性能都有所下降,这表明长尾问题不是纯粹的

类间长尾问题。虽然类内长尾分类更具挑战性,但 Cognisance 仍然可以成功地改进所有现有的主流长尾方法,并且在 F1-Score\* 指标上对所有方法的改进都高出近 5%。最后,表 3 中记录了各种方法在长尾分布测试集上的实验结果,可以看到,与其他方法相比,本文的方法仍然可以在绝大多数评估指标上获得最佳结果。

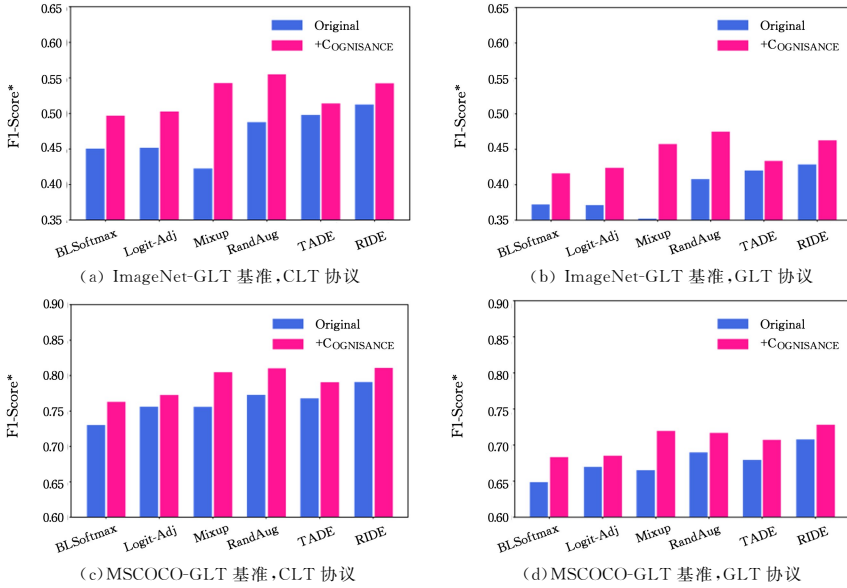


图 7 Cognisance 框架对主流长尾分类方法的提升

Fig. 7 Improvement of mainstream long-tailed classification methods by Cognisance framework

表 3 在两种基准的长尾测试集上各方法的性能对比

Table 3 Performance comparison of various methods on long-tailed test set of two benchmarks

(%)

基准 评估指标	ImageNet-GLT			MSCOCO-GLT			
	准确率	精确率	F1-Score*	准确率	精确率	F1-Score*	
重平衡	Baseline	53.93	44.46	48.74	85.74	79.98	82.76
	BBN <sup>[14]</sup>	58.60	48.9	53.32	84.84	78.04	81.30
	BLSoftmax <sup>[23]</sup>	51.73	41.97	46.34	83.69	73.81	78.44
	Logit-Adj <sup>[6]</sup>	50.94	40.58	45.17	85.28	73.6	79.01
	GLTv1 <sup>[8]</sup>	58.28	50.43	54.07	86.67	<b>81.88</b>	<b>84.21</b>
	* Baseline+Ours	<b>58.85</b>	50.98	<b>54.64</b>	<b>86.86</b>	81.24	83.96
	* BLSoftmax+Ours	56.24	47.58	51.55	86.55	78.48	82.32
	* Logit-Adj+Ours	43.61	<b>55.15</b>	48.71	85.52	73.36	78.98

(续表)

	基准 评估指标	ImageNet-GLT			MSCOCO-GLT		
		准确率	精确率	F1-Score <sup>*</sup>	准确率	精确率	F1-Score <sup>*</sup>
信息 增强	Mixup <sup>[34]</sup>	55.25	46.76	50.66	86.95	82.00	84.40
	RandAug <sup>[32]</sup>	59.88	51.28	55.25	87.64	81.02	84.20
	* Mixup+Ours	64.09	56.57	60.09	88.81	81.55	85.03
	* RandAug+Ours	<b>65.15</b>	<b>56.85</b>	<b>60.72</b>	<b>89.12</b>	<b>84.02</b>	<b>86.49</b>
模块 改进	TADE <sup>[35]</sup>	55.21	46.52	50.49	86.57	79.91	83.10
	RIDE <sup>[36]</sup>	60.17	48.28	53.58	88.03	81.79	84.80
	* TADE+Ours	58.06	47.44	52.21	88.00	81.72	84.74
	* RIDE+Ours	<b>62.11</b>	<b>50.80</b>	<b>55.89</b>	<b>89.02</b>	<b>81.85</b>	<b>85.28</b>

#### 4.4 实验说明

在本方案中,有两种类型的参数可以进行调整:1)与整个训练过程相关的参数,例如,用于预热的 *epoch* 的数量、用于更新样本采样权重的 *epoch* 步长以及训练环境的数量等;2)与聚类相关的参数,CLF 构建过程中需要用到两个参数  $d_{\min}$  和  $d_{\max}$ ,前者用于控制粗粒度节点的半径,后者用于控制子树分裂的精细度。该算法预设了两个相对通用的默认值,但它们仍然可以根据具体的数据集进行调整。需要指出的是,这些参数在一定范围内对整体效果影响不大,但仍然存在进一步优化的空间。

另外,距离度量在两个关键步骤中发挥作用:1)CLF 的构建,在进行 CLF 聚类时需要计算距离矩阵,这里的距离度量可以进行灵活切换;2)多中心损失,在进行 MCL 的优化时,需要计算样本到其归属中心的距离,同样,这里的距离度量也可以根据需求自由切换,只需与 CLF 构建时所使用的距离度量保持一致即可。

在本文中,采用欧氏距离作为默认的距离度量,但根据实际情况,切换到其他距离度量可能会带来更好的结果。鉴于模型训练的时间开销,本文仅采用了一个距离度量,但这并不妨碍核心思想的阐述。本文实验代码可于 GitHub 获取<sup>1)</sup>。

**结束语** 本文从两个粒度层次(类间和类内)关注了长尾分类问题,并且提出了一种新的长尾分类框架 Cognisance,该框架包含两个重要组件:粗粒度引领森林(CLF)和多中心损失(MCL)。CLF 作为一种无监督的学习方法,旨在刻画类别内部属性的分布,从而指导训练过程中多个环境的构建来支持不变性特征的学习;MCL 作为中心损失的改进版本,旨在取代传统的 IRM 损失,进一步增强模型在真实世界数据集上的鲁棒性。

在广义长尾分类的现有基准数据集 MSCOCO-GLT 和 ImageNet-GLT 上进行的广泛实验详尽地展示了所提方案的显著效果。特别值得注意的是, Cognisance 框架的优势在于其低耦合的设计,这意味着它可以与其他长尾分类方法轻松集成,从而显著提升这些长尾分类模型的性能。

此外,任何真实世界的数据集都是不完美的,在长尾数据集上进行噪声识别更是一项极为困难的挑战,而本文方案提出的 CLF 实际上具有识别噪声的潜力。如图 4 粉色框所示,如果 CLF 中有一棵树只包含一个粗粒度节点,并且该节点只包含一个样本,那么该样本很有可能是噪声样本。如图 4 红色框所示,如果某样本位于一颗很深的树的叶子节点中,那么该样本也很有可能是噪声样本。直观上来讲,当一个样本越

孤立或深度越深,说明该节点越离群,可以对这些样本进行进一步处理以确定它们是否为噪声,这一思想将在未来的研究工作中进一步探索。

#### 参考文献

- [1] ZHANG Y, KANG B, HOOI B, et al. Deep Long-Tailed Learning: A Survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 10795-10816.
- [2] YANG Y, WANG H, KATABI D. On multi-domain long-tailed recognition, imbalanced domain generalization and beyond[C]// Proceedings of European Conference on Computer Vision. Tel-Aviv, Israel, 2022: 57-75.
- [3] BUDA M, MAKI A, MAZUROWSKI M A. A systematic study of the class imbalance problem in convolutional neural networks [J]. Neural Networks, 2018, 106: 249-259.
- [4] ANDO S, HUANG C Y. Deep over-sampling framework for classifying imbalanced data[C]// Machine Learning and Knowledge Discovery in Databases: European Conference. Skopje, Macedonia, 2017: 18-22.
- [5] YANG Y, ZHANG G, KATABI D, et al. Me-net: Towards effective adversarial robustness with matrix estimation[C]// International Conference on Machine Learning. California, USA, 2019: 7025-7034.
- [6] CAO K, WEI C, GAIDON A, et al. Learning imbalanced datasets with label-distribution-aware margin loss[C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 1567-1578.
- [7] SUN L, WANG K, YANG K, et al. See clearer at night: towards robust nighttime semantic segmentation through day-night image conversion[C]// Artificial Intelligence and Machine Learning in Defense Applications. Strasbourg, France, 2019: 77-89.
- [8] TANG K, TAO M, QI J, et al. Invariant feature learning for generalized long-tailed classification[C]// European Conference on Computer Vision. Tel-Aviv, Israel, 2022: 709-726.
- [9] DRUMMOND C, HOLTE R C. C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling[C]// Proceedings of the ICML'03 Workshop on Learning from Imbalanced Data Sets II. Washington, USA, 2003: 1-8.
- [10] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning[C]// International Conference on Intelligent Computing. Berlin, Heidelberg, 2005: 878-887.
- [11] HE H, GARCIA E A. Learning from imbalanced data [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1263-1284.
- [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense

<sup>1)</sup> <https://github.com/jinyery/Cognisance>

- object detection[C] // Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017; 2980-2988
- [13] KANG B, XIE S, ROHRBACH M, et al. Decoupling Representation and Classifier for Long-Tailed Recognition[C] // International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020.
- [14] ZHOU B, CUI Q, WEI X S, et al. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Washington, USA, 2020; 9719-9728.
- [15] ZHU B, NIU Y, HUA X S, et al. Cross-domain empirical risk minimization for unbiased long-tailed classification[C] // Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2022; 3589-3597.
- [16] CUI P, ATHEY S. Stable learning establishes some common ground between causal inference and machine learning[J]. Nature Machine Intelligence, 2022, 4(2): 110-115.
- [17] ARJOVSKY M, BOTTOU L, GULRAJANI I, et al. Invariant risk minimization[J]. arXiv:1907.02893, 2019.
- [18] CREAGER E, JACOBSEN J H, ZEMEL R. Environment inference for invariant learning[C] // International Conference on Machine Learning, 2021; 2189-2200.
- [19] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C] // Computer Vision—ECCV 2016; 14th European Conference, Amsterdam, The Netherlands, 2016; 499-515.
- [20] WANG T, LI Y, KANG B, et al. The devil is in classification: A simple framework for long-tail instance segmentation [C] // Computer Vision—ECCV 2020; 16th European Conference, Glasgow, UK, 2020; 728-744.
- [21] ESTABROOKS A, JO T, JAPKOWICZ N. A multiple resampling method for learning from imbalanced data sets[J]. Computational Intelligence, 2004, 20(1): 18-36.
- [22] ZHANG Z, PFISTER T. Learning fast sample re-weighting without reward data[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021; 725-734.
- [23] REN J, YU C, MA X, et al. Balanced meta-softmax for long-tailed visual recognition[J]. Advances in Neural Information Processing Systems, 2020, 33: 4175-4186.
- [24] ELKAN C. The foundations of cost-sensitive learning[C] // International Joint Conference on Artificial Intelligence, Seattle, USA, 2001; 973-978.
- [25] JAMAL M A, BROWN M, YANG M H, et al. Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020; 7610-7619.
- [26] TAN J, WNAG C, LI B, et al. Equalization loss for long-tailed object recognition[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020; 11662-11671.
- [27] TAN J, LU X, ZHANG G, et al. Equalization loss v2: A new gradient balance approach for long-tailed object detection[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021; 1685-1694.
- [28] MENON A K, JAYASUMANA S, RAWAT A S, et al. Long-tail learning via logit adjustment[C] // International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020.
- [29] CHU P, BIAN X, LIU S, et al. Feature space augmentation for long-tailed data[C] // Computer Vision—ECCV 2020; 16th European Conference, Glasgow, UK, 2020; 694-710.
- [30] WANG J, LUKASIEWICZ T, HU X, et al. Rsg: A simple but effective module for learning imbalanced datasets[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021; 3784-3793.
- [31] HU X, JIANG Y, TANG K, et al. Learning to segment the tail [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020; 14045-14054.
- [32] CUBUK E D, ZOPH B, SHLENS J, et al. Randaugment: Practical automated data augmentation with a reduced search space [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020; 702-703.
- [33] LIU J, SUN Y, HAN C, et al. Deep representation learning on long-tailed data: A learnable embedding augmentation perspective[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020; 2970-2979.
- [34] ZHANG H, CISCHE M, DAUPHIN Y N, et al. mixup: Beyond Empirical Risk Minimization[C] // International Conference on Learning Representations, Vancouver, Canada, 2018.
- [35] WANG X, LIAN L, MIAO Z, et al. Long-tailed Recognition by Routing Diverse Distribution-Aware Experts[C] // International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020.
- [36] ZHANG Y, HOOI B, HONG L, et al. Self-supervised aggregation of diverse experts for test-agnostic long-tailed recognition [J]. Advances in Neural Information Processing Systems, 2022, 35: 34077-34090.
- [37] XU J, WANG G, DENG W. DenPEHC: Density peak based efficient hierarchical clustering [J]. Information Sciences, 2016, 373: 200-218.
- [38] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191): 1492-1496.
- [39] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115: 211-252.
- [40] PATTERSON G, HAYS J. Coco attributes: Attributes for people, animals, and objects[C] // Computer Vision—ECCV 2016; 14th European Conference, Amsterdam, The Netherlands, 2016; 85-100.



**YANG Jinye**, born in 1998, postgraduate, is a member of CCF (No. N8701G). His main research interests include granular computing and machine learning.



**XU Ji**, born in 1979, Ph.D., professor, is a member of CCF (No. 12919M). His main research interests include data mining, granular computing and machine learning.