

基于样本分布损失的图像多标签分类研究

朱旭东, 熊贇

引用本文

朱旭东, 熊贇. 基于样本分布损失的图像多标签分类研究[J]. 计算机科学, 2022, 49(6): 210-216.

ZHU Xu-dong, XIONG Yun. Study on Multi-label Image Classification Based on Sample Distribution Loss[J].

Computer Science, 2022, 49(6): 210-216.

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)

基于解耦-检索-生成的图像风格化描述生成模型

Stylized Image Captioning Model Based on Disentangle-Retrieve-Generate

计算机科学, 2022, 49(6): 180-186. <https://doi.org/10.11896/jsjcx.211100129>

基于特征注意力融合网络的遥感变化检测研究

Remote Sensing Change Detection Based on Feature Fusion and Attention Network

计算机科学, 2022, 49(6): 193-198. <https://doi.org/10.11896/jsjcx.210500058>

多分支 RA 胶囊网络及在图像分类中的应用

Multi-branch RA Capsule Network and Its Application in Image Classification

计算机科学, 2022, 49(6): 224-230. <https://doi.org/10.11896/jsjcx.210400087>

机器学习在金融资产定价中的应用研究综述

Application of Machine Learning in Financial Asset Pricing:A Review

计算机科学, 2022, 49(6): 276-286. <https://doi.org/10.11896/jsjcx.210900127>

基于注意力门控图神经网络的文本分类

Text Classification Based on Attention Gated Graph Neural Network

计算机科学, 2022, 49(6): 326-334. <https://doi.org/10.11896/jsjcx.210400218>

基于样本分布损失的图像多标签分类研究

朱旭东 熊 贇

复旦大学计算机科学与技术学院 上海 200433

上海市数据科学重点实验室(复旦大学) 上海 200433

(18212010052@fudan.edu.cn)

摘 要 与一般图像分类场景下的数据分布情况不同,在图像多标签分类问题的场景下,不同标签类别之间存在样本数量分布不均衡,少量头部类别通常占据大多数样本数量的情况。而由于多个标签间同时标记的相关性,再加上多标签下困难样本的分布还与数据分布和类别分布相关,使得单标签问题中解决数据不平衡的重采样等方法在多标签场景下无法有效适用。文中提出了一种基于图像多标签场景下样本分布损失和深度学习的分类方法。首先对多标签数据不均衡分布设置类别相关重采用损失,并通过动态学习方式防止分布过度异化,然后设计非对称样本学习损失,设置对正负样本和困难样本的不同学习能力,同时通过软化样本学习权重减少信息丢失。相关数据集的实验显示,所提算法在解决多标签数据分布不均衡场景下的样本学习问题时取得了很好的效果。

关键词: 多标签;标签关系;重采样;深度学习;图像分类

中图法分类号 TP391

Study on Multi-label Image Classification Based on Sample Distribution Loss

ZHU Xu-dong and XIONG Yun

School of Computer Science and Technology, Fudan University, Shanghai 200433, China

Research Center of Dataology and Data Science, Fudan University, Shanghai 200433, China

Abstract Different from the data distribution in general image classification scenarios, in the scenario of multi label image classification, the sample number distribution among different label categories is unbalanced, and a small number of head categories often account for the majority of sample size. However, due to the correlation between multiple labels, and the distribution of difficult samples under multiple labels is also related to the data distribution and category distribution, the re-sampling and other methods for solving the data imbalance in the single label problem cannot be effectively applied in the multi label scenario. This paper proposes a classification method based on the loss of sample distribution in multi label image scene and deep learning. Firstly, the unbalanced distribution of multi label data is set with category correlation, and the loss is re-used, and the dynamic learning method is used to prevent the excessive alienation of distribution. Then, the asymmetric sample learning loss is designed, and different learning abilities for positive and negative samples and difficult samples are set. At the same time, the information loss is reduced by softening the sample learning weight. Experiments on related data sets show that the algorithm has achieved good results in solving the sample learning problem in the scene of uneven distribution of multi-label data.

Keywords Multi-Label, Label relation, Re-sample, Deep learning, Image classification

1 引言

多标签分类问题的目的在于构建样本空间和标签空间一对多的对应关系。作为多标签问题的一个重要方向,图像多标签分类在计算机视觉领域的内容理解、目标检测等多方面有着广泛应用。

传统的图像多标签分类方法通常将多标签问题转化为

训练多个相互独立的单标签分类器,分别学习不同标签的分类并最后进行汇总。一些方法尝试捕捉多标签相关性,以为分类提供更多有效信息。CNN-RNN^[1]通过卷积神经网络学习图像特征,同时通过循环神经网络构建标签序列的结构以引入标签相关性。MLGCN^[2]最先使用图神经网络进行标签关系的建模,图中节点表示标签,边表示标签相关度权重,通过统计标签共现信息来获得。

到稿日期:2021-03-26 返修日期:2021-08-15

基金项目:国家自然科学基金(U1636207)

This work was supported by the National Natural Science Foundation of China(U1636207).

通信作者:熊贇(yunx@fudan.edu.cn)

图像多标签分类中通常存在不同标签类别数据分布不均衡的情况,特别是数据长尾分布的现象,将会造成不同类别分类能力存在较大差异。一般通过数据重加权或者重采样的方式来解决数据不均衡分布的问题,如BBN^[3]构建双学习分支多重采样分布和原始数据分布进行融合,但多个标签共同标记同一样本,单标签场景下的重采样会造成相关标签的采样,会引入新的数据不平衡。为了解决该问题,DBL^[4]采用构建标签共现和重加权的方式,但该方法对于原有数据存在分布异化的风险,而在多标签场景下存在数据分布过度拟合的风险。对于正负样本和困难样本,Focal Loss^[5]通过可调样本权重进行不同类别样本的学习。ASL^[6]设置动态样本边界进行样本区分学习,但这些方法没有较好地解决多标签不均衡数据分布场景下的分类问题。

本文提出了一种基于样本分布损失的图像多标签分类方法DAB(Distribution Asymmetric Balance),本文的创新点如下:1)提出了在图像多标签场景下考虑多标签相关性的样本重采样方式,减小了标签相关性对数据期望采样数量的影响;2)提出了多标签场景下关注困难负样本学习的损失模块,对正负样本的学习方式进行了解耦,并为其分别设计了不同的学习率变化函数,同时对负样本的学习进行了软化处理;3)将基于多标签的重采样模块和正负样本学习模块统一在一个框架下,并提出通过动态训练的方式来完成实验分析。

2 相关工作

2.1 图像多标签分类

图像多标签分类的传统解决方法如DCMI^[7]通过特征提取器获得对应的图像信息,并将图像信息输入不同的独立的二分类器,以学习不同类别的分类能力。部分方法^[8-10]尝试对图像中与标签分类相关度较高的部分进行重点学习,以更有效地获得分类信息,其中对视觉注意力机制的研究受到了众多学者的重视。MLRRD^[9]通过基于循环神经网络的注意力模块来提取不同图像区域的特征信息,并输出神经网络隐含层参数,以隐含层参数构建多标签分类器。RCP^[10]采用图像随机裁剪合并方式对多标签图像中的复杂内容信息进行最大化应用,很大程度上解决了多标签数据不充分的问题,但是该算法对标签相关信息的挖掘不够,并且随机裁剪使得模型的泛化性较差。

文献^[11-14]尝试使用跨模态模型将标签语义和图像特征进行融合学习。MS-CMA^[12]使用度量学习方法将图像语义和图像特征进行融合。SSGRL^[13]通过双线性池化对标签语义和图像特征信息进行多模态学习,对融合后的特征使用图门控网络进行学习。

也有一些研究者借助序列模型或是图网络模型挖掘图像多标签中的标签关系,为图像分类提供更多信息。CNN-RNN^[1]利用序列神经网络对标签序列进行编码,同时利用并行的卷积神经网络提取图像特征,与标签序列进行线性变换,最终输出分类结果。F-GCN^[15]构建基于图卷积神经网络的快速模型,以标签共现信息为输入,融合图像信息以进行端到端训练。

2.2 数据不均衡分布问题

数据不均衡分布是现实场景下一个常见的现象,特别是存在数据长尾分布。而在多标签分类任务中,由于标签相关性带来的标签分布不均衡的问题,将数据分布不均衡和标签分布问题相结合,为解决数据不均衡分布带来了新的思考。常见的处理数据不均衡分布的方法分为以下3种:数据重采样、数据重加权和迁移学习知识。

数据重采样的方法可以分为头部数据欠采样和尾部数据过采样^[16],它们都是根据部分数据在整体数据中的比重进行反向加权抽样,如式(1)所示,有时也会采用部分数据增强的方式。

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

其中, C 是数据集的类别数量, n_i 是类别为 i 的样本数量, p_i 为从类别 i 中抽取数据的概率。

重加权方法的重点在于在计算损失函数时对不同类别分别加权,相比重采样在计算形式上更灵活。CBLBE^[17]加强了数据中头部数据损失函数的权重,同时减小了尾部数据损失函数的学习权重。MCIL^[18]通过度量学习平衡头部和尾部分类区域的损失权重,来强制各类别数据中心均匀间隔。基于知识迁移的方法通过将具有丰富数据量的头部类别中学习到的特征信息转移到表征信息不足的尾部类别中,来解决分布不均衡的问题。OTLR^[16]通过两阶段的训练框架,使用实例平衡采样首先学习表征信息,再在第二阶段加入记忆模块对分类器进行微调。

2.3 困难样本学习

一些学习较为困难的样本,通常包含对模型重要的有效信息或者是无关信息,充分挖掘困难样本对提升模型的影响具有重要意义。传统机器学习方法中的困难样本学习方法有基于支持向量机^[19]的样本交替训练方式、利用规则设定删去简单样本,以及基于增强决策树^[20]在正样本和随机负样本上进行收敛。深度学习相关的困难样本的处理方式主要包含困难样本数据增强、困难样本损失函数设计、在线困难样本挖掘等方法。Focal Loss^[5]在目标检测任务中通过设计损失函数,来增加学习困难样本的损失权重和增加困难样本对损失梯度的权重。OHEM^[21]通过在线选择的方式对困难样本进行更新挖掘。

相比其他图像多标签分类方法,DAB算法的优势在于:1)对于现有图像多标签分类场景下数据分布不均衡和正负样本学习的问题,提出了创新有效的解决方法;2)将解决数据分布不均衡的方法和正负样本学习的方法统一在同一个学习框架下,减少了训练成本并降低了难度。

3 DAB 算法

本节将对本文提出的DAB算法进行详细描述,DAB算法主要分为以下模块:1)类别分布重采样模块;2)样本学习损失平衡模块。

下文将对DAB中的各个模块进行介绍,DAB的整体结构如图1所示。

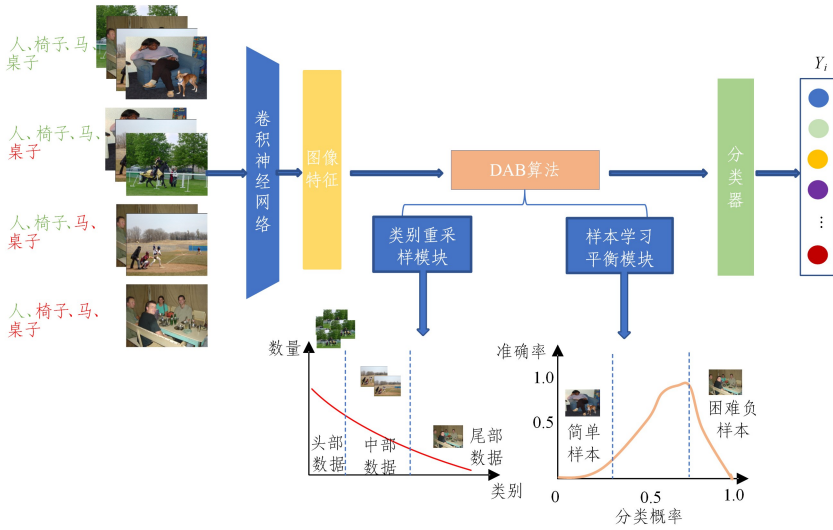


图1 DAB模型的结构

Fig. 1 DAB model structure

3.1 类别分布重采样

对于多标签样本数据集,假定一共有 C 个类别,全体训练样本数据量为 N ,设每张图片的标签为 $Y_k = [y_1^k, \dots, y_C^k] \in \{0, 1\}$,那么 $n_i = \sum_{k=1}^N y_i^k$ 为包含类别 i 的训练样本数量。由于每张图片的标签数量 $Y_k > 1$,可以得到 $N \leq \sum_{i=0}^C n_i$ 。对于一般算法训练中采用的均衡采样方式,类别 i 的采样概率为 $p_i = n_i / N$,而在多标签场景下,类别 i 的采样概率受到其他类别影响,导致 $p_i \neq n_i / N$ 。由此可知,多标签场景下训练数据的直接采样并不满足均匀采样。

进一步分析,假设条件概率 $p(i|j)$ 为某一样本已经被标签 j 标记的情况下,同时被标签 i 标注的概率,即 $p(i, j) = n_{i,j} / n_j$,其中 $n_{i,j}$ 即为标签 i 与标签 j 共同标记的样本数量。因此,当从训练数据中随机选取某一类别的某一样本时,该类别被标记为类别 i 的概率为:

$$\hat{p}_{ij} = \sum_{j=0}^C p(i|j) = \frac{\sum_{j=0}^C n_i \cap n_j}{n_j} \quad (2)$$

从 $\hat{p}_{ij}$ 出发,构建类别非均衡采样损失模块,首先对于类别 i 中的数据样本 x_k 有标签标注概率 y_i^k 。定义从 $\hat{p}_{ij}$ 出发,构建类别非均衡采样损失模块,定义类别抽样概率为:

$$p_i^c(x_k) = \frac{1}{n_i} \quad (3)$$

同时,对于样本 x_k 和该样本对应的标签 y_k ,样本 x_k 的采样次数受到 y_k 中标记的标签个数影响, y_k 中标记的标签个数增加导致 x_k 被重复采样。定义样本抽样概率为:

$$p_i^f(x_k) = \frac{1}{\sum_{y_i^k=1} n_i} \quad (4)$$

相应地,为了平衡实际采样概率和期望采样概率,定义重采样率 s_i^k 如式(5)所示。重采样率定义的原理在于将样本的采样概率从多标签中进行抽离,关注整体样本频次中单个样本的抽样概率。

$$s_i^k = \frac{p_i^c(x_k)}{p_i^f(x_k)} \quad (5)$$

然而,由于不同标签之间的样本数量存在不均衡性,采样数量较少类别的采样概率可能趋近于 0,这增加了分类器的学习难度。为了使不同采样数量的类别对应的分类器具有更均衡的分类能力,使得采样函数更为平滑,减小函数优化的难度,本文在重采样率 s_i^k 的基础上进一步定义了类别平衡重采样率 $\hat{s}_i$ 。

$$\hat{s}_i = \beta + \frac{1}{1 + \exp(-\mu \times (s_i - \epsilon))} \quad (6)$$

其中, β 为采样偏移量参数, μ 和 ϵ 控制采样偏移率。此处参数设计的原理在于获得了变化较为平缓的采样率,同时保证了采样率在合理的范围内。在原有的二分类交叉熵损失函数 L_{BCE} 的基础上,设计了一个重采样损失 L_{S-BCE} 。

$$L_{BCE} = -(y_n \times \log(\delta(z_n)) + (1 - y_n) \times \log(1 - \delta(z_n))) \quad (7)$$

$$L_{S-BCE} = \frac{1}{C} \sum_{i=1}^C [y_i^k \times \log(\delta(z_n)) + (1 - y_i^k) \times \log(1 - \delta(z_n))] \times \frac{\hat{s}_i}{s_i^k} \quad (8)$$

L_{S-BCE} 重采样损失主要解决了多标签对应数据样本采样不均衡的问题。在模型训练的过程中,为了减小重采样对原有数据分布的影响,平衡重采样分布和原始数据的分布,不同于以往的重采样方法^[3-4, 17], DAB 对得到的重采样损失进行了动态设定。定义动态平衡参数为 φ , E_m 为总共的训练周期数量, E_i 为当前的训练周期。

$$\varphi = 1 - \left(\frac{E_i}{E_m} \right)^2 \quad (9)$$

对重采样数据分布和原始数据分布进行融合,动态训练设置的原理和依据在于,模型在训练初期学习率较大,此时梯度下降较快,更多关注原始真实数据分布,而在训练后期增加对重采样信息的关注,相当于模型已经对整体数据信息有了初步的整体性的学习后,对部分数量较少类别进行了重点学习。因此,可以在总体保持原有分类能力的基础上增加对数量较少类别的分类能力。最后,可以得到类别非均衡重采样损失 L_{SR-BCE} 为:

$$L_{\text{SR-BCE}} = (1-\varphi)L_{\text{S-BCE}} + \varphi L_{\text{BCE}} \tag{10}$$

3.2 样本学习平衡模块

多标签图像数据不仅存在样本采样不均匀的情况,而且正负样本和难易样本不平衡也影响了分类器的分类能力,特别是困难负样本的学习对分类器提出了很大挑战。有效学习困难负样本的信息并且减少对易分类样本的过度学习,对于分类器分类能力的提升有很大帮助。与单标签分类问题不同,多标签中标签相关性的存在使得正负样本和难易样本的采样受到了样本中共现标签的影响。

Focal Loss^[5]是目标检测领域解决正负样本不平衡和学习困难样本的有效方法,但无法有效适应多标签分类的场景。本文针对多标签中正负样本分布的情况,同时参考文献[5]中的模型设计了样本非均衡损失。Focal Loss 中损失函数的实现如下:

$$FL(p_t) = -\alpha(1-p_t)^\gamma \log(p_t) \tag{11}$$

其中, p_t 是模型输出的概率; α 和 γ 是设定的学习加权参数,当 $\gamma=0$ 时, Focal Loss 与普通的二分类交叉熵损失函数相同。分析公式可知,当 γ 设定的值增加时,减小对简单样本的学习,更多关注困难样本的信息, α 则决定了正负样本训练损失占全部样本损失的比重。参考 Focal Loss 中对不同难易样本和正负样本的不同处理方式, DAB 希望通过改变多标签图像中正样本和负样本的学习权重水平, 设 γ_{pos} 和 γ_{neg} 分别为多标签图像正样本和多标签图像负样本的类别学习加权参数。本文希望模型能从正样本中学习更多有效信息, 因而设置正负学习权重关系为 $\gamma_{\text{pos}} > \gamma_{\text{neg}}$, 则正负样本的训练损失 l_{pos} 和 l_{neg} 分别为:

$$l_{\text{pos}} = (1-p)^{\gamma_{\text{pos}}} \log(p) \tag{12}$$

$$l_{\text{neg}} = p^{\gamma_{\text{neg}}} \log(1-p) \tag{13}$$

我们注意到, 文献[6]在对负样本的判定中设置截断静态参数阈值, 将低于阈值的简单负样本学习权重设置为 0。但是, 直接对样本进行权重截断, 设置硬分类, 将会导致很多样本信息无法得到利用, 并且实际样本信息也并不是按阈值分布的。为了能充分获取不同损失程度的样本信息, 同时考虑到多标签场景下标签分布的相关性, DAB 算法设置了可调节的软化样本调节概率学习机制。具体来说, 调节概率 $\hat{p}$ 的定义如式(14)所示:

$$\hat{p} = \sigma(M \cdot p) \tag{14}$$

其中, M 调节因子是一个可调参数, $\sigma(\cdot)$ 为 Sigmoid 函数, 该机制的原理在于利用 Sigmoid 的平滑曲线, 当待分类样本的负样本预测值较低时, 易于判定该样本为负样本, 此时减少样本的学习, 而负样本的预测值较高时, 则该样本可能为困难负样本, 此时增加对样本信息的学习。

$$L_{1A} = \begin{cases} (1-p)^{\gamma_{\text{pos}}} \log(p), & L_{\text{pos}} \\ (1-\hat{p})^{\gamma_{\text{neg}}} \log(1-\hat{p}), & L_{\text{neg}} \end{cases} \tag{15}$$

3.3 DAB 模型损失

综合类别分布重采样损失和样本学习平衡损失, 本文将 $L_{\text{SR-BCE}}$ 和 L_{1A} 融合到统一的端到端学习框架 L_{DAB} 中, 具体如式(16)所示。

$$L_{\text{DAB}} = \begin{cases} \hat{s}_i^k (1-y_i^k) (1-\hat{p})^{\gamma_{\text{neg}}} [\varphi \log(1-\hat{p}) + (1-\varphi) \log(\log(1-\hat{p}))], & L_{\text{neg}} \\ \hat{s}_i^k y_i^k (1-\hat{p})^{\gamma_{\text{pos}}} [\varphi \log(p) + (1-\varphi) \log(\log(p))], & L_{\text{pos}} \end{cases} \tag{16}$$

4 实验分析

本节将首先介绍 DAB 进行实验的相关数据集和对比算法, 之后介绍实验设置, 最后对实验结果进行分析总结。实验中使用的软硬件配置及其参数如表 1 所列。

表 1 实验环境的软硬件设置

Table 1 Software and hardware configuration of experimental environment

软硬件配置项目	配置参数
CPU	Intel Xeon Gold 5118 CPU@2.30 GHz
显卡	8 * NVIDIA 1080Ti
硬盘/TB	16
内存容量/GB	250
操作系统	CentOS Linux 7.5.1804(core)
编程语言	Python 3.6.9
编程框架	Pytorch 1.0

4.1 实验数据集

本文实验使用的数据集为 VOC-MLT 和 COCO-MLT, 为了更好地验证 DAB 算法在数据不平衡场景下的效果, 参考文献[4]中的方法, 从 VOC-2012^[22] 和 COCO-2017^[23] 中按照 prareto 分布 $pd f(x) = \frac{x_{\min}^\alpha}{x^{\alpha+1}}$ 进行抽取^[24], 其中 α 为数据规模衰减参数。通过不同类别间的抽样互动约束尾部样本数量, 对数据集中的每个类别 C 进行抽样概率排序, 通过引用期望的分布增减类别实例, 使得新构建数据集中各个类别的样本数排序与原有数据集中的样本数排序保持一致。数据集的具体介绍如下:

(1)VOC-MLT。设置抽样参数 α 为 6, 选取 VOC-2012 数据集 20 个类别中的 1142 张图像, 每个类别最多有 775 张图像, 最少有 4 张图像。头部、中部和尾部数据的类别数分别为 6, 6, 8, 使用含 4952 张图片的 VOC-2012 测试集进行测试。

(2)COCO-MLT。设置抽样参数 α 为 6, 选取 COCO-2017 数据集中全部 80 个类别的 1909 张图像, 每个类别最多有 1128 张, 最少有 6 张, 使用含 5000 张图片的 COCO-2017 测试集进行测试。整理完成的 COCO-MLT 数据集头部、中部和尾部的数据的类别数分别为 22, 33, 25。

4.2 对比算法

本文选取部分在图像多标签分类和长尾分布图像分类算法中表现较好的方法与本文提出的 DAB 算法进行实验对比。

(1)ERM。该算法对所有样本设置相同权重和相同抽样概率, 保持所有样本对模型的学习权重相同。

(2)Re-weighting(RW)^[3]。该算法通过平滑的重加权使抽样权重类别样本频率的平方根成反比, 并在每个小学习批次中对权重进行标准化。

(3)Re-sampling(RS)^[23]。通过类别频率设置重采样, 采样率与类别数据频率成反比。

(4)Focal-Loss^[12]。该算法构建困难样本损失函数, 通过

设置样本信息共享参数对数据中的负样本比重进行控制,同时通过设置调制参数来对困难样本进行挖掘。

(5)MLGCN^[2]。该算法通过以图像多标签中标签为节点、标签相关性为边权重的图模型,对多标签中各个标签关系进行建模。通过 GCN 层对多标签图模型进行更新,最后由图模型分类层和图像特征层相乘获得最终的分类结果。

(6)DBL^[4]。通过设置多标签数量采样对样本分布进行平衡,同时设定负样本阈值以减小负样本对算法学习的影响。

4.3 实验相关设置

实验过程中,采用 ResNet101 模型作为图像特征提取的基准模型。设置训练批数据大小为 64,使用优化器为 SGD,动量参数设置为 0.9,权重衰退设置为 0.0001,学习率设置为 0.01,全部训练周期均设置为 100,同时每训练 40 个周期学习率降为原来的 10%,软化调节因子 M 设置为 1。对训练输入图像进行随机裁剪,尺寸设定为 448×448 ,同时采用随机水平翻转进行训练数据增强。

4.4 对比实验

实验采用的对比指标是平均均值准确率。在 VOC-MLT 数据集上的相关实验结果如表 2 所列,在 COCO-MLT 数据集上的实验结果如表 3 所列。可以看到,与当前的对比算法相比,本文提出的 DAB 算法在解决数据不平衡分布场景下的图像多标签分类问题中取得了明显的效果。

表 2 VOC-MLT 数据集的 mAP 指标结果
Table 2 mAP of dataset VOC-MLT
(单位: %)

Algorithm	All Data	Head Data	Middle Data	Tail Data
ERM	70.86	68.91	80.20	65.31
RW	74.70	67.58	82.81	73.96
RS	75.38	70.95	82.94	73.05
Focal-Loss	73.88	69.41	81.43	71.56
MLGCN	68.92	70.14	76.41	62.39
DBL	78.94	73.22	84.18	79.30
DAB	79.19	73.41	84.37	80.12

表 3 COCO-MLT 数据集的 mAP 指标结果
Table 3 mAP of dataset COCO-MLT
(单位: %)

Algorithm	All Data	Head Data	Middle Data	Tail Data
ERM	41.27	48.48	49.06	24.25
RW	42.27	48.62	45.80	32.02
RS	46.97	47.58	50.55	41.70
Focal-Loss	49.46	49.80	54.77	42.14
MLGCN	44.24	44.04	48.36	38.96
DBL	53.55	51.13	57.05	51.06
DAB	53.71	52.02	57.87	51.73

根据表 2 和表 3 的实验结果,可以总结如下:

(1)DAB 在非均衡场景下的图像多标签分类任务中,VOC-MLT 和 COCO-MLT 数据集的 mAP 指标均高于对比算法,相比表现最好的对比算法,DBA 算法在全部数据、头部数据、中间数据、尾部数据 4 类分布的 mAP 指标分别提高了 1.6、1.0、1.1 和 2.8,这说明 DAB 模型在解决数据分布不平衡场景下的图像多标签分类是有效的。

(2)VOC-MLT 数据集中尾部数据通常会出现单标签,降低分类任务难度和复杂性,甚至使得尾部数据指标高于头部

数据指标的情况存在。这一特点一定程度地减小了头尾数据分布不平衡带来的影响,VOC-MLT 中的尾部数据与其他类别的指标差异较 COCO-MLT 中的更小。

(3)COCO-MLT 数据集中存在较强的头尾数据相关性,即某些尾部数据类别与特定头部数据共现相关性较高,因此通过重采样方式训练时尾部数据的 mAP 指标提升明显。通过 DAB 算法重平衡的设置,头部、中部和尾部数据的 MAP 指标均有不同程度的提升。

(4)与传统解决数据不平衡分布的重采样相比,DAB 可以更好地避免重采样带来的数据冗余问题,防止头部数据被过度忽略和尾部数据被过度采用,同时保证数据整体信息的有效利用。

4.5 消融实验

为了更好地分析 DAB 中各个模块的作用,对 DAB 进行消融实验。实验中使用的数据集是 VOC-MLT,相关参数和硬件配置同对比实验保持一致。消融实验的配置如下。

(1)移除样本损失模块。移除样本学习平衡模块,只使用类别非均衡损失模块作为补充学习损失,其他不变。

(2)移除类别重采样模块。移除类别非均衡重采样模块,只使用样本学习平衡模块作为补充学习损失,其他部分不变。

(3)DAB 模型。保持 DAB 算法中全部模块以及相关参数不变,用于消融实验中的分析比较。

通过对 DAB 中各模块消融实验的配置,图 2 给出了消融实验的相关结果,经过分析可以总结如下。

(1)DAB 中各模块的移除均会造成整体 mAP 指标的下降,说明 DAB 中各模块均具有有效性,同时也证明 DAB 各模块融合可以达到更好的效果。

(2)对各模块的实验结果进行比较,非对称重采样的类别分布对指标提升的贡献最大,说明样本分布对提升分类指标有很大影响,该采样方式相比各个样本平均权重的采样方式,增加了对尾部数据的采样权重,较好地解决了图像多标签场景下数据分布不平衡,尤其是尾部数据学习不充分的问题。

(3)样本学习平衡模块在类别非均衡重采样的基础上,对多标签场景下的困难样本和正负样本信息进行了进一步的有效挖掘,提升了分类效果。但是,单独的样本学习平衡模块的效果不如样本重采样模块,说明多标签下的数据分布失衡问题比多标签相关的样本学习失衡问题对分类的影响更大。

(4)移除类别重采样后,中间部分数据指标比原有 DAB 模型指标略高,原因在于中间部分数据原本分布较均衡,而 DAB 重采样在平衡整体全部数据分布信息特别是头部和尾部数据时,对中间部分数据的学习和相应指标有一定的负面影响。

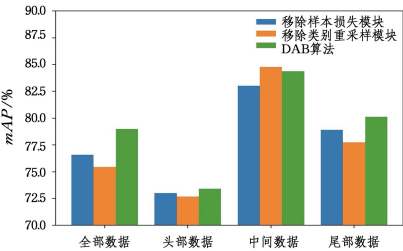


图 2 DAB 消融实验的结果

Fig. 2 Result of DAB ablation study

4.6 动态训练验证

对于 DAB 中的动态训练机制,本文将分别使用 L_{SR-BCE} (动态重采样二分类交叉熵损失), L_{S-BCE} (重采样二分类交叉熵损失), L_{BCE} (交叉熵损失) 代替原有 L_{SR-BCE} 融入 L_{DAB} 架构进行实验。实验结果如图 3 所示。

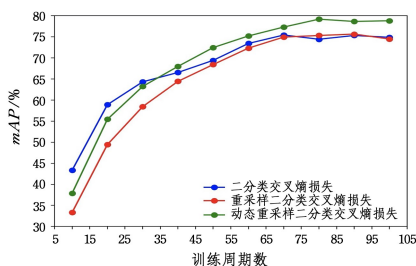


图 3 动态训练实验结果

Fig. 3 Experimental results of dynamic training

(1) 相比仅使用原有分布训练方式和仅使用重采样训练方式,使用动态训练机制的 DAB 算法的指标表现更好。通过分析实验结果得知,其原因是仅使用原有分布的训练无法对尾部数据进行采样分布的均衡,而仅使用重采样的方式过度改变了原有数据的信息分布。而 DAB 的训练机制可以较好地平衡原有数据分布和重采样数据分布,减少了信息损失。

(2) 训练初期, DAB 算法和仅使用重采样的方式指标提升较慢,而使用原数据分布的方式指标提升较快。通过分析实验结果得知,其原因是训练初期梯度下降较快,而仅使用原数据分布的方式在训练初期更能表达整体数据的分布信息。这也说明了 DAB 在训练初期梯度下降快时先对整体数据分布进行大致学习,训练后期再重点对重采样分布后的信息进行学习,在保留对原有数据分布表达能力的基础上提升了对尾部数据的表达能力。

(3) 相比仅使用原分布训练的方式,虽然仅使用重采样的训练的方式在训练初期指标提升较慢,但在训练末期两者相差不大。这说明在数据集范围一定的情况下,随着训练的充分进行,重采样的设置更能够表达真实分布数据中的信息。

4.7 参数分析

本节对 DAB 中非对称重采样参数 μ 和样本学习平衡参数 γ 进行了敏感性分析。实验时保持算法其他部分参数不变,对非对称重采样参数 μ 从 1~15 遍历,对样本学习平衡参数 γ_{pos} 和 γ_{neg} 从 0~3 遍历,参数分析实验采用 VOC-MLT 数据集。实验结果如图 4 与图 5 所示。

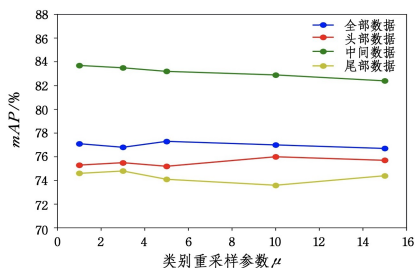


图 4 参数 μ 的敏感性分析

Fig. 4 Sensitivity analysis of parameter μ

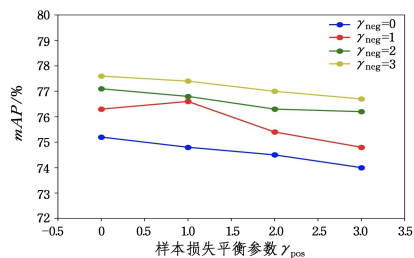


图 5 参数 γ 的敏感性分析

Fig. 5 Sensitivity analysis of parameter γ

分析参数敏感性实验的结果,可以得出如下结论:

(1) 对多标签各类别重采样参数 μ 时可以看到,中间部分数据 mAP 指标表现最好,中间部分数据分布较为均衡,使得模型充分学习。当 μ 的数值由 1~5 变化时, mAP 指标逐渐提升,而当 μ 大于 5 时, mAP 有所下降。通过分析实验结果得知,其原因是当 μ 数值变化时,重采样率相应提高,而 μ 为 5 时重采样使多标签数据集分布处于均衡状态,继续提升时部分类别过度采样影响了指标效果。

(2) 仅将 γ_{pos} 设置为 0 时,模型取得了最好的指标效果,这也可以说明增加对困难样本学习的梯度可以有效提升学习效果。而当 γ_{pos} 增加时,对简单样本学习效果的提升导致整体指标受到影响。

(3) 当 γ_{neg} 逐渐增加时,模型中负样本所贡献的损失权重逐渐减小,正样本对模型整体损失的贡献比例增大,使得相关指标数值增加。

(4) 实验数据中,当保持 $\gamma_{neg} > \gamma_{pos}$ 时,测试指标在 γ_{pos} 和 γ_{neg} 增加的情况下保持上升;当 $\gamma_{neg} > \gamma_{pos}$ 时,模型学习困难样本时得到的损失大于学习简单样本时得到的损失,更多的是学习样本数据中的有效信息。

结束语 针对图像多标签分类问题中存在的分布不均和样本学习不均衡的问题,本文提出了 DAB 算法,通过构造适用于多标签分类场景下的双分支数据采样结构,平衡不同类别样本的数据频率,同时设计困难样本非对称损失函数,以提升对不同样本的学习能力。

本文对实验使用的数据集、实验设置和对比算法进行了介绍,选取了 6 种在图像多标签分类和长尾分布图像分类算法中表现较好的方法,与本文提出的 DAB 算法进行了对比分析,验证了 DAB 算法的有效性。之后对算法进行消融实验和参数敏感性实验,验证了 DAB 算法中模块设置和参数设置的有效性。

参考文献

- [1] WANG J, YI Y, MAO J H, et al. Cnn-rnn: A unified framework for multi-label image classification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Press, 2016: 2285-2294.
- [2] CHEN Z M, WEI X S, WANG P, et al. Multi-Label Image Recognition With Graph Convolutional Networks. [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Press, 2016: 5177-5186.

[3] 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. 2020: 9719-9728.

[4] WU T, HUANG Q, LIU Z, et al. Distribution-balanced loss for multi-label classification in long-tailed datasets[C]// European Conference on Computer Vision. Cham: Springer, 2020: 162-178.

[5] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980-2988.

[6] BEN-BARUCH E, RIDNIK T, ZAMIR N, et al. Asymmetric Loss For Multi-Label Classification [EB/OL]. (2020-09-29) [2021-05-18]. <https://arxiv.org/abs/2009.14119>.

[7] GONG Y, JIA Y, LEUNG T, et al. Deep convolutional ranking for multilabel image annotation[EB/OL]. (2013-12-17) [2014-04-14]. <https://arxiv.org/abs/1312.4894>.

[8] SHENG L, MA J F, YANG R X. Research on CNN Image Classification Algorithm Based on Feature Exchange [J]. Computer Engineering, 2016, 29(6): 927-933.

[9] WANG Z, CHEN T, LI G, et al. Multi-label image recognition by recurrently discovering attentional regions[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 464-472.

[10] WANG M, LUO C, HONG R, et al. Beyond object proposals: Random crop pooling for multi-label image recognition[J]. IEEE Transactions on Image Processing, 2016, 25(12): 5678-5688.

[11] WANG Y B, ZHENG W J, CHEN Y S. Multi label classification algorithm based on PLSA learning probability distribution semantic information [J]. Journal of Nanjing University(Natural Science), 2016, 29(6): 927-933.

[12] YOU R, GUO Z, CUI L, et al. Cross-modality attention with semantic graph embedding for multi-label classification[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(7): 12709-12716.

[13] CHEN T, XU M, HUI X, et al. Learning semantic-specific graph representation for multi-label image recognition [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 522-531.

[14] GU G H, CAO Y Y, LI G. Image Hierarchical Classification Based on Semantic Label Generation and Partial Order Structure [J]. Journal of Software, 2016, 29(6): 927-933.

[15] WANG Y, XIE Y, LIU Y, et al. Fast Graph Convolution Network Based Multi-label Image Recognition via Cross-modal Fusion[C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020: 1575-1584.

[16] LIU Z, MIAO Z, ZHAN X, et al. Large-scale long-tailed recognition in an open world[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2537-2546.

[17] CUI Y, JIA M, LIN T Y, et al. Class-balanced loss based on effective number of samples[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9268-9277.

[18] HAYAT M, KHAN S, ZAMIR W, et al. Max-margin class imbalanced learning with gaussian affinity[J]. arXiv:1901.07711, 2019.

[19] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32(9): 1627-1645.

[20] LI B, LIU Y, WANG X. Gradient harmonized single-stage detector[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(1): 8577-8584.

[21] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 761-769.

[22] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[23] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]// European Conference on Computer Vision. Cham: Springer, 2014: 740-755.

[24] SHEN L, LIN Z, HUANG Q. Relay backpropagation for effective learning of deep convolutional neural networks[C]// European Conference on Computer Vision. Cham: Springer, 2016: 467-482.

ZHU Xu-dong, born in 1995, postgraduate. His main research interests include computer vision and graph neural network.

XIONG Yun, born in 1980, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include data mining and graph neural network.

(责任编辑:喻黎)