



## 长尾识别研究进展

张明<sup>1,2</sup> 翟俊海<sup>1,2\*</sup> 许垒<sup>1,2</sup> 高光远<sup>1,2</sup>

1 (河北大学数学与信息科学学院 河北保定 071002)

2 (河北省机器学习与计算智能重点实验室 (河北大学) 河北保定 071002)

(\*翟俊海 E-mail: mczjh@126.com)

### 1 引言

近年来,深度学习在图像分类、目标检测、人脸识别等计算机视觉任务中取得了巨大的成功。然而,当深度学习模型遇到长尾分布数据时,往往会表现不佳。这是因为长尾分布数据的头部少数类占据了大多数数据,而尾部多数类却占据了很少的一部分数据。因此,深度学习模型在处理长尾分布数据时会偏向头部类。但我们对它的期望是可以在所有类上都表现良好,而不是偏向头部类。长尾识别的本质便是这两者之间的不匹配问题。

我们将现有的解决方案分为五类:一是重采样,其策略是试图构造出平衡的数据集;二是重加权,其策略是在损失函数中给小的类别分配大的权重;三是迁移学习,分别对头部类和尾部类建模,将学到的头部类的信息、表示、知识迁移给少数类使用;四是解耦特征学习和分类器学习,由于类别重新平衡方法(重采样和重加权)都会损害深度网络学习到的深度特征的表示能力,该方法把不平衡学习分为两个阶段,在特征学习阶段正常采样,在分类器学习阶段平衡采样,以此来克服类别重新平衡方法的缺点;五是其它方法,是一些长尾学习在其它领域的相关研究工作。

### 2 主要方法

#### 2.1 数据重采样

过采样中的数据合成方法是利用已有样本生成更多样本,其中最常见的一种方法为Smote<sup>[1]</sup>,它利用少数类样本在特征空间的相似性来生成新样本。对于少数类样本 $x_i \in S_{min}$ ,从它属于少数类的K近邻中随机选取一个样本点 $\hat{x}_i$ ,生成一个新的少数类样本 $x_{new}$ :

$$x_{new} = x_i + (\hat{x}_i - x_i) * \delta \quad (1)$$

其中 $\delta \in [0,1]$ ,是一个随机数。

Smote为每个少数类合成相同数量的新样本,这样就存在一些潜在的问题:一方面是增加了类别之间重叠的可能性,另一方面是生成一些没有提供有益信息的样本。为了解决这个问题,Han提出了Borderline-Smote<sup>[2]</sup>方法。

Borderline-Smote为每个少数类样本计算K近邻,但只为其中K近邻中有一半以上多数类样本的少数类样本生成新样本。直观地讲,只为那些周围大部分是多数类样本的少数类样本生成新样本,因为这些样本往往是边界样本。在确定了为哪些少数类样本生成新样本后再利用Smote生成新样本。

总的来说,重采样就是在已有数据不平衡的情况下,人为的让模型学习时接触到的训练数据是类别平衡的。不过由于尾部类的少量数据往往被反复学习,缺少足够多的样本差异,鲁棒性较差;而头部拥有丰富多样性的大量数据又往往得不到充分学习。因此,重采样在非平衡程度相对较小的长尾分布数据集上往往可以取得比较好的效果;但是如果在极端失衡的长尾分布下,这种方法的效果就微乎其微了,所以重采样也并非是个真正完美的解决方案。

#### 2.2 重加权

在直觉上,我们认为数据越多越好,但是由于数据之间存在信息重叠,随着样本数量的增加,模型可以从数据中提取的有益信息逐渐减少。基于此观察,Cui等提出了一种考虑数据重叠来帮助量化有效样本数的方法,它与每个类别的有效样本数成反比来重新加权损失<sup>[3]</sup>。

给定一个类,将其特征空间中所有可能的数据集表示为S。我们假设S的体积为N,  $N \geq 1$ 。将每个样本表示为S的一个子集,其单位体积为1,并且可能与其他样本重叠。文中定义,样本的有效数量是样本的期望体积。将样本的有效数量(期望体积)表示为 $E_n$ ,其中n是样本数量。可以用一个简单的公式来计算:

$$E_n = (1 - \beta^n) / (1 - \beta) \quad (2)$$

其中 $\beta \in [0, 1]$ ,是一个超参数, $\beta = (N-1)/N$ 。

当N较大时,有效样本数与样本数n趋近相同。这时唯一的原型数N较大,因此不存在数据重叠,每个样本都是唯一的。在另一个极端,如果N=1,这意味着只有一个原型,所以这个类中的所有数据都可以通过数据扩充、转换等由这个原型表示。

为了解决不平衡数据的训练问题,文中还提出了一种类别平衡损失。引入一个与有效样本数成反比的加权因子,该因子与i类样本的有效样本数量成反比: $\alpha_i = 1/E_{n_i}$ 。类别平衡损失为:

$$\mathcal{L}(p, y) = \frac{1}{E_{n_y}} \mathcal{L}(p, y) = \frac{1 - \beta}{1 - \beta_{n_y}} \mathcal{L}(p, y) \quad (3)$$

其中 $n_y$ 是真实类别y中的样本数。 $\beta = 0$ 对应没有重新加权损失,而 $\beta \rightarrow 1$ 对应通过相反类别频率重新加权损失。

文中在大规模数据集上进行了对比实验,证实了类别平衡损失比Softmax损失、Sigmoid损失、Focal损失<sup>[4]</sup>等有了显著的性能提升。但是文中的方法现在还做不到给每个类别都找到超参数。接下来,文中计划通过在数据分布中纳入合理的假设或设计,用基于自适应的方法来扩展当前方法的框架。

相对而言,我比较喜欢纯粹的重加权相关的拓展研究,通过对损失函数的改进来解决长尾识别问题。它们大部分实现简单,往往只需修改下损失函数,就可以取得非常具有竞争力的结果。但是重加权方法也是具有局限性的,因为它以牺牲深度网络的特征学习模块的性能为代价,显著提升了深度网络的分类器学习模块的性能,所以重加权也并非是个真正完美的解决方案。

#### 2.3 迁移学习

2019年,UC Berkeley的研究人员们深入研究了视觉识别问题的背景和设定,重新定义了开放长尾识别问题,通过融合不平衡分类、少样本学习和开放集识别三方面,大幅度提升长尾数据识别的表现<sup>[5]</sup>。长尾识别不仅要处理封闭世界中的不平衡分类和少样本学习,而且还要处理开放集识别。基于此观察,Liu等提出了一种集成的长尾识别模型,来处理开放集识别。

文中提出的长尾识别模型主要由两个模块组成:动态元嵌入和调节注意力。前者在头部类和尾部类之间建立联系并传递知识,而后者则在头部类和尾部类之间保持区别。由于长尾分布的头部类数据足够丰富,因此特征提取器一般可以很好地提取出头部类的特征。而尾部类严重缺乏数据,导致特征提取器无法很好地提取出这些类别的特征。

为了解决尾部类问题,文中引入了一个记忆模块,里面存储着每个类的原

型。通过类似中心损失的方式考虑了类别内与类别间信息构造出来的。当用特征提取器提取特征时,会从记忆模块中借鉴一些有用的信息来扩充当前的特征表示。

文中还提出了用概念选择器来控制融合特征的数量和种类。由于头部类别已经具有丰富的数据,只需小部分特征用于针对它们的融合。而对于尾部类别来说,数据较少所以记忆特征中对它们有着较大的提升作用。

文中通过实验发现,对特征图添加空间注意力可以进一步增强特征的判别能力。对于不同类别的图像,具有判别能力的信息往往分布在图片的不同位置上,如果能自适应地给出一个注意力更关注这些位置,或许会使最终学到的特征更适合分类这样。所以为了让模型能具备这种自适应的提取判别特征的效果,文中提出了调节注意力,它由图像上的自注意力和空间注意力构成。

文中将这一模型应用到ImageNet-LT、Places-LT数据集上,比Focal损失<sup>[4]</sup>、距离损失<sup>[6]</sup>、Plain Model<sup>[7]</sup>等有了显著的性能提升。

基于开放长尾识别问题的视觉任务更适合于数据的自然分布,能够更准确贴切地描述真实状况,将会为目标检测、分割和强化学习带来新的提升。

对于迁移学习方法而言,从头部类中学习通用知识,然后迁移到尾部类中的这种理念和实际效果都非常好。但是这类方法,它们大部分实现起来比较困难,因为它们往往还需要设计额外的比较复杂的模块。不过目前的偏好也并非绝对,也许未来我们可以设计出简单有效的迁移模型。

#### 2.4 解耦特征学习和分类器学习

2020年,Zhou等首次揭示了重采样和重加权这类类别重新平衡方法,其奏效的原因实际在于它们显著提升了深度网络的分类器学习的性能。但是,它们也意外地损害了所学习的深层特征的表示能力<sup>[8]</sup>。

基于这一观察,Zhou等提出了一个双边分支网络(BBN),以同时兼顾特征学习和分类器学习。将深度模型的这两个重要模块进行解耦,从而保证两个模块互不影响,共同达到优异的收敛状态,协同促进深度网络在长尾数据分布上的泛化性能。

BBN主要由三部分组成:常规学习分支、再平衡分支、累计学习策略。常规学习分支和再平衡分支分别用于特征学习和分类器学习。这两个分支使用了同样的残差网络结构,除最后一个残差模块,两个分支的网络参数是共享的。文中为这两个分支分别配备了均匀采样器和逆向采样器,得到两个样本 $(x_c, y_c)$ 和 $(x_r, y_r)$ 作为输入数据,其中前者用于常规学习分支,后者用于再平衡分支;

然后,将这两个样本送入各自对应的分支后,通过残差网络和全局平均池化(GAP)得到特征向量 $f_c$ 和 $f_r$ ;最后,通过使用一个自适应权衡参数 $\alpha$ 来控制

$f_c$ 和 $f_r$ 的权重,将加权特征向量 $\alpha f_c$ 和 $(1-\alpha) f_r$ 分别发送到分类器 $W_c$ 和 $W_r$ ,然后再通过逐元素累加的方式将其输出整合到一起。

经实验验证,BBN在多个长尾分布的标准数据集(iNaturalist2017/2018、CIFAR-10-LT和CIFAR-100-LT)上均取得了目前最佳的视觉识别性能,彻底改善了长尾任务的识别性能。

在深度学习中,特征学习和分类器学习通常被耦合在一起进行端到端的模型训练。但在长尾分布数据的极度不平衡因素影响下,特征学习和分类器学习的效果均会受到不同程度干扰。解耦特征学习和分类器学习这类方法,将深度模型这两个模块进行解耦,从而保证两个模块互不影响,共同达到优异的收敛状态,协同促进深度网络在长尾数据分布上的泛化性能。这种方法简单有效,

### 3 总结与展望

本文对解决长尾识别问题的若干方法进行了介绍和讨论,之前的一些传统的非平衡学习方法也可以用来解决长尾识别问题。例如,重采样和重加权方法,都可以用来解决长尾识别问题。这些方法可以直接影响深度网络中分类器权重的更新,从而促进了分类器的学习。但是,它们也在一定程度上损害所学习深层特征的表示能力。所以本文还介绍了一些迁移学习、解耦特征学习和分类器学习的方法。对于长尾识别问题的研究,我认为在以下方面有待进一步的探索:生成对抗网络和对抗样本在最近几年中取得了长足的进步。我们能否利用对抗本来提高生成式对抗网络生成的图片的多样性,对生成对抗网络添加小的扰动来生成多样性更高的数据,从而丰富尾部类的多样性。

目前,长尾识别处理的都是封闭世界中的不平衡分类和少样本学习,针对开放集识别的方法还很少。当我们考虑到整体时,性能就会大大下降。我们需要将新提出的方法应用于更多不同的长尾分布数据集,将结果与多类的性能评价指标进行比较,将有助于为长尾识别的未来应用确定首选的深度学习方法。

目前,我们用在长尾识别问题上的评价指标大多都是多标签分类的性能评价指标,因此提出专门用在长尾识别问题上的评价指标就是未来我们必须要考虑的了。

### 4 参考文献

- Chawla N V, Bowyer K W, Hall L O, et al. Smote: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002: 321–357.
- Han Hui, Wang Wen-Yuan, and Mao Bing-Huan. Borderline-smote: a new over-sampling method in imbalanced data sets learning[C]. International conference on intelligent computing, Hefei, 2005: 878–887.
- Cui Yin, Jia Menglin, Lin Tsung-Yi, et al. Class-Balanced Loss Based on Effective Number of Samples[C]. In IEEE Conference on Computer Vision and Pattern Recognition, Los Angeles, 2019: 9268-9277.
- Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. In International Conference on Computer Vision, Venice, 2017: 2980–2988.
- Liu Ziwei, Miao Zhongqi, Zhan Xiaohang, et al. Large-Scale Long-Tailed Recognition in an Open World[C]. In IEEE Conference on Computer Vision and Pattern Recognition, Los Angeles, 2019: 2532-2541.
- Zhang Xiao, Fang Zhiyuan, Wen Yandong, et al. Range Loss for Deep Face Recognition with Long-Tailed Training Data[C]. In IEEE International Conference on Computer Vision, Venice, 2017, pp: 5419-5428.
- He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. In IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- Zhou Boyan, Cui, Quan, Wei Xiu-Shen, et al. BBN: Bilateral-Branch Network with Cumulative Learning for Long-Tailed Visual Recognition[C]. In Computer Vision and Pattern Recognition, Seattle, 2020: 9716-9724.

投稿人 姓名: 张明

单位: 河北大学