

小样本学习研究综述^{*}

赵凯琳^{1,2}, 靳小龙^{1,2}, 王元卓^{1,2}



¹(中国科学院 计算技术研究所 网络数据科学与技术重点实验室, 北京 100190)

²(中国科学院大学 计算机与控制学院, 北京 100049)

通讯作者: 赵凯琳, E-mail: zhaokailin17z@ict.ac.cn

摘 要: 小样本学习旨在通过少量样本学习到解决问题的模型. 近年来, 在大数据训练模型的趋势下, 机器学习和深度学习在许多领域中取得了成功. 但是在现实世界中的很多应用场景中, 样本量很少或者标注样本很少, 而对大量无标签样本进行标注工作将会耗费很大的人力. 所以, 如何用少量样本进行学习就成为目前人们需要关注的问题. 系统地梳理了当前小样本学习的相关工作, 具体来说介绍了基于模型微调、基于数据增强和基于迁移学习这 3 大类小样本学习模型与算法的研究进展; 将基于数据增强的方法细分为基于无标签数据、基于数据合成和基于特征增强这 3 类, 将基于迁移学习的方法细分为基于度量学习、基于元学习和基于图神经网络这 3 类; 总结了目前常用的小样本数据集和代表性的小样本学习模型在这些数据集上的实验结果; 随后对小样本学习的现状和挑战进行了概述; 最后展望了小样本学习的未来发展方向.

关键词: 小样本学习; 微调; 数据增强; 迁移学习; 度量学习; 元学习

中图法分类号: TP18

中文引用格式: 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述. 软件学报, 2021, 32(2): 349–369. <http://www.jos.org.cn/1000-9825/6138.htm>

英文引用格式: Zhao KL, Jin XL, Wang YZ. Survey on few-shot learning. Ruan Jian Xue Bao/Journal of Software, 2021, 32(2): 349–369 (in Chinese). <http://www.jos.org.cn/1000-9825/6138.htm>

Survey on Few-shot Learning

ZHAO Kai-Lin^{1,2}, JIN Xiao-Long^{1,2}, WANG Yuan-Zhuo^{1,2}

¹(Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

²(School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Few-shot learning is defined as learning models to solve problems from small samples. In recent years, under the trend of training model with big data, machine learning and deep learning have achieved success in many fields. However, in many application scenarios in the real world, there is not a large amount of data or labeled data for model training, and labeling a large number of unlabeled samples will cost a lot of manpower. Therefore, how to use a small number of samples for learning has become a problem that needs to be paid attention to at present. This paper systematically combs the current approaches of few-shot learning. It introduces each kind of corresponding model from the three categories: fine-tune based, data augmentation based, and transfer learning based. Then, the data augmentation based approaches are subdivided into unlabeled data based, data generation based, and feature augmentation based approaches. The transfer learning based approaches are subdivided into metric learning based, meta-learning based, and graph neural network based methods. In the following, the paper summarizes the few-shot datasets and the results in the experiments of the

* 基金项目: 国家重点研发计划; 国家自然科学基金(U1836206, 61772501, 61572473, 61572469)

Foundation item: National Key Research and Development Program of China; National Natural Science Foundation of China (U1836206, 61772501, 61572473, 61572469)

收稿时间: 2019-10-09; 修改时间: 2020-01-01; 采用时间: 2020-02-17; jos 在线出版时间: 2020-09-10

aforementioned models. Next, the paper summarizes the current situation and challenges in few-shot learning. Finally, the future technological development of few-shot learning is prospected.

Key words: few-shot learning; fine-tune; data augmentation; transfer learning; metric learning; meta-learning

随着大数据时代的到来,深度学习模型已经在图像分类、文本分类等任务中取得了先进成果.但深度学习模型的成功,很大程度上依赖于大量训练数据.而在现实世界的真实场景中,某些类别只有少量数据或少量标注数据,而对无标签数据进行标注将会消耗大量的时间和人力.与此相反,人类只需要通过少量数据就能做到快速学习.例如,一个五六岁的小孩子从未见过企鹅,但如果给他看过一张企鹅的图像,当他进入动物园看到真正的企鹅时,就会马上认出这是自己曾经在图像上见过的“企鹅”,这就是机器学习和人类学习之间存在的差距.受到人类学习观点的启发^[1],小样本学习(few-shot learning)^[2,3]的概念被提出,使得机器学习更加靠近人类思维.

早在20世纪八九十年代,就有一些研究人员注意到了单样本学习(one-shot learning)的问题.直到2003年,Li等人^[4]才正式提出了单样本学习的概念.他们认为:当新的类别只有一个或几个带标签的样本时,已经学习到的旧类别可以帮助预测新类别^[5].小样本学习也叫做少样本学习(low-shot learning)^[6],其目标是从少量样本中学习到的解决问题的方法.与小样本学习相关的概念还有零样本学习(zero-shot learning)^[7]等.零样本学习是指在没有训练数据的情况下,利用类别的属性等信息训练模型,从而识别新类别.

小样本学习的概念最早从计算机视觉(computer vision)^[8]领域兴起,近几年受到广泛关注,在图像分类任务中已有很多性能优异的算法模型^[9-11].但是在自然语言处理领域(natural language processing)^[12]的发展较为缓慢,原因在于图像和语言特性不同.图像相比文本更为客观,所以当样本数量较少时,图像的特征提取比文本更加容易^[13].不过近年来,小样本学习在自然语言处理领域也有了一些研究和发展^[14-16].根据所采用方法的不同,本文将小样本学习分为基于模型微调、基于数据增强和基于迁移学习这3种.基于模型微调的方法首先在含有大量数据的源数据集上训练一个分类模型,然后在含有少量数据的目标数据集上对模型进行微调.但这种做法可能导致模型过拟合,因为少量数据并不能很好地反映大量数据的真实分布情况.为解决上述过拟合的问题,基于数据增强和基于迁移学习的小样本学习方法被提出.基于数据增强的方法是利用辅助数据集或者辅助信息增强目标数据集中样本的特征或扩充对目标数据集,使模型能更好地提取特征.本文根据学习方法的不同,将基于数据增强的小样本学习方法进一步细分为基于无标签数据、基于数据合成和基于特征增强这3类方法.基于迁移学习的方法是当前比较前沿的方法,是指将已经学会的知识迁移到一个新的领域中.本文根据学习框架,将基于迁移学习的方法细分为基于度量学习、基于元学习和基于图神经网络(graph neural network)的方法.在度量学习的框架下,目前已有许多性能较好的小样本学习模型,例如比较著名的原型网络(prototypical network)^[9]和匹配网络(matching network)^[17]等.基于元学习的方法不仅在目标任务上训练模型,并且从许多不同的任务中学习元知识,当一个新的任务到来时,利用元知识调整模型参数,使模型能够快速收敛.近年来,随着图神经网络的兴起,研究者将图神经网络也应用到小样本学习中,取得了先进的结果.

除了图像分类和文本分类这两个主要任务,许多其他任务也面临着小样本问题.在计算机视觉应用中,利用小样本学习进行人脸识别^[8,18,19]、食品识别^[20]、表情识别^[21]、手写字体识别^[22,23]以及其他的图像识别^[24].在自然语言处理应用中,使用小样本方法实现对话系统^[25]、口语理解^[26],或者完成NLP的基本任务,例如word embedding^[27].在多媒体领域应用中,可以使用小样本方法实现影像提取^[28]和声纹识别^[29]等.在生物与医学领域,可以应用于疾病诊断^[30,31]、临床实验^[32,33]、护士能力评价^[34]、农作物病害识别^[35,36]、水量分析^[37]等.在经济领域,可应用于产品销量预测^[38]等.在工业与军事领域,可应用于齿轮泵寿命预测^[39]、军事目标识别^[40]和目标威胁评估^[41]等.

本文首先从基于模型微调、基于数据增强和基于迁移学习这3种方法介绍小样本学习的研究进展,总结小样本学习的几个著名数据集以及已有模型在这些数据集上的实验结果;接下来,本文对小样本学习的研究现状和主要挑战进行总结;最后展望了未来的发展趋势.

1 基于模型微调的小样本学习

基于模型微调的方法是小样本学习较为传统的方法,该方法通常在大规模数据上预训练模型,在目标小样本数据集上对神经网络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.若目标数据集和源数据集分布较类似,可采用模型微调的方法.

为了使微调后的小样本分类模型取得较好的效果,使用何种微调方法需要被考虑.Howard 等人^[14]在 2018 年提出了一个通用微调语言模型(universal language model fine-tuning,简称 ULMFit).与其他模型不同的是,此方法使用了语言模型而非深度神经网络.该模型分为 3 个阶段:(1) 语言模型预训练;(2) 语言模型微调;(3) 分类器微调.该模型的创新点在于改变学习速率来微调语言模型,主要体现在两个方面.

- 1) 传统方法认为,模型每一层学习速率相同;而 ULMFit 中,语言模型的每一层学习速率均不相同.模型底层表示普遍特征,这些特征不需要很大调整,所以学习速率较慢;而高层特征更具有独特性,更能体现出任务和数据的独有特征,于是高层特征需要用更大的学习速率学习.总体看来,模型底层到最高层学习速率不断加快.
- 2) 对于模型中的同一层,当迭代次数变化时,自身学习率也会相应地产生变化.作者提出了斜三角学习率的概念,当迭代次数从 0 开始增加时,学习速率逐渐变大;当迭代次数增长到某个固定值时,此时已经学习到了足够知识,固定值之后的学习率又开始逐步下降.

论文从纵向和横向两个维度学习速率的变化对语言模型进行微调,让模型更快地在小样本数据集上收敛;同时,让模型学习到的知识更符合目标任务.另外,Nakamura 等人^[42]提出了一种微调方法,主要包含以下几个机制:(1) 在小样本类别上再训练的过程使用更低的学习率;(2) 在微调阶段使用自适应的梯度优化器;(3) 当源数据集和目标数据集之间存在较大差异性时,可以通过调整整个网络来实现.

基于模型微调的方法较简单,但是在真实场景中,目标数据集和源数据集往往并不类似,采用模型微调的方法会导致模型在目标数据集上过拟合.为解决模型在目标数据集上过拟合的问题,两种解决思路被提出:基于数据增强和基于迁移学习的方法.这两种方法将在接下来的两个章节中依次介绍.

2 基于数据增强的小样本学习

小样本学习的根本问题在于样本量过少,从而导致样本多样性变低.在数据量有限的情况下,可以通过数据增强(data augmentation)^[43]来提高样本多样性.数据增强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.数据扩充是向原有数据集添加新的数据,可以是无标签数据或者合成的带标签数据;特征增强是在原样本的特征空间中添加便于分类的特征,增加特征多样性.基于上述概念,本文将基于数据增强的方法分为基于无标签数据、基于数据合成和基于特征增强的方法三种.接下来,就这 3 种方法分别介绍小样本学习的进展.

2.1 基于无标签数据的方法

基于无标签数据的方法是指利用无标签数据对小样本数据集进行扩充,常见的方法有半监督学习^[44,45]和直推式学习^[46]等.半监督学习是机器学习领域研究的重点问题^[47],将半监督方法应用到小样本学习现在已经有了许多尝试.2016 年,Wang 等人^[48]在半监督学习的思想下,同时受到 CNN 可迁移性的启发,提出利用一个附加的无监督元训练阶段,让多个顶层单元接触真实世界中大量的无标注数据.通过鼓励这些单元学习无标注数据中低密度分离器的 diverse sets,捕获一个更通用的、更丰富的对视觉世界的描述,将这些单元从与特定的类别集的联系中解耦出来(也就是不仅仅能表示特定的数据集).作者提出了一个无监督的 margin 最大化函数来联合估计高密度区域的影响并推测低密度分离器.低密度分离器(LDS)模块可以插入任何标准的 CNN 架构的顶层.除此之外,Boney 等人^[49]在 2018 年提出使用 MAML^[11]模型来进行半监督学习,利用无标签数据调整嵌入函数的参数,用带标签数据调整分类器的参数.MAML 算法本文将在第 3.2 节详细介绍.2018 年,Ren 等人^[50]在原型网络^[9]的基础上进行改进,加入了无标注数据,取得了更高的准确率.此模型本文将在第 3.1 节进行详细介绍.

直推式学习可看作半监督学习的子问题.直推式学习假设未标注数据是测试数据,目的是在这些未标记数据上取得最佳泛化能力.Liu 等人^[51]使用了直推式学习的方法,在 2019 年提出了转导传播网络(transductive propagation network)来解决小样本问题.转导传播网络分为 4 个阶段:特征嵌入、图构建、标签传播和损失计算.该模型在特征嵌入阶段,将所有的标注数据和无标注数据通过嵌入函数 f 映射到向量空间中;在图构建阶段,使用构建函数 g 将嵌入向量构建为无向图中的节点,连边权重由两个节点计算高斯相似度得到;随后,根据公式 $F^*=(I-\alpha S)^{-1}Y$ 来进行标签传播(其中, F^* 是标签预测结果, S 是归一化之后的连边权重, Y 是初始标签的矩阵),让标签从标注数据传播到无标注数据;最后,通过交叉熵函数计算损失,用反向传播更新嵌入函数和构建函数的参数.另外,Hou 等人^[52]也提出了一个交叉注意力网络(cross attention network),基于直推式学习的思想,利用注意力机制为每对类特征和查询生成交叉注意映射对特征进行采样,突出目标对象区域,使提取的特征更具鉴别性.其次,提出了一种转换推理算法,为了缓解数据量过少的问题,迭代地利用未标记的查询集以增加支持集,从而使类别特性更具代表性.

2.2 基于数据合成的方法

基于数据合成的方法是指为小样本类别合成新的带标签数据来扩充训练数据,常用的算法有生成对抗网络(generative adversarial net)^[53]等.Mehrotra 等人^[54]将 GAN 应用到小样本学习中,提出了生成对抗残差成对网络(generative adversarial residual pairwise network)来解决单样本学习问题.算法使用基于 GAN 的生成器网络对不可见的分布提供有效的正则表示,用残差成对网络作为判别器来度量成对样本的相似性,如图 1 的流程图所示.

- 输入样本为 x_t ,为了防止生成器简单复制原样本,首先对 x_t 进行破坏得到损坏样本 \tilde{x}_t ,将 \tilde{x}_t 输入到生成器 G 中得到生成样本 x , x 的另一个来源是真实的数据集.
- 其次, x_t 空间变换后输入到判别器 D 中,同时输入的还有样本 x .
- 随后,判别器会给出 3 种判别结果:(1) x 是真实样本,并且 x 与 x_t 不同;(2) x 是真实样本,并且 x 与 x_t 相似;(3) x 是假样本.

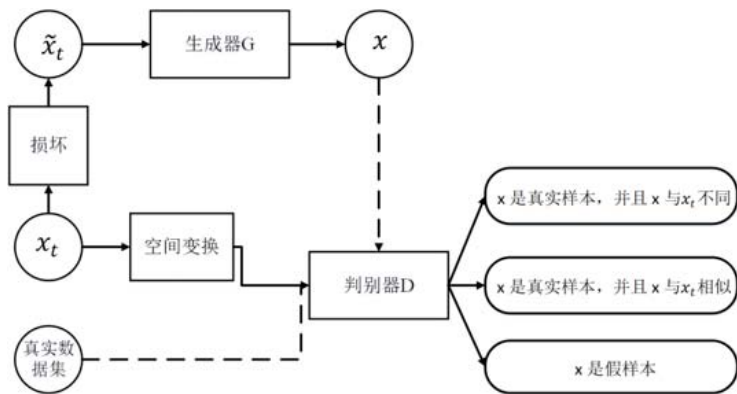


Fig.1 An indicative figure of generative adversarial residual pairwise network^[54]

图 1 生成对抗残差成对网络示意图^[54]

除了基于 GAN 的数据合成方法,Hariharan 等人^[55]提出了一种新的方法,该方法分为两个阶段:表示学习阶段和小样本学习阶段.

- 表示学习阶段是指在含有大量数据的源数据集上学习一个通用的表示模型,在此阶段中,作者提出一个新的损失函数——平方梯度量级损失(SGM) $L_D^{SGM}(\phi, W) = \frac{1}{|D|} \sum_{(x,y) \in D} \alpha(W, (x,y)) \|\phi(x)\|^2$ 来提升表示学习的效果.

- 小样本学习阶段是指在少量数据的新类别中微调模型,在此阶段中,本文提出了生成新数据的方法类来为小样本类别进行数据增强.

作者认为,属于同一类别的两个样本之间存在着一个转换.那么给定新类别的一个样本 x ,通过这个转换,生成器 G 可以生成属于该类别的新样本.

Wang 等人^[56]将元学习与数据生成相结合,提出了通过数据生成模型生成虚拟数据来扩充样本的多样性,并结合当前比较先进的元学习方法,通过端到端方法共同训练生成模型和分类算法.通过让现有图像的一些属性和特征发生变化,如拍照姿态改变、光照改变、位置迁移等,迁移到新的样本上,从而生成具有不同变化的新样本图像,实现数据的扩充.此外,本模型可以套用任何元学习模型,具有灵活性.

但是现有的数据生成方法具有以下缺点:(1) 没有捕捉到复杂的数据分布;(2) 不能泛化到小样本的类别;(3) 生成的特征不具有可解释性.Xian 等人^[57]为解决上述问题,将变分编码器(VAE)和 GAN 进行结合,充分利用了两者的优势集成了一个新的网络 f-VAEGAN-D2.这个网络再完成小样本学习图像分类的同时,能够将生成样本的特征空间通过自然语言的形式表现出来,具有可解释性.Chen 等人^[58]对此继续研究,提出可以利用元学习对训练集的图像对支持集进行插值,形成扩充的支持集集合:首先,从元训练集的集合中每一类随机选择几个样本,形成集合 G ;其次,针对某个任务提取支持集的特征,形成最近邻分类器,对集合 G 中的所有图像分类,找到概率最高的 N 个图像,将对应的图像加权得到扩充的图像,图像标签与原图像保持一致;最后,用扩充的支持集与查询样本计算分类损失,用来优化权重生成子网络.

2.3 基于特征增强的方法

以上两种方法都是利用辅助数据来增强样本空间,除此之外,还可通过增强样本特征空间来提高样本的多样性,因为小样本学习的一个关键是如何得到一个泛化性好的特征提取器.Dixit 等人^[59]提出了 AGA(attributed-guided augmentation)模型学习合成数据的映射,使样本的属性处于期望的值或强度.然而,将基于合成数据的网络应用到真实图像中具有迁移学习的问题,但之前的方法都不适用于具有姿态物体的迁移,AGA 是一个解决方法,但是它的轨迹是离散的,不能连续.所以基于此,Liu 等人^[60]提出了特征迁移网络(FATTEN),用于描述物体姿态变化引起的运动轨迹变化.与其他特征提取不同的是,该方法对物品的外观和姿态分别有一个预测器.网络包括一个编码器和一个解码器,编码器将 CNN 对目标图像的特征 x 映射为一对外观 $A(x)$ 和姿态 $P(x)$ 参数,然后,解码器需要这些参数产生相应的特征向量 x .

此外,Schwartz 等人^[61]提出了 Delta 编码器,通过看到少量样本来为不可见的类别合成新样本,将合成样本用于训练分类器.该模型既能提取同类训练样本之间可转移的类内变形,也能将这些增量应用到新类别的小样本中,以便有效地合成新类样本.但此方法的特征增强过于简单,无法显著改善分类边界^[1].为解决这个问题,Chen 等人^[62]提出了一个双向网络 TriNet,他们认为,图像的每个类别在语义空间中具有更丰富的特征,所以通过标签语义空间和图像特征空间的相互映射,可以对图像的特征进行增强.该模型用一个 4 层卷积神经网络 ResNet-18 提取图像的视觉特征,再通过 TriNet 的编码器将视觉特征映射到语义空间,在语义空间中进行数据增强;之后,通过 TriNet 的解码器将增强后的语义特征映射回图像的特征空间.例如,shark 这个类别在语义空间中与已知的一些类别 fish,whale_shark 和 halobios 距离较近,即代表它们的语义相似,TriNet 可借用这些邻近类别的特征来调整网络中用于特征提取的参数,帮助模型更好地提取图像特征.

但是在上面的方法中,分类网络通常提取的特征只关注最具有判别性的区域,而忽略了其他判别性较弱的区域,不利于网络的泛化.为了解决这个问题,Shen 等人^[63]提出可以把固定的注意力机制换成不确定的注意力机制 M .输入的图像经提取特征后进行平均池化,分类得到交叉熵损失 l .用 l 对 M 求梯度,得到使 l 最大的更新方向从而更新 M .其次,将提取的特征与更新后的 M 相乘,得到对抗特征,得到分类损失 l_1 ;将初始特征再经过多个卷积,得到一维特征对其分类,得到分类损失 l_2 .这两个分类器共享参数,从而使得高层特征对底层特征具有一定的指导作用,从而优化网络.

通过梳理基于数据增强的小样本学习模型的研究进展,可以思考未来的两个改进方向.

- 1) 更好地利用无标注数据.由于真实世界中存在着大量的无标注数据,不利用这些数据会损失很多信

息,更好、更合理地使用无标注数据,是一个非常重要的改进方向.

- 2) 更好地利用辅助特征.小样本学习中,由于样本量过少导致特征多样性降低.为提高特征多样性,可利用辅助数据集或者辅助属性进行特征增强,从而帮助模型更好地提取特征来提升分类的准确率.

3 基于迁移学习的小样本学习

迁移学习是指利用旧知识来学习新知识,主要目标是将已经学会的知识很快地迁移到一个新的领域中^[64].举例说明:一个程序员在掌握了 C 语言的前提下,能够更快地理解和学习 Python 语言.迁移学习主要解决的一个问题是小样本问题.基于模型微调的方法在源数据集和目标数据集分布大致相同时有效,分布不相似时会导致过拟合问题.迁移学习则解决了这个问题.迁移学习只需要源领域和目标领域存在一定关联,使得在源领域和数据中学习到的知识和特征能够帮助在目标领域训练分类模型,从而实现知识在不同领域之间的迁移.一般来说,源领域和目标领域之间的关联性越强,那么迁移学习的效果就会越好^[65].近年来,迁移学习这个新兴的学习框架受到了越来越多研究人员的关注,很多性能优异的小样本算法模型被提出.在迁移学习中,数据集被划分为 3 部分:训练集(training set)、支持集(support set)和查询集(query set).其中,训练集是指源数据集,一般包含大量的标注数据;支持集是指目标领域中的训练样本,包含少量标注数据;查询集是目标领域中的测试样本.

随着深度学习的发展,深度神经网络常被用来学习样本的嵌入函数,如卷积神经网络(convolutional neural network)^[66]、循环神经网络(recurrent neural network)^[67]和长短期记忆网络(long short-term memory)等.在图像分类任务中,常使用卷积神经网络作为嵌入网络,比较常用的变型包括 VGG, Inception, Resnet 等. Wang 等人^[68]基于迁移学习的思想,在 2016 年提出了回归网络(regression network)来解决小样本问题.他们认为:一个由少量样本训练的模型和一个由大量样本训练的模型之间存在一个通用的忽略类别的转换 T , 这个转换 T 由回归网络学习得到.通过 T 的转换,可以把由小样本训练得到的效果不佳的模型,映射为由大量样本训练得到的效果较好的模型.两个模型的转换实质上是模型参数的映射,即,将一个模型的权重映射到另一个模型.

近年来,随着迁移学习的兴起,与之相关的模型也不断涌现.但是在以前的迁移学习算法中,源网络中的某一层迁移到目标网络中的某一层是人工给定的,并且是将所有的特征映射都迁移,没有考虑迁移到哪里和迁移多少的问题. Jang 等人^[69]专注于迁移学习中迁移什么(what)和迁移到哪里(where)的问题,提出利用元学习来学习迁移特征映射的权重和迁移层的权重来解决这个问题,同时提出了一步学习的策略,只用一步来适应目标数据集.该方法同时在小样本数据集上进行了实验,但是效果提升没有大规模数据集多.根据迁移学习的方法不同,本文将分为基于度量学习、基于元学习和基于图神经网络的方法这 3 类.接下来就这 3 类方法中典型的算法模型及研究进展进行介绍.

3.1 基于度量学习的方法

在数学概念中,度量指衡量两个元素之间距离的函数,也叫做距离函数^[70].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^[71].在深度学习中,我们通常采用欧氏距离、马氏距离和余弦相似度等^[72,73]作为距离函数.将度量学习的框架应用到小样本学习上,顾名思义,就是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.基于度量学习方法的通用流程如图 2 所示,该框架具有两个模块:嵌入模块和度量模块,将样本通过嵌入模块嵌入向量空间,再根据度量模块给出相似度得分.基于度量学习的方法通常采用 episodic training,是指将数据集分为多个任务进行训练,每个任务从训练集中随机采样 C -way K -shot 的样本,即选出 C 个类别,每个类别含有 K 个样本,通过多次采样构建多个任务.当进行训练时,将多个任务依次输入到模型中,这就是 episodic training.在测试时,一般从剩余的样本中选取一个 batch 来进行测试.

Koch 等人^[74]在 2015 年最先提出使用孪生神经网络(siamese neural network)进行单样本图像识别.孪生神经网络是一种相似性度量模型,当类别数多但每个类别的样本数量少的情况下,可用于类别的识别.孪生神经网络从数据中学习度量,进而利用学习到的度量比较和匹配未知类别的样本,两个孪生神经网络共享一套参数和权重.其主要思想是:通过嵌入函数将输入映射到目标空间,使用简单的距离函数进行相似度计算.孪生神经网络

络在训练阶段最小化一对相同类别样本的损失,最大化一对不同类别样本的损失.该模型使用两个 CNN 提取输入图像的特征,将图像映射成向量.输入是一对样本而不是单个样本,同一类样本标签为 1,不同类为 0;然后,通过交叉熵函数计算损失.对于单样本学习,训练集中每个类别只有一个样本,所以测试集中的每张图像和训练集中的每个样本都组成一个样本对,依次输入到孪生神经网络中,得到每对样本的距离.选取距离最小的训练样本的标签作为测试样本的类别,从而完成分类.

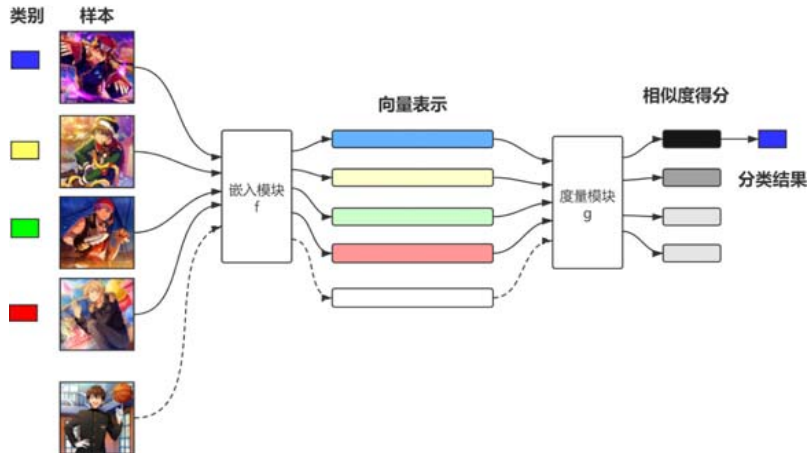


Fig.2 Generalized flow chart of the metric learning based models^[17]

图 2 基于度量学习的模型通用流程图^[17]

Vinyals 等人^[17]继续就单样本学习问题进行深入探讨,在 2016 年提出了匹配网络(matching network),该网络可将带标签的小样本数据和不带标签的样本映射到对应的标签上.针对单样本学习问题,该模型使用 LSTM 将样本映射到低维向量空间中,新样本与每个带标签样本计算相似度,使用核密度估计函数(kernel density estimation)输出预测标签.核密度估计函数专注于从数据样本本身出发来研究数据的分布特征,是在概率论中用来估计未知的密度函数,属于非参数检验方法.该模型在两个方面进行了创新.

- 一是在模型层面提出了匹配网络的概念.匹配网络使用公式 $\hat{y} = \sum_{i=1}^k a(\hat{x}, x_i) y_i$ 计算目标样本与已知标签样本之间的相似度,其中,支持集 $S = \{(x_i, y_i)\}_{i=1}^k$, \hat{x} 是待分类样本, \hat{y} 是待分类样本的预测标签.在此公式中, a 可以看作注意力机制, y_i 看作约束于 x_i 的记忆单元,即匹配网络在计算中引入了外部记忆和注意力机制.
- 二是训练过程中保持训练集和测试集的匹配.这与机器学习中独立同分布的概念相似,在保证训练集和测试集独立同分布的条件下,训练模型在测试集上也能取得较好的效果.

Jiang 等人^[75]基于匹配网络的思想,将嵌入函数改进为 4 层的卷积神经网络,分别采用双向 LSTM 和基于注意力机制的 LSTM 算法深入提取训练样本和测试样本中更加关键和有用的特征并进行编码;最后,在平方欧氏距离上利用 softmax 非线性分类器对测试样本分类.实验结果表明,改进的匹配网络在类别数更多而样本数较少的复杂场景下具有更好的分类效果.

在基于度量学习的单样本图像分类方面,上述方法都是从图像本身特征出发,没有考虑到分类标签这一信息.Wang 等人^[76]认为需要将图像的分类标签纳入考虑,并提出了多注意力网络模型(multi-attention network).该模型使用 GloVe Embedding 将图像的标签嵌入到向量空间,通过构建标签语义特征和图像特征之间的注意力机制,得到一张图像属于该标签的特征主要集中于哪一个部分(单注意力)或哪几个部分(多注意力),利用注意力机制更新该图像的向量,最后通过距离函数计算相似度得到分类结果.

尽管上述模型已经取得了较好的成果,但它们针对的都是单样本学习问题.为了进一步深入解决小样本问

题,Snell 等人^[9]在 2017 年提出了原型网络(prototypical network).作者认为,每个类别在向量空间中都存在一个原型(prototype),也称作类别中心点.原型网络使用深度神经网络将图像映射成向量,对于同属一个类别的样本,求得这一类样本向量的平均值作为该类别的原型.通过不断训练模型和最小化损失函数,使得同一类别内的样本距离更为靠近,不同类别的样本更为远离,从而更新嵌入函数的参数.原型网络的思路如图 3 所示,输入样本 x ,比较 x 的向量和每个类别原型的欧式距离.根据计算发现, x 与原型 2 的距离更近,也就代表 x 与原型 2 所代表的类别更为相似,于是将 x 分到类别 2.

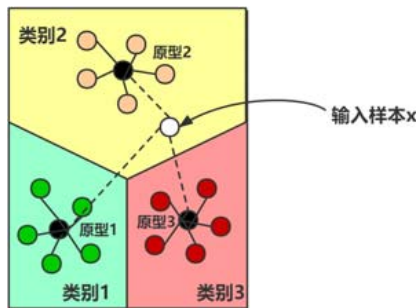


Fig.3 A case study of prototypical networks^[9]

图 3 原型网络样例^[9]

原型网络的想法和实现方法都十分简单明晰,但效果与之前的工作相比得到了一定的提升.但是仅仅使用标注数据得到的结果不一定是准确的,因为样本量太少会导致分类边界偏差.针对这个不足之处,Ren 等人^[50]于 2018 年在原型网络的基础上进行扩展,使用了半监督学习的思想,在训练集中加入了不带标签的数据来改善分类边界.实验证明:无标签数据的加入,提高了分类效果.原因在于:原型网络只使用带标签数据进行原型的计算,但是带标签数据数量较少导致了类别中心的计算不准确;而改进的网络加入了无标签数据之后,对类别的原型进行了修正,使得待分类样本能够得到正确的分类.论文中对这种使用半监督方法的原型网络提出了 3 种变型.

- 第 1 种,所有的无标签数据都属于带标签的数据所属的类别.在这种假设下,将无标签数据和带标签数据一起计算新的原型.
- 第 2 种,无标签数据要么属于带标签数据所属的类别,要么属于一个另外的类——干扰类(distractor class).干扰类开始以原点(0,0)作为原型,模型学习的是干扰类的半径.
- 第 3 种,无标签数据要么属于已知的类别,要么被掩盖(masked).这种变型的提出,是因为第 2 种假设的不合理性.第 2 种假设下,把所有没有标签的数据分到同一个类显然是不符合实际场景的,所以作者又提出了一种新的算法:masked K -means 算法.在这种假设下,论文用一个多层感知机(multi-layer perception,简称 MLP)来学习每个类的半径和斜率,用这两个参数和距离得到一个没有标签的数据属于每个类别的概率.

作者认为:一个样本离原型越远,它就越容易被掩盖.但是上面的网络在计算时均没有考虑样本的权重,只是做了一个简单的平均数计算.在很多情况下,用来计算原型的样本的重要程度是不同的,尤其当样本是噪声数据的时候,体现的更为明显.受到这个问题的驱动,Gao 等人^[77]提出了基于人工注意力的原型网络.相比于传统的原型网络,该模型多了两个部件:样本级别的注意力机制和特征级别的注意力机制,来分别捕捉对分类更重要的样本和特征.作者分别用含有 0%,10%,30%和 50%的噪声数据集进行了效果评测,均比 baseline 取得了更好的效果;并且噪声越多,提升效果越好.说明该模型具有很好的鲁棒性.Sun 等人^[78]也为解决这个问题提出了层次注意力原型网络(HAPN),比起传统的原型网络添加了特征级别、词语级别和样本级别的 3 种注意力机制.词语级别的注意力机制是文本分类中常用的方法,在这里不再赘述.基于人工注意力的原型网络和层次注意力原型网络均添加了样本级别和特征级别的注意力机制,表明不同的样本和特征对于分类任务的重要性确实不同,只做简单的平均计算是远远不够的,需要对样本的特征进行加权处理.

上述模型都是基于距离函数来计算相似度,这样虽然简单易操作,但有时候距离函数却并不适用于一些特定的任务.针对这个问题,一些研究人员提出可以使用深度神经网络来进行度量^[1].Sung 等人^[79]在 2018 年提出了一个新的模型——关系网络(relation network,简称 RN),该模型分为两个模块:嵌入模块和关系模块.其中,嵌入模块 f 是一个 4 层的卷积神经网络,用来学习样本到低维向量空间的嵌入;关系模块 g 是一个相似度比较模块,使用 ReLU 来进行相似度计算,用来输出两个样本的相似度得分.关系网络在 3 种问题上作了讨论.

- 1) 单样本学习.每种类别有一个支持样本 x ,嵌入向量为 $f(x)$,针对查询样本 y 获得嵌入向量 $f(y)$, $C(f(x), f(y))$ 表示两个向量的连接,将这个连接后的向量放到关系模块 g 中,得到相似度打分,完成分类.
- 2) 小样本学习.对于每一类的支持样本,将它们的嵌入向量相加作为整个类别的特征映射.剩下的过程和单样本学习相同.
- 3) 零样本学习.对于没有标注样本的问题,利用每个类别的语义特征嵌入向量 v ,使用新的嵌入函数 f_2 ,得到这个类别的特征映射 $f_2(v)$,剩下过程与上面相同.

在关系网络的基础上,Zhang 等人^[10]提出了深度比较网络(deep comparison network,简称 DCN),将嵌入学习分解为一系列模块,并将每个模块与一个关系模块配对.关系模块利用相应嵌入模块的表示计算一个非线性度量对匹配进行打分.为了保证所有嵌入模块的特征都被使用,关系模块被深度监控.最后,通过学习噪声调节器进一步提高泛化性.Hilliard 等人^[80]也使用了一种新的体系结构,抛弃了传统的度量学习方法,通过训练一个网络来执行类别之间的比较,而不是依赖于静态度量比较.该网络可以决定一个类的哪些方面对于分类比较重要,从而更好地地区分类别边界.

之前的方法注意力集中在一阶统计量的概念表示上,Li 等人^[81]提出了一个协方差度量网络(CovaMNet),在基于小样本分类任务的分布一致性上,利用了协方差表示和协方差矩阵,其中,协方差表示用来捕获二阶统计信息,协方差矩阵用来衡量 query 样本与新类别之间的分布一致性.但由于现有的方法忽略了局部特征的信息,为了捕捉局部特征,Li 等人^[82]又提出了深度最近邻神经网络(DN4).与其他方法的最大不同是:在最后一层用图像到类别的局部描述符来代替图像级别的特征测量,查询样本在进行特征映射时,为每个空间特征计算一个相似性.针对一个查询样本特征映射的每个空间特征,找到支持特征映射中最相近的 K 个特征来计算相似性,最后将所有位置相似性相加,得到此查询样本的相似性.

但是上面的方法是针对每个任务提取不同的特征,仍旧忽略了支持集中所有图像之间的语义关系.受到这个想法的驱动,Li 等人^[83]提出可以利用模型整合支持集中所有图像的信息,从而找到最具有判别性的特征.

- 首先,根据支持集得到一个 channel attention;随后,对所有的图像应用 channel attention,对于支持集中的图像提取特征,经过一个卷积层求得原型;将所有类连接,得到一个特征;再经过卷积,得到一个 channel attention.
- 其次,将支持集特征和查询样本特征经过一个卷积层,与上述得到的 attention 相乘,得到更具有判别性的特征.
- 最后,对于更新后的特征做度量学习.

由上面的模型可以看出:基于度量学习的方法经历了从解决单样本问题到解决小样本问题再到同时解决小样本问题和零样本问题的变革,模型逐渐趋近于成熟;同时,也经历了从基于传统距离函数的方法到基于深度网络的方法的改进.继续采用基于传统距离函数的方法很难在小样本分类准确率方面得到较大的提升,所以加强对基于神经网络进行度量方法的研究,将是今后关注的方向.

3.2 基于元学习的方法

元学习(meta-learning)也叫做学会学习(learning to learn)^[84],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.元知识指在模型训练过程之外可以学习到的知识,比如模型的超参数、神经网络的初始参数、神经网络的结构和优化器等^[85].在小样本学习中,元学习具体指从大量的先验任务中学习得到元知识,利用以往的先验知识来指导模型在新任务(即小样本任务)中更快地学习.元学习中的数据集中通常分为元训练集和元测试集,二

者均包含了原始模型所需要的训练集和测试集.如图 4 所示,分类模型的数据集包括训练集和测试集;元学习模型的数据集包括元训练集和元测试集,其中,元训练集和元测试集均包含训练集和测试集.



Fig.4 Example of meta-learning dataset^[86]

图 4 元学习数据集实例^[86]

早在 2001 年,Hochreiter 等人^[87]就证明了记忆神经网络可适用于元学习.在此工作的基础上,Santoro 等人^[88]在 2016 年提出了基于记忆增强的神经网络(memory-augmented neural networks,简称 MANN)来解决单样本学习问题.作者使用了神经图灵机(neural Turing machine,简称 NTM)^[89]作为 MANN 的基本模型,因为 NTM 作为一种可微的 MANN,可以直接通过梯度下降进行训练.神经图灵机既可以通过缓慢权重更新来实现长期存储,又能够通过记忆库实现短期存储,这与元学习的思想完全一致.作者致力于让神经图灵机学习到一种策略,这个策略可以指导 NTM 将样本类型放入到记忆库中,同时指导它今后如何通过这些类型进行预测或者分类.

MANN 是元学习方法在单样本学习问题上较早且较成功的一次尝试,此后,应用元学习框架解决小样本问题逐渐进入人们的视野.在 2017 年,Munkhdalai 等人^[90]继续采用元学习的框架来解决单样本分类的问题,并提出了一个新的模型——元网络(meta network).元网络主要分为两个部分:base-learner 和 meta-learner,还有一个额外的记忆块,这个记忆块可以帮助模型快速学习.

Base-learner 在任务空间中学习,meta-learner 在抽象的元空间中持续学习并且从不同的任务中获取元知识.当新任务到来时,base-learner 对当前任务进行分析,并将元信息反馈给 meta-learner;Meta-learner 收到元信息之后,根据元信息对自身和 base-learner 快速参数化.具体来说,元网络分为一个缓慢权重化的过程和一个快速权重化的过程,在学习不同任务之间的泛化信息时权重更新缓慢,而当对一个新任务快速适应时,则需要快速权重化.

虽然上面两个工作已经将元学习的框架带进了小样本学习中,但是它们解决的都是单样本问题.为了使模型更加适用到小样本分类问题上,Finn 等人^[11]在 2017 年提出了未知模型的元学习方法(model-agnostic meta-learning,简称 MAML).使用这个模型,从很少的数据中进行少步数的训练,就可以得到较好的分类效果.MAML 首先使用 RNN 从所有任务的分布中来学习知识,而不是仅仅学习单一任务.MAML 致力于找到神经网络中对每个任务较为敏感的参数,通过微调这些参数,让模型的损失函数快速收敛.模型的初始参数为 θ ,针对不同的任

务分别计算损失,根据每个任务的损失更新对应的模型参数.MAML 的目标是求得初始化参数 θ ,使得模型在面对新任务时,能够使用梯度下降的方法在很少的步数内得到收敛.为了避免元学习器的偏移,并且提高元学习器的泛化性,Jamal 等人^[91]提出了算法未知任务元学习法(task-agnostic meta-learning,简称 TAML).在这项工作中,为了避免元学习模型对训练任务过拟合,作者在输出预测时加入了一个正则化项.正则化要么会使预测具有更高的熵(即预测的概率不会看起来像一个独热矢量),要么使模型在不同任务之间的差异更小(即,在不同任务上表现相同).显然,对于小样本学习来说,有一个强大的正则化机制是十分重要的.本文作者在 MAML 的基础上测试了该方法,得到了更优的性能.

MAML 在小样本图像分类任务中已经得到了广泛使用,但是在自然语言处理领域,小样本学习的问题还亟待解决.相比图像分类的任务,文本分类的任务更具有挑战性,这和文本本身的性质有关.例如,使用微博评论做情感分析任务,将会面临拼写错误、缩略词等难题;而针对一张图像,模型只需要提取它的图像特征即可,因为图像是一个客观的实体.由于上述原因,文本分类比图像分类更具有挑战性^[13].不过,近几年文本方面的小样本学习仍取得了一些突破性进展^[14-16].Xiang 等人^[15]在 2018 年将 MAML 迁移到了文本领域,并且在方法中加入了注意力机制,提出了基于注意力机制的未知任务元学习法(attentive task-agnostic meta-learner,简称 ATAML).在文本分类中,不同的词对于分类的重要程度不同.例如,对书的评论进行情感分析的任务中,“我觉得这本书十分有趣”中的“有趣”,对于将这条评论判定为正面情感起着更重要的作用.注意力机制可以将文本中的词语赋予不同的权重,使得对分类起着更重要作用的词语能够有更高的影响力.ATAML 主要分为两个部分:一是忽略任务的表示学习,二是面向任务的注意力学习.第 1 部分是在许多任务中学习一个通用的嵌入模型,将一段文本中的每个词都表示为向量,模型参数为 θ_E ;第 2 部分是针对特定的任务学习特定的参数 $\theta_T, \theta_T = \{\theta_W, \theta_{ATT}\}$,其中, θ_{ATT} 为注意力机制参数, θ_W 为分类器参数.

但是 MAML 存在着一些缺点:一是训练时需要数量足够多的任务才可以收敛;二是这种方法一般只适用于浅层网络,在深层网络中泛化性较差,容易过拟合.Sun 等人^[92]提出:可以让 MAML 只学习最后一层作为分类器,用所有训练数据预训练一个特征提取器,固定特征提取器.该算法利用了 MAML 的思想,随机初始化分类器 W 的参数;然后针对每个任务,利用支持集来优化 W ,得到更新后的 W ;计算查询集的损失,梯度更新 W ,以得到新的 W .Liu 等人^[93]也在 MAML 上做了一些改进,他们认为:对于一个元学习的任务,超参数的设置是十分重要的.可以利用元学习对网络中每一层学习一个超参数,并且在通常情况下,一个分类器具有不稳定性,可以在 MAML 的机制上学习如何融合多个分类器:首先,MAML 内循环更新初始参数多次,得到多个分类器;其次,在 MAML 外循环优化分类器的初始参数、超参数和多个分类器融合系数.其中,测试集上的预测类别为多个分类器预测值加权求和,利用测试集的损失函数更新上述参数.

为了更好地融入语义信息,Wang 等人^[94]提出了任务感知特征嵌入网络(TAFE-Net).在这项工作中,标签嵌入被用来预测数据特征提取模型的权重.该方法通过权重分解,做到只需要预测一个较低维的权重向量,使得权重预测更加简便.此外,该方法还通过嵌入损失使得语义嵌入和图像嵌入对齐.除此之外,还有很多元学习的方法.例如,Ravi 等人^[86]在 2017 年提出了利用优化器的元学习模型进行小样本图像分类,该模型使用基于 LSTM 的元学习器学习优化算法,用优化算法的参数更新规则更新分类器的网络参数,使得分类器在小样本数据上能取得较好的分类效果.同样地,使用上述方法也可以帮助分类器学习到一个较好的初始化参数,使得模型能够在新的小样本数据集上快速收敛.模型的具体流程如下:首先,元学习器给分类器一个初始化参数 θ 将第 1 个 batch 的数据输入分类器进行训练,得到当前的损失和斜率,并将其反馈给元学习器;元学习器根据损失更新模型参数,将更新后的参数传给分类器.按照这个步骤循环迭代.因为元学习器参数更新和 LSTM 细胞状态更新的过程十分相似,所以在此方法中,可以把 LSTM 细胞单元更新的计算方法应用到元学习器的参数更新中.Gidaris 等人^[95]提出了一种方法,该方法包含一个基于注意力机制的权重生成器,同时,在特征表示和类别权重向量之间重新设计一个 CNN 作为余弦相似度函数.先使用训练集训练得到特征提取器;然后对于新的小样本数据,通过一个元学习器来生成对应的参数权重.在含有多个样本时,使用了注意力机制来选择对应的初始权重,而不是做一个简单的平均.

在上述基于元学习的方法中,元学习器从多个任务中学习知识,但是对于不同的任务学习到的模型初始参数是相同的,忽略了不同任务之间的差异性.而在现实世界中,不同任务之间千差万别,基于这个前提,Yu 等人^[16]在 2018 年提出了多任务聚类的元学习法.该模型对所有任务进行聚类,将不同的任务分成不同的簇,同一簇中的任务较为相似,它们共享一套分类器参数.当新任务到来时,计算每个簇的分类器在当前小样本数据上的适应效果,该适应效果由适应参数 α 表示;随后,当前分类任务的模型参数由所有簇的参数与 α 线性组合得到;最后输出分类结果.

通过梳理近年来基于元学习的小样本学习模型,不难看出:随着元学习的兴起,各种元学习方法层出不穷.元学习方法经历了从单样本学习到小样本学习的转变,同时也从图像领域迁移到了文本领域.元学习方法为小样本学习带来了很大突破,到今后很长一段时间都将是小样本学习的主流方法.研究人员可以设计新的元学习器,让分类器在少量样本上学得更快更好.

3.3 基于图神经网络的方法

在计算机科学中,图作为一种数据结构,由点和边构成.图这种数据结构,具有表现力强和展示直观的优点.随着近年来机器学习的兴起,机器学习逐渐被应用到图的分析上.图神经网络是一种基于深度学习的处理图领域信息的模型,由于其较好的性能和可解释性,它最近已成为一种广泛应用的图分析方法^[96].图神经网络有很多种变体,比较常用的有图卷积神经网络(graph convolutional network)、门控图神经网络(gated graph neural network)和图注意力网络(graph attention network)等.

Garcia 等人^[97]在 2018 年使用图卷积神经网络实现小样本图像分类.在图神经网络里,每一个样本被看作图中的一个节点,该模型不仅学习每个节点的嵌入向量,还学习每条边的嵌入向量.卷积神经网络将所有样本嵌入到向量空间中,将样本向量与标签向量连接后输入图神经网络,构建每个节点之间的连边;然后通过图卷积更新节点向量,再通过节点向量不断更新边的向量,这就构成了一个深度的图神经网络.如图 5 所示,5 个不同的节点输入到 GNN 中,根据公式 A 构建边,然后通过图卷积更新节点向量,再根据 A 更新边,再通过一层图卷积得到最后的点向量,最后计算概率.利用公式

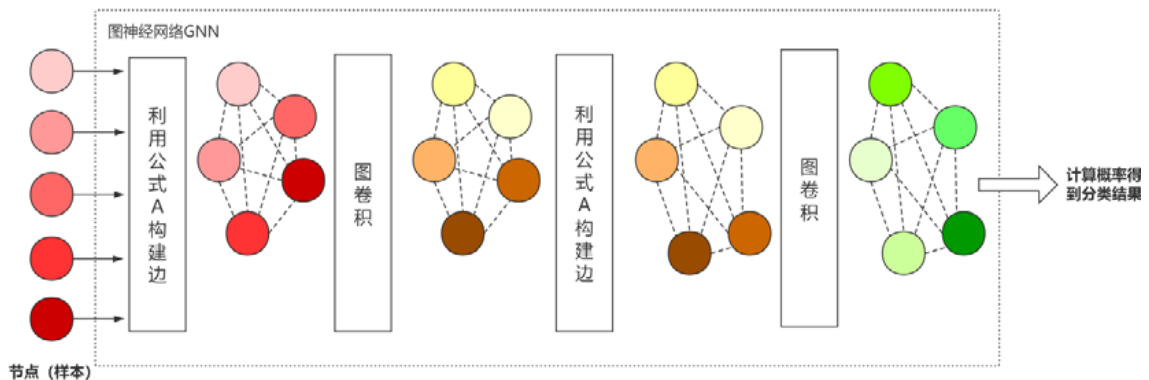


Fig.5 Model process of GNN^[97]

图 5 GNN 模型流程图^[97]

上面的方法是对图中的节点进行分类,Kim 等人^[98]从另一个方面进行考虑,对图中的边进行分类.首先,对图中的边特征向量进行初始化,边的特征向量有两维:第 1 维表示相连的两个节点属于同一类的概率,第 2 维表示它们不属于同一类的概率.随后,用边的特征向量更新节点向量,边的两维特征分别对应节点的类内特征和类间特征.经过多次更新后,对边进行二分类,得到两个节点是否属于同一类.为了对传统的 GNN 进行改进,Gidaris 等人^[99]在 GNN 中加入了降噪自编码器(DAE),以修正小样本类别的权重.DAE 理论指出:对于被高斯噪声干扰的输入,DAE 能够估算其输入 w 的密度的能量函数 $p(w)$ 的梯度.首先,在进行训练时加入部分的高斯噪声数据来防止过拟合,样本经过嵌入网络之后,输入到图神经网络中作为节点.根据未加入高斯噪声的各类初始权值向量

的余弦相似性,将最近的类连接起来,而图的两个节点之间的边的边缘强度(边的权值)就是两个节点的余弦相似性的 softmax 函数值,根据 softmax 函数输出结果.基于图神经网络的方法相比基于度量学习和基于元学习的方法较少,但图神经网络可解释性强并且性能较好,可以思考如何对其进行改进,从而提高分类准确率.

4 数据集与实验

在小样本图像分类任务中,一些标准数据集被广泛使用.单样本学习最常用的是 Omniglot 数据集^[100],小样本学习最常用的数据集是 miniImageNet^[101].除此之外,常用数据集还有 CUB^[102]、tieredImageNet^[50]等,同时, CIFAR-100、Stanford Dogs 和 Stanford Cars 常用作细粒度小样本图像分类.接下来对每个数据集进行简单介绍.

- (1) Omniglot 包含 50 个字母的 1 623 个手写字符,每一个字符都是由 20 个不同的人通过亚马逊的 Mechanical Turk^[103]在线绘制的.
- (2) miniImageNet 是从 ImageNet^[104]分割得到的,是 ImageNet 的一个精缩版本,包含 ImageNet 的 100 个类别,每个类别含有 600 个图像.一般 64 类用于训练,16 类用于验证,20 类用于测试.
- (3) tieredImageNet 是 Mengye 等人^[50]在 2018 年提出的新数据集,也是 ImageNet 的子集.与 miniImageNet 不同的是,tieredImageNet 中类别更多,有 608 种.
- (4) CUB(caltech-UCSD birds)是一个鸟类图像数据集,包含 200 种鸟类,共计 11 788 张图像.一般 130 类用于训练,20 类用于验证,50 类用于测试.
- (5) CIFAR-100 数据集:共 100 个类,每个类包含 600 个图像,分别包括 500 个训练图像和 100 个测试图像. CIFAR-100 中的 100 个子类所属于 20 个父类,每个图像都带有一个子类标签和一个父类标签.
- (6) Stanford Dogs:一般用于细粒度图像分类任务.包括 120 类狗的样本共计 20 580 个图像,一般 70 类用于训练,20 类用于验证,30 类用于测试.
- (7) Stanford Cars:一般用于细粒度图像分类任务.包括 196 类车的样本共计 16 185 个图像,一般 130 类用于训练,17 类用于验证,49 类用于测试.

本文选取一些著名模型在上述数据集上的实验结果进行总结,所有方法都选取了 5-way 1-shot(即 5 个类别,每个类别具有 1 个样本)和 5-way 5-shot(即 5 个类别,每个类别具有 5 个样本)的结果进行对比.图像分类的任务一般使用 CNN 作为嵌入网络,常用的有 VGG,Inception 和 Resnet 等.具体见表 1.

Table 1 Comparison of precision in few-shot learning methods
表 1 小样本学习方法准确率对比

方法	Omniglot		miniImageNet	
	1-shot	5-shot	1-shot	5-shot
MAML ^[11]	98.7±0.4	99.9±0.1	48.70±1.84	63.11±0.92
孪生神经网络 ^[74]	97.3	98.4	—	—
匹配网络 ^[17]	98.1	98.9	43.56±0.84	55.31±0.73
原型网络 ^[9]	97.4	99.3	49.42±0.78	68.20±0.66
改进的原型网络 ^[50]	97.30±0.08	—	50.41±0.31	64.39±0.24
关系网络 ^[79]	99.6±0.2	99.8±0.1	57.02±0.92	71.07±0.69
Dual TriNet ^[62]	—	—	58.12±1.37	76.92±0.69
GNN ^[97]	99.2	99.7	50.33±0.36	66.41±0.63
Meta-Learner LSTM ^[86]	—	—	43.44±0.77	60.60±0.71
TPN ^[51]	—	—	55.51	69.86
CovaMNet ^[81]	—	—	51.19±0.76	67.65±0.63
TAML ^[91]	99.37±0.3	99.77±0.1	49.4±1.9	66.0±0.89
EGNN ^[98]	—	—	—	76.37

表 1 选取了 Omniglot 和 miniImageNet 数据集的实验结果作为对比参考,因为其他的数据集使用次数较少,所以在此不多加讨论.由表 1 可看出,每个数据集中 5-shot 的准确率均比 1-shot 的高.表明训练数据越多,学到的特征也越多,分类效果越好.在 Omniglot 数据集上,所有模型在 1-shot 场景下的准确率都达到了 97%,在 5-shot 任务下的准确率均达到了 98%,可提升空间较少;在 miniImageNet 数据集上,不同模型之间的提升较大,在 1-shot

任务下效果最好的模型比效果最差的准确率提升了 15%左右,在 5-shot 任务下准确率提升了高达 21%,表明在此数据集上还有较大的提升空间。

5 小样本学习总结与展望

5.1 小样本学习总结

由于真实世界的某些领域中样本量很少或标注样本很少,而样本标注工作会耗费大量时间和人力,近年来,小样本学习逐渐成为人们重点关注的问题.本文介绍了图像分类和文本分类两个任务中小样本学习的研究进展,总体上看,小样本图像分类已有了许多性能优异的算法模型,但小样本文本分类仍是个亟待解决的问题.根据小样本学习方法的不同,本文将其分为基于模型微调、基于数据增强和基于迁移学习的方法这 3 类,其中,基于数据增强的方法可以细分为基于无标签数据、基于数据合成和基于特征增强的方法这 3 种,基于迁移学习的方法可以细分为基于度量学习、基于元学习和基于图神经网络的方法这 3 种.本文对以上几种方法做了总结,并且比较了它们的优点和缺点,具体见表 2.

Table 2 Comparison of advantages and disadvantages in different few-learning methods

表 2 小样本学习方法优缺点对比

分类		优点	缺点
基于模型微调		操作简单,只需重新调整模型的参数	在目标数据集和源数据集不类似的情况下,会导致模型在目标数据集上过拟合
基于数据增强	基于无标签数据	不需要对模型进行调整,只需利用辅助数据或者辅助信息扩充数据或增强特征	有可能会引入许多噪声数据或者特征,对分类效果产生负面影响
	基于数据合成		
	基于特征增强		
基于迁移学习	基于度量学习	便于计算和公式化	在样本数量较少的情况下,简单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	使模型具有学习能力,能够学习到一些训练过程之外的知识	复杂度较高,由于最近几年才兴起,需要改进和发展的方面还有很多
	基于图神经网络	性能较好,可解释性强,展示更为直观	当样本总数变大时,图神经网络中边的数量会变多,导致计算复杂度变高

总体来说,小样本学习研究已有很大进展,但和人类分类准确率相比还有很大差距.为了解决基于模型微调方法带来的过拟合问题,基于数据增强和基于迁移学习的方法被提出.基于数据增强的方法是对小样本数据集进行数据扩充或特征增强,这种方法可以不对模型进行参数调整,但是容易引入噪声数据.基于迁移学习的方法是将旧领域学到的知识迁移到新领域,并且不需要两者之间有很强的关联性,但关联性越强,迁移效果越好.在基于迁移学习的方法中,基于度量学习的方法最简单、容易操作,只需要通过距离来衡量样本之间的相似度,但是学习到的知识太少.基于元学习比基于度量学习的方法学习能力更强,可以学习到更多知识.基于图神经网络的方法在 3 种方法中展示最为直观、可解释性较强,但样本总数变大时,会导致计算复杂度增高.

5.2 小样本学习挑战

尽管近年来小样本学习已经得到深入研究,并且取得了一定进展,但仍面临着一些挑战.

(1) 强制的预训练模型^[4]

在已有的小样本学习方法中,不管是基于模型微调的方法还是基于迁移学习的方法,都需要在大量的非目标数据集上对模型进行预训练,致使“小样本学习”一定程度上变成个伪命题.因为模型的预训练依旧需要大量标注数据,从本质上来看,与小样本学习的定义背道而驰.从根本上解决小样本问题,就要做到不依赖预训练模型,可以研究利用其他先验知识而非模型预训练的方法.

(2) 深度学习的可解释性

由于深度学习模型本身是一个黑盒模型,在基于迁移学习的小样本深度学习模型中,人们很难了解到特征迁移和参数迁移时保留了哪些特征,使得调整参数更加困难^[105].提高深度学习的可解释性,能帮助理解特征迁

移,在源领域和目标领域之间发现合适的迁移特征^[106].在此方面已有了一些工作^[107].

(3) 数据集挑战

现有的小样本学习模型都需要在大规模数据上预训练,图像分类任务中已经有了 ImageNet 作为预训练数据集,而文本分类中,缺少类似的预训练数据集,所以需要构建一个被各种任务广泛使用的小样本文本分类预训练数据集.同时,在小样本图像分类任务中,miniImageNet 和 omniglot 是两个被广泛使用的标准数据集.但在小样本文本分类任务中,不同工作所采用的目标数据集千差万别,有很多都是网上爬取的数据集.所以,构建一个适用于文本分类任务的小样本目标数据集是需要考虑的问题.为解决这个问题,Han 等人^[108]在 2018 年提出了一个小样本关系抽取数据集 FewRel,其中包含 100 种分类,共 70 000 个实例,规模和 miniImageNet 相当.

(4) 不同任务之间复杂的梯度迁移^[90]

在基于元学习的小样本学习方法中,从不同任务中学习元知识的过程梯度下降较慢.将模型迁移到新任务中时,由于样本数量较少,所以期望模型能在目标数据集上快速收敛,在此过程中,梯度下降较快.针对基于元学习的方法设计合理的梯度迁移算法,也是目前需要研究并亟待解决的问题.

(5) 其他挑战

在小样本文本分类中,不同语言的分类难度不同.英文的文本分类比较成熟,但是中文分类由于分词等问题,目前还不是很成熟.另外,跨语种或者多语种的文本分类是一个难题,由于源语言与目标语言的特征空间相差甚大,同时,各国的语言、文字又包含了不同的语言学特征,这无疑加大了跨语言文本分类的难度^[105].

5.3 小样本学习展望

通过对当前小样本学习研究进展的梳理,可以展望未来小样本学习的发展方向.

- 1) 在数据层面,尝试利用其他先验知识训练模型,或者更好地利用无标注数据.为了使小样本学习的概念更靠近真实,可以探索不依赖模型预训练、使用先验知识(例如知识图谱)就能取得较好效果的方法.虽然在很多领域中标注样本数量很少,但真实世界中存在的大量无标注数据蕴含着大量信息,利用无标注数据的信息训练模型,这个方向也值得深入研究.
- 2) 基于迁移学习的小样本学习面临着特征、参数和梯度迁移的挑战.为更好地理解哪些特征和参数适合被迁移,需要提高深度学习的可解释性;为使模型在新的领域新任务中快速收敛,需要设计合理的梯度迁移算法.
- 3) 针对基于度量学习的小样本学习,提出更有效的神经网络度量方法.度量学习在小样本学习中的应用已经相对成熟,但是基于距离函数的静态度量方法改进空间较少,使用神经网络来进行样本相似度计算将成为以后度量方法的主流,所以需要设计性能更好的神经网络度量算法.
- 4) 针对基于元学习的小样本学习,设计更好的元学习器.元学习作为小样本学习领域刚兴起的方法,目前的模型还不够成熟.如何设计元学习器,使其学习到更多或更有效的元知识,也将是今后一个重要的研究方向.
- 5) 针对基于图神经网络的小样本学习,探索更有效的应用方法.图神经网络作为这几年比较火热的方法,已经覆盖到很多领域,并且可解释性强、性能好,但是在小样本学习中应用的模型较少.如何设计图网络结构、节点更新函数和边更新函数等方面,值得进一步探究.
- 6) 尝试不同小样本学习方法的融合.现有小样本学习模型都是单一使用数据增强或迁移学习的方法,今后可以尝试将二者进行结合,从数据和模型两个层面同时进行改进,以达到更好的效果.同时,近年来,随着主动学习(active learning)^[109]和强化学习(reinforcement learning)^[110]框架的兴起,可以考虑将这些先进框架应用到小样本学习上.

6 结束语

由于真实世界中样本稀缺的问题,小样本学习越来越受到人们的重视.随着机器学习和深度学习的不断发展,基于小样本学习的分类问题已经在计算机视觉和自然语言处理领域有了深入的研究,并将在医疗领域(如疾

病诊断)、金融领域(如资金异常)等不同领域展现出良好的应用前景.本文对目前小样本学习的研究进展进行了详细阐述,同时分析了目前小样本学习面临的挑战,最后对其进行了前景展望.

References:

- [1] Li XY, Long SP, Zhu J. Survey of few-shot learning based on deep neural network. *Application Research of Computers*, 2020, 37(8):2241–2247 (in Chinese with English abstract).
- [2] Jankowski N, Duch W, Grąbczewski K. *Meta-learning in Computational Intelligence*. Springer Science and Business Media, 2011. 97–115.
- [3] Lake B, Salakhutdinov R. One-shot learning by inverting a compositional causal process. In: *Advances in Neural Information Processing Systems*. 2013. 2526–2534.
- [4] Li FF, *et al.* A Bayesian approach to unsupervised one-shot learning of object categories. In: *Proc. of the 9th IEEE Int'l Conf. on Computer Vision*. 2003.
- [5] Feifei L, Fergus R, Perona P. One-shot learning of object categories. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006, 28(4): 594–611.
- [6] Wang YX, Girshick R, Hebert M, *et al.* Low-shot learning from imaginary data. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 7278–7286.
- [7] Fu Y, Xiang T, Jiang YG, *et al.* Recent advances in zero-shot recognition: Toward data-efficient understanding of visual content. *IEEE Signal Processing Magazine*, 2018, 35(1):112–125.
- [8] Yang J, Liu YL. The latest advances in face recognition with single training sample. *Journal of Xihua University (Natural Science Edition)*, 2014, 33(4):1–5, 10 (in Chinese with English abstract).
- [9] Snell J, Swersky K, Zemel RS. Prototypical networks for few-shot learning. In: *Advances in Neural Information Processing Systems*. 2017. 4077–4087.
- [10] Zhang X, Sung F, Qiang Y, *et al.* Deep comparison: Relation columns for few-shot learning. *arXiv preprint arXiv:1811.07100*, 2018.
- [11] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: *Proc. of the 34th Int'l Conf. on Machine Learning*, Vol.70. 2017. 1126–1135.
- [12] Manning C. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [13] Bailey K, Chopra S. Few-shot text classification with pre-trained word embeddings and a human in the loop. *arXiv preprint arXiv:1804.02063*, 2018.
- [14] Howard J, Ruder S. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.
- [15] Xiang J, Havai M, Chartrand G, *et al.* On the importance of attention in meta-learning for few-shot text classification. *arXiv preprint arXiv:1806.00852*, 2018.
- [16] Yu M, Guo X, Yi J, *et al.* Diverse few-shot text classification with multiple metrics. In: *Proc. of the NAACL-HLT*. 2018. 1206–1215.
- [17] Vinyals O, Blundell C, Lillicrap T, *et al.* Matching networks for one shot learning. In: *Advances in Neural Information Processing Systems*. 2016. 3630–3638.
- [18] Huang ZW, Xie K, Wen C, Sheng GQ, Wen FQ. Small sample face recognition algorithm based on transfer learning model. *Journal of Changjiang University (Natural Science Edition)*, 2019, 16(7):88–94 (in Chinese with English abstract).
- [19] Hu ZP, He W, Wang M, Sun Z, Ren DW. Deep subspace joined sparse representation for single sample face recognition. *Journal of Yanshan University*, 2018, 42(5):409–415 (in Chinese with English abstract).
- [20] Lv YQ, Min WQ, Duan H, Jiang SQ. Few-shot food recognition combining triplet convolutional neural network with relation network. *Computer Science*, 2020, 47(1):136–143 (in Chinese with English abstract).
- [21] Lin KZ, Bai JX, Li HT, Li W. Facial expression recognition with small samples fused with different models under deep learning. *Computer Science and Exploration*, 2020, 14(3):482–492 (in Chinese with English abstract).
- [22] Zhou TY, Zhao L. Research of handwritten Chinese character recognition model with small dataset based on neural network. *Journal of Shandong University of Technology (Natural Science Edition)*, 2019, 33(3):69–74 (in Chinese with English abstract).

- [23] Quan ZN, Lin JJ. Text-independent writer identification method based on Chinese handwriting of small samples. *Journal of East China University of Science and Technology (Natural Science Edition)*, 2018,44(6):882–886 (in Chinese with English abstract).
- [24] Liu JZ. Small sample bark image recognition method based on convolutional neural network. *Journal of Northwest Forestry University*, 2019,34(4):230–235 (in Chinese with English abstract).
- [25] Liu JM, Meng YL, Wan XY. Cross-task dialog system based on small sample machine learning. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, 2019,31(3):299–304 (in Chinese with English abstract).
- [26] Upadhyay S, Faruqui M, Tur G, *et al.* (Almost) Zero-shot cross-lingual spoken language understanding. In: *Proc. of the 2018 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2018)*. IEEE, 2018. 6034–6038.
- [27] Lampinen AK, McClelland JL. One-shot and few-shot learning of word embeddings. *arXiv preprint arXiv:1710.10280*, 2017.
- [28] Li CK, Fang J, Wu N, Song JY, Zhou Q, Zhou QL. A road extraction method for high resolution remote sensing image with limited samples. *Science of Surveying and Mapping*, 2020,45(4):81–88 (in Chinese with English abstract).
- [29] Sun CW, Wen C, Xie K, He JB. Voiceprint recognition method of small sample based on deep migration model. *Computer Engineering and Design*, 2018,39(12):3816–3822 (in Chinese with English abstract).
- [30] Zhao Y. Convolutional neural network based carotid plaque recognition over small sample size ultrasound images [MS. Thesis]. Wuhan: Huazhong University of Science and Technology, 2018 (in Chinese with English abstract).
- [31] Cheng L, Yuan Q, Wang Y, Feng W, Dai XT. A small sample exploratory study of autogenous bronchial basal cells for the treatment of chronic obstructive pulmonary disease. *Chongqing Medical*, 2019,48(23):4012–4016 (in Chinese with English abstract).
- [32] Zhu FN, Ma Y. Comparison of small sample of local complications between femoral artery sheath removal and vascular closure device. *Journal of Modern Integrated Chinese and Western Medicine*, 2010,19(14):1748–1820 (in Chinese with English abstract).
- [33] Liu YF, Zhou Y, Liu X, Dong F, Wang C, Wang ZH. Wasserstein GAN-based small-sample augmentation for new-generation artificial intelligence: A case study of cancer-staging data in biology. *Engineering*, 2019,5(1):338–354 (in Chinese with English abstract).
- [34] Jia LZ, Qin RR, Chi RX, Wang JH. Evaluation research of nurse's core competence based on a small sample. *Medical Higher Vocational Education and Modern Nursing*, 2018,1(6):340–342 (in Chinese with English abstract).
- [35] Yan B, Zhou P, Yan L. Disease identification of small sample crop based on transfer learning. *Modern Agricultural Sciences and Technology*, 2019,(6):87–89 (in Chinese with English abstract).
- [36] Sun YY, Jiang ZH, Dong W, Zhang LP, Rao Y. Image recognition of tea plant disease based on convolutional neural network and small samples. *Jiangsu Journal of Agricultural Sciences*, 2019,35(1):48–55 (in Chinese with English abstract).
- [37] Wang X, Ma TM, Yang T, Song P, Xie QJ, Chen ZG. Moisture quantitative analysis with small sample set of maize grain in filling stage based on near infrared spectroscopy. *Journal of Agricultural Engineering*, 2018,34(13):203–210 (in Chinese with English abstract).
- [38] He XJ, Ma S, Wu YY, Jiang GR. E-commerce product sales forecast with multi-dimensional index integration under small sample. *Computer Engineering and Applications*, 2019,55(15):177–184 (in Chinese with English abstract).
- [39] Liu XP, Guo B, Cui DJ, Wu ZY, Zhang LJ. Q-precentile life prediction based on bivariate wiener process for gear pump with small sample size. *China Mechanical Engineering*, 2020,31(11):1315–1322 (in Chinese with English abstract).
- [40] Chen L, Zhang F, Jiang S. Deep forest learning for military object recognition under small training set condition. *Journal of Chinese Academy of Electronics*, 2019,14(3):232–237 (in Chinese with English abstract).
- [41] Sun HW, Xie XF, Sun T, Zhang LJ. Threat assessment method of warships formation air defense based on DBN under the condition of small sample data missing. *Systems Engineering and Electronics*, 2019,41(6):1300–1308 (in Chinese with English abstract).
- [42] Nakamura A, Harada T. Revisiting fine-tuning for few-shot learning. *arXiv preprint arXiv:1910.00216*, 2019.
- [43] Royle JA, Dorazio RM, Link WA. Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, 2007,16(1):67–85.
- [44] Liu JW, Liu Y, Luo XL. Semi-supervised learning methods. *Chinese Journal of Computers*, 2015,38(8):1592–1617 (in Chinese with English abstract).

- [45] Chen WJ. Semi-supervised learning study summary. *Academic Exchange*, 2011,7(16):3887–3889 (in Chinese with English abstract).
- [46] Su FL, Xie QH, Huang QQ, Qiu JY, Yue ZJ. Semi-supervised method for attribute extraction based on transductive learning. *Journal of Shandong University (Science Edition)*, 2016,51(3):111–115 (in Chinese with English abstract).
- [47] Tu EM, Yang J. A review of semi-supervised learning theories and recent advances. *Journal of Shanghai Jiaotong University*, 2018,52(10):1280–1291 (in Chinese with English abstract).
- [48] Wang YX, Hebert M. Learning from small sample sets by combining unsupervised meta-training with CNNs. In: *Advances in Neural Information Processing Systems*. 2016. 244–252.
- [49] Boney R, Ilin A. Semi-supervised few-shot learning with MAML. In: *Proc. of the ICLR (Workshop)*. 2018.
- [50] Ren MY, Triantafillou E, Ravi S, *et al.* Meta-learning for semi-supervised few-shot classification. *arXiv preprint arXiv:1803.00676*, 2018.
- [51] Liu Y, Lee J, Park M, *et al.* Learning to propagate labels: Transductive propagation network for few-shot learning. *arXiv preprint arXiv:1805.10002*, 2018.
- [52] Hou RB, Chang H, Ma BP, *et al.* Cross attention network for few-shot classification. In: *Advances in Neural Information Processing Systems*. 2019. 4003–4014.
- [53] Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, 2014,27:2672–2680.
- [54] Mehrotra A, Dukkipati A. Generative adversarial residual pairwise networks for one shot learning. *arXiv preprint arXiv:1703.08033*, 2017.
- [55] Hariharan B, Girshick R. Low-shot visual recognition by shrinking and hallucinating features. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2017. 3018–3027.
- [56] Wang YX, Girshick R, Hebert M, *et al.* Low-shot learning from imaginary data. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 7278–7286.
- [57] Xian Y, Sharma S, Schiele B, *et al.* f-VAEGAN-D2: A feature generating framework for any-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2019. 10275–10284.
- [58] Chen Z, Fu Y, Kim YX, *et al.* Image deformation meta-networks for one-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2019. 8680–8689.
- [59] Dixit M, Kwitt R, Niethammer M, *et al.* AGA: Attribute guided augmentation. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 7455–7463.
- [60] Liu B, Wang X, Dixit M, *et al.* Feature space transfer for data augmentation. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 9090–9098.
- [61] Schwartz E, Karlinsky L, Shtok J, *et al.* Delta-encoder: An effective sample synthesis method for few-shot object recognition. In: *Advances in Neural Information Processing Systems*. 2018. 2845–2855.
- [62] Chen Z, Fu Y, Zhang Y, *et al.* Semantic feature augmentation in few-shot learning. *arXiv preprint arXiv:1804.05298*, 2018.
- [63] Shen W, Shi Z, Sun J. Learning from adversarial features for few-shot classification. *arXiv preprint arXiv:1903.10225*, 2019.
- [64] Liu XP, Luan XD, Xie YX, *et al.* Transfer learning research and algorithm review. *Journal of Changsha University*, 2018,32(5): 33–36,41 (in Chinese with English abstract).
- [65] Wang H. Research review on transfer learning. *Computer Knowledge and Technology*, 2017,13(32):203–205 (in Chinese with English abstract).
- [66] Liu J, Yuan Q, Wu G, Yu X. Review of convolutional neural networks. *Computer Era*, 2018,(11):19–23 (in Chinese with English abstract).
- [67] Yang L, Wu YQ, Wang JL, Liu YL. Research on recurrent neural network. *Computer Application*, 2018,38(S2):1–6,26 (in Chinese with English abstract).
- [68] Wang YX, Hebert M. Learning to learn: Model regression networks for easy small sample learning. In: *Proc. of the European Conf. on Computer Vision*. Springer Int'l Publishing, 2016.

- [69] Jang YH, Lee HK, Hwang SJ, *et al.* Learning what and where to transfer. In: Proc. of the Int'l Conf. on Machine Learning. 2019. 3030–3039.
- [70] Shen YY, Yan Y, Wang HZ. Recent advances on supervised distance metric learning algorithms. *Acta Automatica Sinica*, 2014, 40(12):2673–2686 (in Chinese with English abstract).
- [71] Aurélien B, Amaury H, Marc S. A survey on metric learning for feature vectors and structured data. *arXiv preprint arXiv:1306.6709*, 2013.
- [72] Kulis B. Metric learning: A survey. *Foundations and Trends® in Machine Learning*, 2013, 5(4):287–364.
- [73] Weinberger KQ. Distance metric learning for large margin nearest neighbor classification. In: *Advances in Neural Information Processing Systems*. 2006. 1473–1480.
- [74] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. In: *Proc. of the ICML Deep Learning Workshop*. 2015.
- [75] Jiang LB, Zhou XL, Jiang FW, Che L. One-shot learning based on improved matching network. *Systems Engineering and Electronics*, 2019, 41(6):1210–1217 (in Chinese with English abstract).
- [76] Wang P, Liu L, Shen C, *et al.* Multi-attention network for one shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 2721–2729.
- [77] Gao TY, Han X, Liu ZY, Sun MS. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In: *Proc. of the AAAI Conf. on Artificial Intelligence*. 2019. 6407–6414.
- [78] Sun SL, Sun QF, Zhou K, Lv TC. Hierarchical attention prototypical networks for few-shot text classification. In: *Proc. of the Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP)*. 2019. 476–485.
- [79] Sung F, Yang Y, Zhang L, *et al.* Learning to compare: Relation network for few-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 1199–1208.
- [80] Hilliard N, Phillips L, Howland S, *et al.* Few-shot learning with metric-agnostic conditional embeddings. *arXiv preprint arXiv:1802.04376*, 2018.
- [81] Li W, Xu J, Huo J, Wang L, Yang G, Luo J. Distribution consistency based covariance metric networks for few-shot learning. In: *Proc. of the AAAI Conf. on Artificial Intelligence*. 2019. 8642–8649.
- [82] Li W, Wang L, Xu J, Huo J, Gao Y, Luo J. Revisiting local descriptor based image-to-class measure for few-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2019. 7260–7268.
- [83] Li H, Eigen D, Dodge S, *et al.* Finding task-relevant features for few-shot learning by category traversal. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2019. 1–10.
- [84] Thrun S, Pratt L. Learning to learn: Introduction and overview. In: *Learning to Learn*. Boston: Springer-Verlag, 1998. 3–17.
- [85] Vilalta R, Drissi Y. A perspective view and survey of meta-learning. *Artificial Intelligence Review*, 2002, 18(2):77–95.
- [86] Ravi S, Larochelle H. Optimization as a model for few-shot learning. In: *Proc. of the ICLR*. 2016.
- [87] Hochreiter S, Younger AS, Conwell PR. Learning to learn using gradient descent. In: *Proc. of the Int'l Conf. on Artificial Neural Networks*. Berlin, Heidelberg: Springer-Verlag, 2001. 87–94.
- [88] Santoro A, Bartunov S, Botvinick M, *et al.* One-shot learning with memory-augmented neural networks. *arXiv preprint arXiv:1605.06065*, 2016.
- [89] Graves A, Wayne G, Danihelka I. Neural Turing machines. *arXiv preprint arXiv:1410.5401*, 2014.
- [90] Munkhdalai T, Yu H. Meta networks. *International Conference on Machine Learning*. In: *Proc. of the PMLR*. 2017. 2554–2563.
- [91] Jamal MA, Qi GJ. Task agnostic meta-learning for few-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2019. 11719–11727.
- [92] Sun Q, Liu Y, Chua TS, *et al.* Meta-transfer learning for few-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2019. 403–412.
- [93] Liu Y, Sun Q, Liu AA, *et al.* LCC: Learning to customize and combine neural networks for few-shot learning. *arXiv preprint arXiv:1904.08479*, 2019.

- [94] Wang X, Yu F, Wang R, *et al.* TAFE-Net: Task-aware feature embeddings for low shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2019. 1831–1840.
- [95] Gidaris S, Komodakis N. Dynamic few-shot visual learning without forgetting. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 4367–4375.
- [96] Zhou J, Cui G, Zhang Z, *et al.* Graph neural networks: A review of methods and applications. arXiv preprint arXiv:1812.08434, 2018.
- [97] Garcia V, Bruna J. Few-shot learning with graph neural networks. In: Proc. of the Int'l Conf. on Learning Representations. 2018.
- [98] Kim J, Kim T, Kim S, *et al.* Edge-labeling graph neural network for few-shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2019. 11–20.
- [99] Gidaris S, Komodakis N. Generating classification weights with GNN denoising autoencoders for few-shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2019. 21–30.
- [100] Fort S. Gaussian prototypical networks for few-shot learning on Omniglot. arXiv preprint arXiv:1708.02735, 2017.
- [101] Malalur P, Jaakkola T. Alignment based matching networks for one-shot classification and open-set recognition. arXiv preprint arXiv:1903.06538, 2019.
- [102] Yin C, Feng Z, Lin Y, *et al.* Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop. In: Proc. of the IEEE conference on computer vision and pattern recognition. 2016. 1153–1162.
- [103] Michael B, Tracy K, Samuel D. Gosling. Amazon's Mechanical Turk. Perspectives on Psychological Science, 2011,6(1):3–5.
- [104] Deng J, Dong W, Socher R, *et al.* ImageNet: A large-scale hierarchical image database. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2009. 248–255.
- [105] Wang K, Liu BS. Research review on text classification. Data Communication, 2019,(3):37–47 (in Chinese with English abstract).
- [106] Long M, Zhu H, Wang J, *et al.* Unsupervised domain adaptation with residual transfer networks. In: Proc. of the 30th Int'l Conf. on Neural Information Processing Systems. 2016. 136–144.
- [107] Koh PW, Liang P. Understanding black-box predictions via influence functions. In: Proc. of the Int'l Conf. on Machine Learning. 2017. 1885–1894.
- [108] Han X, Zhu H, Yu P, *et al.* FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. 2018. 4803–4809.
- [109] Cohn DA, Ghahramani Z, Jordan MI. Active learning with statistical models. Journal of Artificial Intelligence Research, 1996, 4(1):705–712.
- [110] Kaelbling LP, Littman ML, Moore AP. Reinforcement learning: A survey. Journal of Artificial Intelligence Research, 1996,4: 237–285.

附中文参考文献:

- [1] 李新叶,龙慎鹏,朱婧.基于深度神经网络的少样本学习综述.计算机应用研究,2020,37(8):2241–2247.
- [8] 杨军,刘妍丽.基于图像的单样本人脸识别研究进展.西华大学学报(自然科学版),2014,33(4):1–5,10.
- [18] 黄振文,谢凯,文畅,盛冠群,文方青.迁移学习模型下的小样本人脸识别算法.长江大学学报(自然科学版),2019,16(7):88–94.
- [19] 胡正平,何薇,王蒙,孙哲,任大伟.深度子空间联合稀疏表示单样本人脸识别算法.燕山大学学报,2018,42(5):409–415.
- [20] 吕永强,闵颀庆,段华,蒋树强.融合三元卷积神经网络与关系网络的小样本食品图像识别.计算机科学,2020,47(1):136–143.
- [21] 林克正,白婧轩,李昊天,李鹭.深度学习下融合不同模型的小样本表情识别.计算机科学与探索,2020,14(3):482–492.
- [22] 周添一,赵磊.基于神经网络的小样本手写汉字识别.山东理工大学学报(自然科学版),2019,33(3):69–74.
- [23] 全志楠,林家骏.文本无关的小样本手写汉字笔迹鉴别方法.华东理工大学学报(自然科学版),2018,44(6):882–886.
- [24] 刘嘉政.基于卷积神经网络的小样本树皮图像识别方法.西北林学院学报,2019,34(4):230–235.
- [25] 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.重庆邮电大学学报(自然科学版),2019,31(3):299–304.
- [28] 李朝奎,方军,吴媛,宋璟毓,周倩,周青蓝.一种针对小样本的高分辨率遥感影像道路提取方法.测绘科学,2020,45(4):81–88.
- [29] 孙存威,文畅,谢凯,贺建巍.深度迁移模型下的小样本声纹识别方法.计算机工程与设计,2018,39(12):3816–3822.
- [30] 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[硕士学位论文].武汉:华中科技大学,2018.

- [31] 程林,袁谦,王瑜,冯伟,孙凤军,戴晓天.自体支气管基底层细胞治疗慢性阻塞性肺疾病的小样本探索性研究.重庆医学,2019,48(23):4012-4016.
- [32] 朱方年,马野.拔除股动脉鞘管手法压迫与应用血管闭合装置的局部并发症小样本比较.现代中西医结合杂志,2010,19(14):1748-1820.
- [33] 刘宇飞,周源,刘欣,董放,王畅,王子鸿.基于 Wasserstein GAN 的新一代人工智能小样本数据增强方法——以生物领域癌症分期数据为例.工程,2019,5(1):338-354.
- [34] 贾玲芝,秦蓉蓉,迟荣香,王吉华.基于小样本的护士核心能力评估研究.医药高职教育与现代护理,2018,1(6):340-342.
- [35] 燕斌,周鹏,严利.基于迁移学习的小样本农作物病害识别.现代农业科技,2019,(6):87-89.
- [36] 孙云云,江朝晖,董伟,张立平,饶元,李绍稳.基于卷积神经网络和小样本的茶树病害图像识别.江苏农业学报,2019,35(1):48-55.
- [37] 王雪,马铁民,杨涛,宋平,谢秋菊,陈争光.基于近红外光谱的灌浆期玉米籽粒水分小样本定量分析.农业工程学报,2018,34(13):203-210.
- [38] 何喜军,马珊,武玉英,蒋国瑞.小样本下多维指标融合的电商产品销量预测.计算机工程与应用,2019,55(15):177-184.
- [39] 刘小平,郭斌,崔德军,吴振宇,张立杰.基于二元维纳过程的小样本齿轮泵可靠寿命预测.中国机械工程,2020,31(11):1315-1322.
- [40] 陈龙,张峰,蒋升.小样本条件下基于深度森林学习模型的典型军事目标识别方法.中国电子科学研究院学报,2019,14(3):232-237.
- [41] 孙海文,谢晓方,孙涛,张龙杰.小样本数据缺失状态下 DBN 舰艇编队防空目标威胁评估方法.系统工程与电子技术,2019,41(6):1300-1308.
- [44] 刘建伟,刘媛,罗雄麟.半监督学习方法.计算机学报,2015,38(8):1592-1617.
- [45] 陈武锦.半监督学习研究综述.电脑知识与技术,2011,7(16):3887-3889.
- [46] 苏丰龙,谢庆华,黄清泉,邱继远,岳振军.基于直推式学习的半监督属性抽取.山东大学学报(理学版),2016,51(3):111-115.
- [47] 屠恩美,杨杰.半监督学习理论及其研究进展概述.上海交通大学学报,2018,52(10):1280-1291.
- [64] 刘鑫鹏,荣悉道,谢毓湘,等.迁移学习研究和算法综述.长沙大学学报,2018,32(5):33-36,41.
- [65] 王惠.迁移学习研究综述.电脑知识与技术,2017,13(32):203-205.
- [66] 刘健,袁谦,吴广,喻晓.卷积神经网络综述.计算机时代,2018,(11):19-23.
- [67] 杨丽,吴雨茜,王俊丽,刘义理.循环神经网络研究综述.计算机应用,2018,38(S2):1-6,26.
- [70] 沈媛媛,严严,王菡子.有监督的距离度量学习算法研究进展.自动化学报,2014,40(12):2673-2686.
- [75] 蒋留兵,周小龙,姜风伟,车俐.基于改进匹配网络的单样本学习.系统工程与电子技术,2019,41(6):1210-1217.
- [105] 汪启,刘柏嵩.文本分类研究综述.数据通信,2019,(3):37-47.



赵凯琳(1995—),女,博士生,主要研究领域为面向知识图谱的小样本学习。



王元卓(1978—),男,博士,研究员,博士生导师,CCF 杰出会员,主要研究领域为大数据分析,开放知识网络,社交演化分析。



靳小龙(1976—),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为大数据知识工程,知识图谱。