



开放科学(OSID)

半监督深度学习图像分类方法研究综述

吕昊远⁺, 俞璐, 周星宇, 邓祥

陆军工程大学 通信工程学院, 南京 210007

+ 通信作者 E-mail: haoyuan0130@163.com

摘要:作为人工智能领域近十年来最受关注的技术之一,深度学习在诸多应用中取得了优异的效果,但目前的学习策略严重依赖大量的有标记数据。在许多实际问题中,获得众多有标记的训练数据并不可行,因此加大了模型的训练难度,但容易获得大量无标记的数据。半监督学习充分利用无标记数据,提供了在有限标记数据条件下提高模型性能的解决思路 and 有效方法,在图像分类任务中达到了很高的识别精准度。首先对于半监督学习进行概述,然后介绍了分类算法中常用的基本思想,重点对近年来基于半监督深度学习框架的图像分类方法,包括多视图训练、一致性正则、多样混合和半监督生成对抗网络进行全面的综述,总结多种方法共有的技术,分析比较不同方法的实验效果差异,最后思考当前存在的问题并展望未来可行的研究方向。

关键词:半监督深度学习;多视图训练;一致性正则;多样混合;半监督生成对抗网络

文献标志码:A **中图分类号:**TP391.4

Review of Semi-supervised Deep Learning Image Classification Methods

LYU Haoyuan⁺, YU Lu, ZHOU Xingyu, DENG Xiang

College of Communication Engineering, Army Engineering University of PLA, Nanjing 210007, China

Abstract: As one of the most concerned technologies in the field of artificial intelligence in recent ten years, deep learning has achieved excellent results in many applications, but the current learning strategies rely heavily on a large number of labeled data. In many practical problems, it is not feasible to obtain a large number of labeled training data, so it increases the training difficulty of the model. But it is easy to obtain a large number of unlabeled data. Semi-supervised learning makes full use of unlabeled data, provides solutions and effective methods to improve the performance of the model under the condition of limited labeled data, and achieves high recognition accuracy in the task of image classification. This paper first gives an overview of semi-supervised learning, and then introduces the basic ideas commonly used in classification algorithms. It focuses on the comprehensive review of image classification methods based on semi-supervised deep learning framework in recent years, including multi-view training, consistency regularization, diversity mixing and semi-supervised generative adversarial networks. It summarizes the common technologies of various methods, analyzes and compares the differences of experimental results of different methods. Finally, this paper thinks about the existing problems and looks forward to the feasible research direction in the future.

Key words: semi-supervised deep learning; multi-view training; consistency regularization; diversity mixing; semi-supervised generative adversarial networks

基金项目:国家自然科学基金(61702543)。

This work was supported by the National Natural Science Foundation of China (61702543).

收稿日期:2020-11-09 **修回日期:**2021-03-01

随着人们对于深度学习的研究热情日益高涨,各种深度学习框架和网络模型被相继提出^[1]。深度学习在很多领域都有着不错的表现,尤其是计算机视觉中取得突出的成果,在图像分类、目标检测、语义分割等各种任务中都达到了极佳性能^[2]。

然而,深度神经网络的训练质量受训练集中有标记样本数目的强烈影响。在实际的问题中,获取数以万计的有标记训练样本的代价是极其昂贵的,但对于无标记样本却很容易获取,人们就用多种方式研究无标记样本的使用。在此之中,基于少量有标记样本和大量无标记样本的半监督学习(semi-supervised learning)更适用于真实条件下的应用^[3],近来已成为深度学习领域热门的新方向。

半监督学习是一种在不需要大量标记的情况下对大量数据进行模型训练的强大方法。半监督学习通过提供一种利用无标记数据的方法来减少对标记数据的需求^[4]。由于无标记的数据通常可以用最少的人力获得,因此半监督学习带来的任何性能提升通常都是低成本的。人们对于半监督学习方法的研究热情也与日俱增,提出了大量基于深度学习框架的半监督学习方法。

在半监督学习中,有一个最基本的前提假设是,在训练过程中使用无标记的数据样本是有益的(参见图1的具体说明)。它可以使标记较少的训练更加健壮,甚至在某些情况下的学习效果超过有监督学习^[5]。

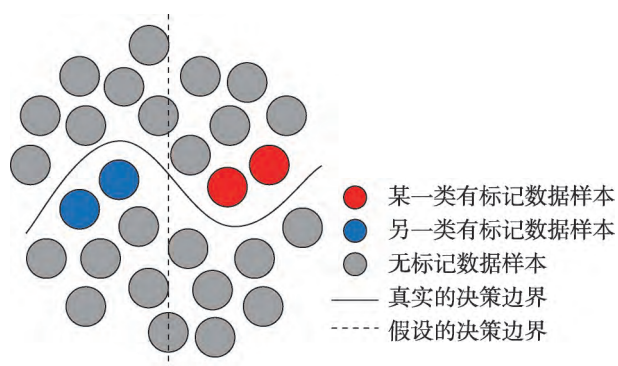


Fig.1 Schematic diagram of semi-supervision function

图1 半监督作用示意图

图1简单说明了半监督学习中使用未标记数据的好处。红色和蓝色圆圈代表不同类别的有标记数据点。灰色圆圈代表无标记的数据点。如果只有少量有标记样本可用,只能对潜在的真实分布(实线)

做出假设(虚线)。只有当考虑未标记的数据点并阐明决策边界时,才能确定这种真正的分布。

通常认为半监督学习的研究于1994年由 Shah-shahani 和 Landgrebe 提出^[6]。半监督学习在20世纪末,21世纪初随着实际应用中对于使用无标记数据的巨大需求的涌现而快速发展。并且在较短的时间里,半监督学习中基于分歧的方法、半监督支持向量机、图半监督学习的代表性工作就取得了巨大的研究成果。本文只关注于基于深度神经网络的半监督学习方法。

Hinton 等在2006年提出了“深度学习”这一概念^[7]。深度学习通过深度神经网络(deep neural networks, DNN)将底层的特征映射到高层,并且通过高层将特征抽象出来,从而发现数据的分布式特征表示。经过十多年的不断发展,深度学习在各个领域都有着重大突破。深度网络也有着越来越强的函数表达能力和特征提取能力。

基于深度学习框架的半监督学习是近几年来提出的研究新思路,其可以理解为,在少量有标记数据和大量无标记数据混合成的训练数据中使用的深度学习算法。通过几年来的不断研究,基于深度学习框架的半监督学习算法已发展为三类。包括无标记数据预训练网络后用有标记数据对网络进行微调;有标记数据训练网络,利用从网络中得到的深度特征来做半监督算法;端到端的半监督深度模型^[8]。

前两类方法虽然都用到了有标记样本和无标记样本,但对于深度神经网络本身而言,它的训练还是在有监督学习的方式上。只有第三种方法,才是网络本身的半监督学习。本文也将围绕这个主题阐述。

1 基本思想

虽然半监督学习方法不尽相同,但都有用到了一些最基本的共有思想,也可以理解为算法训练中的一些技巧手段,包括损失函数类型和样本集增强方式等。在本章中,将列举这些常用的基本想法,并作简要介绍。

1.1 常用的损失函数

1.1.1 KL 散度

Kullback-Leibler 散度(KL)是用来衡量同一个随机变量的两个不同分布之间的距离^[9]。设 $p(x)$ 、 $q(x)$ 分别是离散随机变量 X 的两个概率分布,则 p 对 q

的相对熵是:

$$D_{KL}(p||q) = \sum_i p(x_i) \lg \frac{p(x_i)}{q(x_i)} \quad (1)$$

在实际的图像分类问题中,假如 $p(x)$ 是目标真实的分布,而 $q(x)$ 是预测得来的分布,为了让这两个分布尽可能相同,就需要最小化KL散度。

1.1.2 交叉熵

交叉熵(cross entropy, CE)是一种常见的图像分类损失函数。被广泛用于有标记数据的有监督学习。其用来测定给定的数据集中,有标记样本的预测标记值和其真实标记值之间的误差^[10]。

$$CE(f(x), z) = \sum_{c=1}^C P_{f(x)}(c) \lg P_z(c) = H(P_z) + KL(P_z | P_{f(x)}) \quad (2)$$

式中, P 是所有类的概率分布; H 为概率分布的熵; KL 为 Kullback-Leibler 散度。分布 P 可以近似于神经网络的输出 $f(x)$ 或给定的标签 z 。交叉熵是 $f(x)$ 和 z 之间的熵和KL散度的总和。一般来说,对于独热编码(one-hot)标记 z ,熵 $H(P_z)$ 为0。

1.1.3 均方误差

对于数据集中的两样本,均方误差(mean-square error, MSE)是其经过神经网络的标记预测值之间的距离度量。与之前的误差函数相比,MSE不是基于概率论测量差异,而是使用输出向量的欧式距离^[11]。

$$MSE(f(x), f(y)) = \|f(x) - f(y)\|_2^2 \quad (3)$$

这个度量的最小化可以使两个输出相互收缩。这个距离度量也可以用于 $f(x)$ 和 $f(y)$ 的任何中间输出(特征空间)^[12]。由此可见,使用MSE不需要用到样本标记,因此MSE方法可以广泛地使用在无标记样本数据的一致性正则中。

1.1.4 熵最小化

通过最小化熵(EntMin)来锐化半监督学习中的输出预测。对于图像 x ,它们基于特定的神经输出 $f(x)$ 对所有概率分布的 $P_{f(x)}$ 最小化熵为 $H(P_{f(x)})$,实际表现为在图像分类问题中,基于熵最小化的思想,通常在给定平均预测量的基础上,应用锐化函数来减小标记分布的熵^[13],如图2所示。不过这种熵最小化只会增强神经网络的预测,不能单独使用。

1.2 样本集增强方式

1.2.1 伪标记

估计未知数据标记的一种简单方法是使用伪标记,可以用深度网络对无标记数据进行分类识别,并

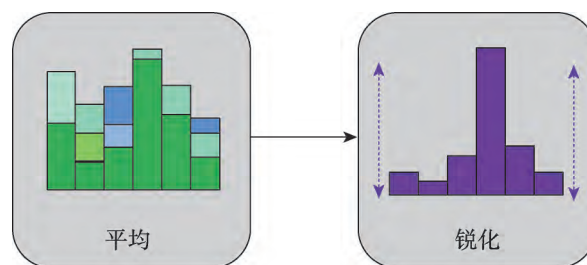


Fig.2 Sharpening prediction label value

图2 预测标签值的锐化

将预测值作为其数据标记^[14]。这种自我实现的假设在实际的图像分类任务中取得了非常好的效果。最新提出的方法都是基于这样的核心思想,即通过自己网络的预测来创建标记,然后作为有标记数据加入到训练集中。

1.2.2 数据增强

深度学习任务中,训练数据一定程度上决定了模型的好坏。在多数情况下,经常面对着数据不够多,质量不够好的问题,而数据增强就是解决这一问题的有效方法^[15]。本文将分为传统增强和强增强两种方式。

传统的数据增强方式又分为两类,其中几何变换选择了图像的一部分或者对像素进行了重分布,包括图像的平移、翻转、旋转、裁剪、比例缩放等。颜色变换则改变了图像本身的内容,常见的包括噪声、模糊、颜色变换、擦除、填充等^[16]。

强增强方式包括随机裁剪(CutOut),这种增强会随机选择图像的某个区域并掩盖。自动增强(Auto-Augment)使用强化学习来自动创建有用的增强^[17],其还有两种变体方法,RandAugment和CTAugment,非常适合缺少标记的半监督环境^[18]。除此之外,还有通过模型学习数据的分布,随机生成与训练数据集分布一致的图片,代表方法为GAN网络。

mixup是一种运用在计算机视觉中的对图像进行混类增强的算法^[19],它可以将不同类之间的图像进行混合,从而扩充数据集。mixup通过将图像相互混合来创建新的凸形组合。再对相应标记的凸组合进行预测。研究表明,相应标记凸组合的预测对一般的监督学习是有益的。

如图3所示,猫和狗的图像被参数化混合在一起,标记也相应地参数化结合在一起。所示图像取自STL-10数据集。

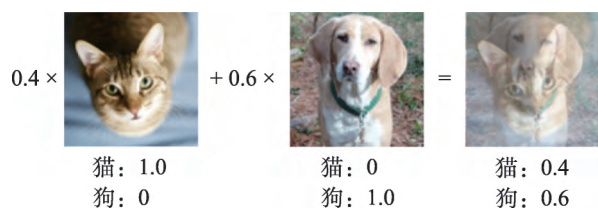


Fig.3 Diagram of mixup

图3 mixup的图解

2 半监督学习方法

本文将基于半监督深度学习的图像分类方法分为四个类别,分别为多视图训练、一致性正则、多样混合训练以及与GAN网络相结合的半监督学习。接下来将分别介绍各个类别,并举出典型网络模型进行分析比较。

2.1 多视图训练

多视图训练假设每个数据可以从不同的角度(view)进行分类,然后用这些从不同角度训练出来的分类器再对无标记样本进行分类,再选出认为可信的无标记样本以及其伪标记加入训练集中^[6]。

在这种情况下,多视图训练的目标是学习独特的预测函数为各视图下的数据建模,并共同优化所有用于提高泛化性能的功能。理想情况下,不同视图的结果相互补充,不同的模型可以相互协作以提高彼此的性能。

2.1.1 Co-training

Co-training 有两个模型,它们分别在数据集上训练。每轮迭代中,如果两个模型里的一个模型认为自己对样本分类的置信度高,分类概率大于某个事先设好的阈值,这个模型就会为它生成伪标记,然后把它放入另一个模型的训练集。简而言之,一个模型会为另一个模型的输入样本提供标记^[20]。

Qiao 等^[21]提出了深度 Co-training 结合半监督学习的神经网络模型。不过其想法略有不同,没有对同一个数据集训练两个模型,而是考虑到视图差异约束。具体做法如图4所示,对于数据集,通过计算其对抗样本从而派生出新的数据集。然后使用对方的对抗样本训练自己,从而很好地解决模型崩溃问题。

2.1.2 Tri-training

Chen 等^[22]提出的 Tri-training 网络模型,如图5所示,首先对有标记示例集进行可重复取样以获得三个有标记训练集,然后从每个训练集产生一个分类

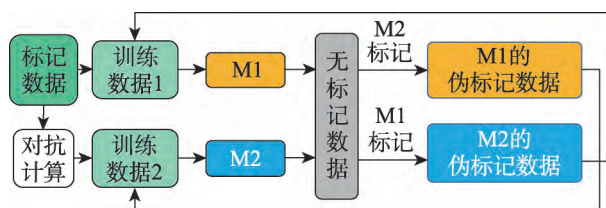


Fig.4 Co-training algorithm flow

图4 Co-training 算法流程

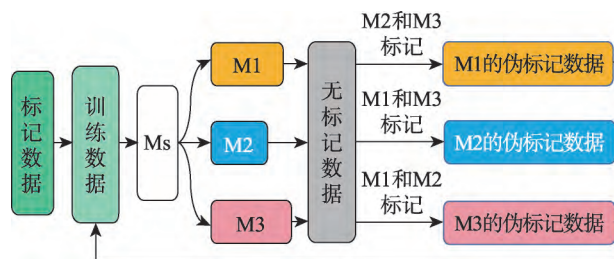


Fig.5 Tri-training algorithm flow

图5 Tri-training 算法流程

器。在协同训练过程中,各分类器所获得的新标记示例都由其余两个分类器协作提供。

具体来说,如果两个分类器对同一个未标记示例的预测相同,则该示例就被认为具有较高的标记置信度,并在标记后被加入第三个分类器的有标记训练集。

多视图训练的缺点也比较明显,算法中需要多个深度模型,训练过程相对复杂,并且靠经验设计。其中 Tri-training 算法中的伪标签需要进行多次预测,也会加重计算负担。

2.2 一致性正则

因为在半监督学习中,无标记数据没有自己的类别标记,所以就研究出了无需标记信息的一致性正则方法。一致性正则的主要思想就是,对于一个输入,即使受到了细微的干扰,其预测都应该是一致的^[23]。

应用在具体的半监督算法中,最基本的思路是,对于有标记样本,使用标准的交叉熵损失函数计算误差,而对于无标记样本,就需要使用一致性正则,比如最小均方误差或者KL散度等,具体到每一种算法的核心思想是没有变化的,即最小化未标记数据与其扰动输出两者之间的距离,但计算输出的形式上有很多变化。最后还需加入权重系数,结合有监督与无监督两部分损失,从而构建出总体的半监督损失函数。

下面就阐述典型的基于一致性正则的半监督学习方法。

2.2.1 π 模型和时序组合模型

Laine 和 Aila^[24]提出了两种类似的半监督学习方法模型,即图6展示的 π 模型和时序组合模型^[25]。这两种方法都结合使用了有监督的交叉熵损失和无监督的最小均方损失。

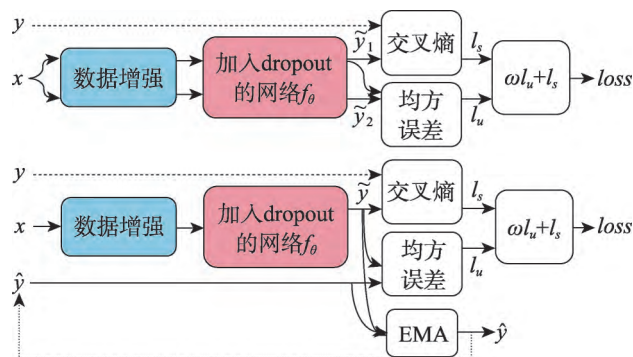


Fig.6 Algorithm flow of π and temporal ensembling models

图6 π 模型与时序组合模型的算法流程

在这两种方法中,无监督一致性损失的第一个输入都是来自随机增强后的图像的网络输出。而第二个输入是不同的,在 π 模型中是相同图像的第二次增强,在时序组合模型中,是对先前迭代周期预测结果的指数移动平均(exponential moving average, EMA)。

π 模型的训练过程中,每轮次的迭代需要网络预测两次,相比之下,时序组合模型的速度快了两倍,而且更加稳定。

2.2.2 平均教师模型

时序组合模型中,指数移动平均线每个时期只更新一次,这会使得网络训练效率较低。Tarvainen 和 Valpola^[26]提出了一种半监督学习的学生教师方法,继承并进一步发展了 π 模型和时序组合模型,也使用均方误差作为两个预测之间的一致性损失。

但创建这些预测不同,如图7所示^[25],他们提议在每个更新步骤中使用基于学生模型平均权重的教师模型,有效解决了时序组合模型中将新信息融入预测的速度太慢的问题。并且经过多次实验表明,均方误差是比KL散度更好的一致性损失。

2.2.3 无监督数据增强

前文中提到的几个基于一致性正则的半监督学习方法中,对于图像增强部分主要采用简单的随机

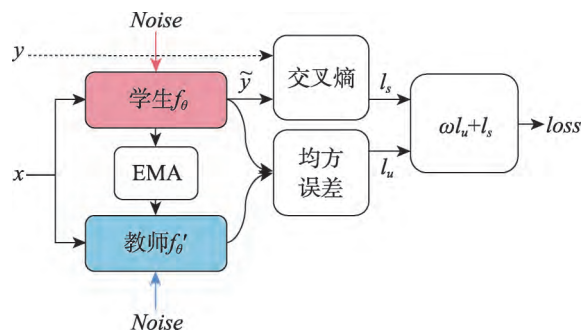


Fig.7 Algorithm flow of mean teacher model

图7 平均教师模型的算法流程

噪声,Xie等^[27]提出了一种半监督学习方法,图8展示的无监督数据增强(unsupervised data augment, UDA)集中使用了最先进的数据增强。也结合有监督损失和无监督损失,有监督损失为交叉熵,无监督损失为输出预测之间的KL散度。这些输出预测是基于原图像和该图像的增强版本。对于图像分类,他们提出使用自动增强结合裁剪的增强方案。

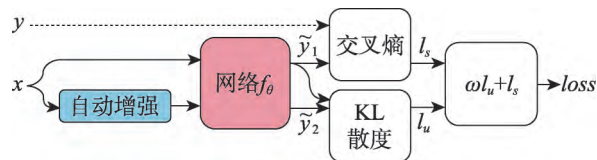


Fig.8 UDA model algorithm flow

图8 UDA 模型算法流程

研究表明,这种组合增强方法比以往单独的裁剪或翻转等方法获得的性能更高。除了不同的增强方法外,他们还提出使用多种其他正则化方法以及训练信号退火算法,可以在训练过程中限制标记样本的影响,有效防止过拟合。

UDA 还结合了熵最小化和伪标记的思想,但使用伪标记是过滤无监督数据的异常值。

从局限性来讲,UDA方法中的强增强方式只运用在了图像基准问题上。其他领域的研究还需要探索新的数据增强方式。

2.3 多样混合方法

多样混合方法试图在一个框架中整合前文提到的半监督学习的多种思想,从而集思广益,在识别任务中获得更好的性能。

2.3.1 MixMatch

MixMatch 同样使用监督损失和非监督损失的组合。Berthelot等^[28]使用交叉熵作为监督损失,使用预

测和生成伪标记之间的最小均方误差作为无监督损失,他们提出了一种新的多次预测再锐化方法来提高伪标记的质量,这些伪标记是根据先前对增强图像的多次预测而创建的(如图9所示)。并且创新之处在于,通过合并生成的标记,将mixup算法扩展到半监督学习。

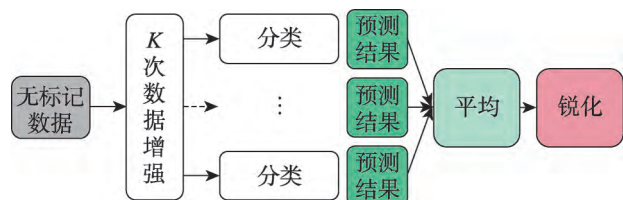


Fig.9 Multi-prediction sharpening

图9 多预测锐化

如图10所示, x' 和 u' 是经过 mixup 后得到的混合样本, 分别是标记数据和无标记数据。 p 和 q 为 mixup 后分别对应的标记。有标记数据使用交叉熵, 预测标记则使用均方误差, 最后将两部分损失加权组合。

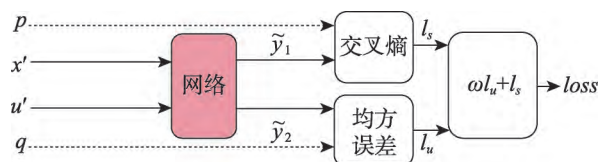


Fig.10 MixMatch model algorithm flow

图10 MixMatch模型算法流程

2.3.2 FixMatch

FixMatch 是 Google Brain 提出的一种多混合的半监督学习方法^[29](如图11所示), 与 MixMatch 方法不同的是, FixMatch 使用交叉熵将弱增强(标准的翻转和平移)和强增强(先使用 RandAugment 或 CT-Augment, 再使用 CutOut 增强)的无标记数据进行比较计算损失。其巧妙之处在于一致性正则化使用的是交叉熵损失函数, 并且取得了不错的效果。

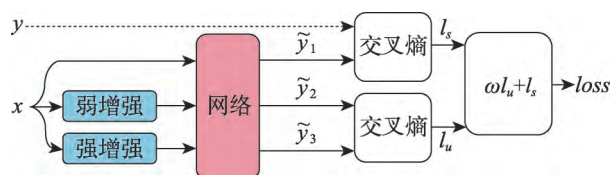


Fig.11 FixMatch model algorithm flow

图11 FixMatch模型算法流程

FixMatch 是在弱增强图像与强增强图像之间进行一致性计算, 但其没有使用两种图像的概率分布一致, 而是将弱增强数据预测结果看作伪标签 (\tilde{y}_2), 再结合强增强数据的预测结果 (\tilde{y}_3), 这样就需要使用交叉熵进行一致性的计算。此外, FixMatch 仅使用具有高置信度的未标记数据参与训练。

多样混合方法中, 多种技术的组合实现了很好的图像分类效果, 但与此同时, 模型结构、算法过程也更为复杂。

2.4 半监督生成对抗网络

生成对抗网络 (generative adversarial networks, GAN) 是 Goodfellow 等^[30]在2014年提出的一种深度生成模型, 可以广泛地用于图像视觉领域, 在图像领域有着突出的效果, 生成高分辨率逼真的图像, 进行图像修复、风格迁移等。在有标记样本数目较少的情况下, 就可以把半监督思想与 GAN 相结合, 这种半监督 GAN 在图像分类中取得很好的实验表现。在本节中, 先来概括 GAN 的结构。

GAN 通过框架中两个模块生成器 (generative model) 和判别器 (discriminative model) 的互相博弈学习, 从而最终产生相当好的输出^[31]。生成器和判别器可以用神经网络或者函数表示, 如图12所示, 用 G 表示生成器, D 表示判别器。

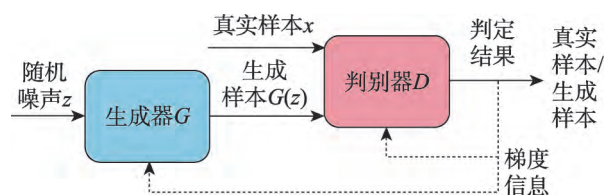


Fig.12 GAN model structure

图12 GAN模型结构

生成器 G 的目的是生成基于数据分布的 p_g , G 的输入为随机噪声 z (例如高斯噪声), z 取源于先验分布 $p_z(z)$ 。判别器对输入的数据进行判别, 如果是真实数据则输出1, 否则的话, 输出为0。

2.4.1 半监督学习 GAN

标准的分类是有监督的学习, 模型通过最小化交叉熵损失, 获得最优的网络参数。Salimans^[32]利用 GAN 完成半监督学习, 面对 K 分类问题时, 把 GAN 的判别器改为 $K+1$ 类的分类器, 其中第 $K+1$ 类是异常类 (生成器生成的假图像)。

如图 13 所示,分类器接收到的图像数据为三部分,有标记样本 (x, y) 、无标记样本 x 和 G 生成的假样本 $G(z)$ 。对于分类器而言,如果是有标记样本,就进行标准的有监督学习,学习任务是 K 分类问题;而对于无标记样本和生成器生成的假样本,分类器就需要将认为是假样本的图像分为第 $K+1$ 类,而将认为是真实样本的分为前 K 类中的某一类,但不需要判定预测正确与否。综上所述,分类器的损失函数由以上三部分构成。而生成器需要骗过分类器,即尽最大努力生成以假乱真的样本,并且希望生成的假样本被分类器判别为真(分为前 K 类中的某一类)。

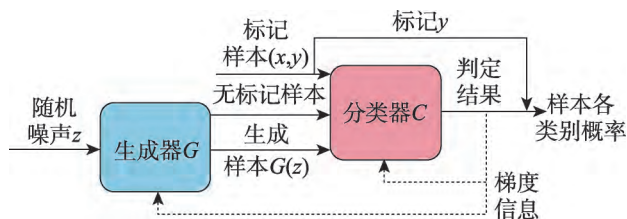


Fig.13 SSL-GAN model structure

图 13 SSL-GAN 模型结构

2.4.2 半监督 GAN

Odena^[3]提出的半监督 GAN(semi-supervised GAN, SGAN)的网络结构与上文的半监督学习 GAN 基本相似,所不同的是,SGAN 的想法更简单明了,如图 14 所示,分类器接受的图像只包括有标记样本 (x, y) 和生成的假样本 $G(z)$ 两部分,并且在具体的实验中,两部分的数目比例相同。

从缺点的角度看,半监督 GAN 网络中,生成的是

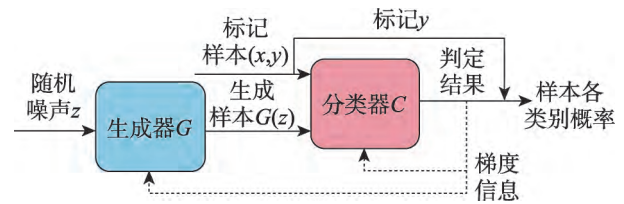


Fig.14 SGAN model structure

图 14 SGAN 模型结构

未带有标记信息的随机样本,小样本条件下会造成类别不均衡的问题,而且没有用到前文中提到的一些基本技术,这也限制了此类方法的性能提升。

3 实验比较

本章将分析前文提到的四类方法之间有哪些异同的思想,并且在几个通用的图像数据集上观察这些方法的实验结果表现。

3.1 方法比较

表 1 展示了不同的半监督方法中包含着的不同思想,所有的 \checkmark 表示此方法中包含有这个思想,最后一列统计了此种方法包含的思想数目,最后一行统计了包含此种思想的方法数目,表中空白位表示方法中未包含这种思想。现在对于一些角标作出解释: \checkmark^s 表示在所有的半监督方法中,有监督学习部分使用交叉熵损失; $\checkmark^{s,u}$ 表示 FixMatch 中有监督与无监督部分都使用交叉熵; $\checkmark^{\text{sharp}}$ 表示 MixMatch 通过对预测标签进行锐化而最小化熵; $\checkmark^{\text{filter}}$ 表示 UDA 预测用于过滤无监督数据的伪标记; \checkmark^{weak} 表示用到传统的数据增强方式; \checkmark^{auto} 表示用到 AutoAugment 强增强;

Table 1 Basic thoughts used in different methods

表 1 不同方法采用的基本思想

方法	KL 散度	均方误差	交叉熵	熵最小化	伪标记	Mixup	数据增强	总数
Co-Training	\checkmark		\checkmark^s		\checkmark		\checkmark^{weak}	4
Tri-Training	\checkmark		\checkmark^s		\checkmark		\checkmark^{weak}	4
π 模型		\checkmark	\checkmark^s				\checkmark^{weak}	3
时序组合模型		\checkmark	\checkmark^s				\checkmark^{weak}	3
平均教师模型		\checkmark	\checkmark^s				\checkmark^{weak}	3
UDA	\checkmark		\checkmark^s	\checkmark	$\checkmark^{\text{filter}}$		\checkmark^{auto}	5
MixMatch		\checkmark	\checkmark^s	$\checkmark^{\text{sharp}}$	\checkmark	\checkmark	\checkmark^{weak}	6
FixMatch			$\checkmark^{s,u}$	\checkmark	\checkmark		\checkmark^{rand}	4
SSL-GAN			\checkmark^s				\checkmark^{gan}	2
SGAN			\checkmark^s				\checkmark^{gan}	2
总数	3	4	10	3	5	1	10	36

$\sqrt{\text{rand}}$ 表示用到 RandAugment 和 CTAugment 强增强;
 $\sqrt{\text{gan}}$ 表示用到 GAN 网络生成分布一致的图像。

3.2 结果比较

本节中,在多种数据集上比较了所提到的半监督方法。根据之前论文公开的结果,选择了在多篇论文中使用的五个数据集进行比较。

MNIST 是手写阿拉伯数字的标准数据集^[34],一共包含 0 到 9 的 10 类数字。数据集中有 60 000 个训练样本和 10 000 个测试样本。这些数字已经过尺寸标准化并位于图像中心,图像是固定大小 28×28 像素,其值为 0 到 1。

CIFAR-10 和 CIFAR-100 是众多小型彩色图像构成的大规模数据集^[35],大小为 32×32。这两个数据集分别包含了属于 10 个或 100 个类别的 60 000 幅图像。两个数据集都提供了 50 000 个训练样本和 10 000 个测试样本,且都为有标记数据。但本文的半监督方法所呈现的结果是仅用训练集中少量有标记样本和一部分抹去标记的样本进行训练。

STL-10 是为无监督和半监督学习设计的数据集。数据集的创建受 CIFAR-10 的启发,但提供的标签较少。它提供了 5 000 个有标记训练样本和 8 000 个有标记测试样本^[36]。然而,还提供了 100 000 个无标记的示例图像。这些无标记的样本属于训练集和一些不同的类别。这些图像是 96×96 彩色图像。

ILSVRC-2012 是 ImageNet 的一个子集^[37],训练集由 120 万幅图像组成,而验证和测试集包括 150 000 幅图像。这些图像属于 1 000 个类别。所呈现的结果是仅用 10% 的标记来训练,以表示半监督情况。

表 2 和表 3 分别展示了当有标记样本数为 1 000 和 2 000 时,不同半监督学习方法以及全监督方法在 MNIST、CIFAR-10 和 CIFAR-100 数据集上的表现,表格最左侧一栏是不同的方法,右侧三栏是前文中提到的三个数据集。表格内容展示的是识别准确率,表中空白位表示方法未在此数据集上进行实验验证。其中无标记样本量是有标记样本的三倍,每个方法的最终识别准确率都选取了最好的实验结果。可以看到,半监督学习方法可以很好地利用无标记样本的信息,在同等有标记样本数据的情况下,识别准确率大幅度提升。

表 4 展示了当有标记样本数目为 4 000 的识别结果。表中空白位表示方法未在此数据集上进行实验

Table 2 Recognition accuracy of 1000 labeled samples

表 2 1 000 个有标记样本的识别准确率 %

方法	MNIST	CIFAR-10	CIFAR-100
全监督	84.12	53.67	39.49
Co-Training		73.10 ^[21]	58.12 ^[21]
Tri-Training	87.23 ^[22]	71.12 ^[22]	
π 模型	89.66	67.23	51.13 ^[24]
时序组合模型	91.25	71.81	
平均教师模型	92.49	76.12	
UDA		81.29	
MixMatch		79.12	58.58 ^[28]
FixMatch		75.71	62.31 ^[29]
SSL-GAN	90.38		
SGAN	88.29		

Table 3 Recognition accuracy of 2000 labeled samples

表 3 2 000 个有标记样本的识别准确率 %

方法	MNIST	CIFAR-10	CIFAR-100
全监督	86.42	55.14	42.98
Co-Training		84.52 ^[21]	59.93 ^[21]
Tri-Training	88.19 ^[22]	86.39 ^[22]	
π 模型	90.27	79.41	53.89 ^[24]
时序组合模型	91.98	81.05	
平均教师模型	93.35	84.12	
UDA		91.01	
MixMatch		91.97	62.29 ^[28]
FixMatch		93.26	68.71 ^[29]
SSL-GAN	91.75		
SGAN	90.19		

Table 4 Recognition accuracy of 4000 labeled samples

表 4 4 000 个有标记样本的识别准确率 %

方法	MNIST	CIFAR-10	CIFAR-100	STL-10	ILSVRC-2012
全监督	91.29	59.68 ^[26]	51.72 ^[38]	48.49 ^[39]	59.65 ^[40]
Co-Training		91.65 ^[21]	61.23 ^[21]		
Tri-Training	92.46 ^[22]	91.75 ^[22]			
π 模型	92.29 ^[24]	87.64 ^[24]	56.37 ^[24]		
时序组合模型	93.85	87.84 ^[24]			
平均教师模型	94.36	89.69 ^[26]			
UDA		94.70 ^[27]			88.52 ^[27]
MixMatch		95.05 ^[28]	74.12 ^[28]	94.41 ^[28]	
FixMatch		95.74 ^[29]	81.71 ^[29]		
SSL-GAN	93.65 ^[32]				
SGAN	93.41 ^[33]				

验证。根据表4,横向比较可以发现在五个数据集中,CIFAR-10的实验效果最好,这是因为十分类问题中,类别数目较少且各类之中样本数目均衡,并且图像的分辨率较低,使得训练也更加容易。纵向比较不同方法的识别结果,发现UDA、MixMatch和FixMatch取得了好的实验结果,结合表1,这是因为三个方法中都结合了多种不同的思想,从而使得这些方法产生了很好的联合反应。两个GAN方法在MNIST上表现很好,这是因为数字的二值图片特征比较好提取,生成器更容易学习到数据分布,从而推动分类器的性能提升。

4 存在的问题及发展趋势

本章将讨论三个主要趋势,这些趋势带来了未来可能的研究机会。

4.1 更少的标记数据

随着层出不穷的新半监督学习方法的提出,半监督学习和有监督学习的差距越来越小,表现在前文提到的数据集中,半监督方法可以利用不到有监督方法1/10的标记数据量,而达到较高的识别准确率。

因此可以预计,未来的研究将集中在使用所有标记中只有1%或更少的标记数据而获得更好的结果。

4.2 更多的想法结合

在很多半监督学习的论文中,由于作者的目的是用尽可能少的标记获得最好的准确率,因此没有做到集思广益,一些基本的想法之间只有很少的重叠。不过在UDA、MixMatch、FixMatch方法中,已经将伪标签、熵最小化和一致性正则等想法结合在一起,取得了很好的实验结果。由此可见,想法结合是有益的,但仍然有许多合理的结合尚未探索,多种想法的结合是一个未来很有前途的研究方向。

4.3 更大的领域扩展

根据图像识别准确度可以看出,许多方法可以在CIFAR-10(简单的低分辨率数据集)上有着90%以上的准确率,然而很少有方法能够在ILSVRC-2012(高分辨率数据集)上达到较高的准确率。因此得出结论,大多数方法都不能扩展到真实世界的图像分类问题上。

取得较高识别准确率的数据集中,不同类别的图像数目都相同,且不同类别之间差异明显。并且

标记数据和未标记数据来自同一分布。这种假设在许多实际应用中很难成立,其中一个常见的情况是,无标记数据包含有标记数据中未知的类别^[41]。因此未来的研究需要解决这些实际问题,做到处理类别不平衡以及含有噪声的有标记数据,以及面对不同分布的数据集时,有着具有更强鲁棒性的半监督学习方法。

本文所提到的半监督方法还是更多地与卷积神经网络以及生成对抗网络相结合。不过在最新的研究中,已经将半监督思想结合在其他深度网络模型方法,例如图神经网络、自编码器以及元学习、对比学习等。如何在更多的深度学习框架中结合半监督学习也是亟待探索的重点。

半监督深度学习方法应用于图像分类问题中已经取得了显著的成绩,而在其他领域,例如文本分类问题中也有着一定的成果^[42],那在自然语言处理、视频识别等问题中,是否也能有着优良的性能表现仍然需要进一步研究。

5 结束语

有监督的深度学习在很多应用领域已经有着很好的性能表现,但在实际问题中面临着海量标记数据难以获取的现状,而有标记数据较少时模型训练很容易出现过拟合现象。基于深度网络的半监督学习方法可以很好地使用少量的有标记数据和大量的无标记数据,在图像分类任务中取得非常好的实验效果。本文提供了一个基于半监督深度学习的图像分类方法概述。总结了多种半监督学习方法的差异、相似性以及组合效果,比较了不同方法在图像分类中的实验结果。并且分析讨论了半监督学习发展趋势和可能的研究领域。

如何利用更少的标记数据而提升识别准确率,未来的研究可能会进一步减少所需标记的数量;而对于类别不平衡的问题,半监督是否仍可以有着较高的识别准确率还有待实验证实;实验中已经发现不同技术的结合有利于整体性能的提升,如何做到想法之间的重叠,也是未来可能的研究机会;另外,还需要研究与注意力机制^[43]等更为复杂的卷积神经网络以及其他深度模型相结合^[44]的半监督方法。而且大多数半监督学习算法的工作是在图像基准上评估,未来的研究还需要探讨在其他领域的有效性。

参考文献:

- [1] ZHANG S, GONG Y H, WANG J J. The development of deep convolution neural network and its applications on computer vision[J]. Journal of Computers, 2019, 42(3): 453-482.
张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- [2] ZHOU F Y, JIN L P, DONG J. Review of convolutional neural network[J]. Journal of Computers, 2017, 40(6): 1229-1251.
周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [3] CHAPPELLE O, SCHOLKOPF B, ZIEN A. Semi-supervised learning[M]. Cambridge: MIT Press, 2006: 34-36.
- [4] MEY A, LOOG M. Improvability through semi-supervised learning: a survey of theoretical results[J]. arXiv:1908.09574, 2019.
- [5] XU J, HENRIQUES J F, VEDALDI A. Invariant information clustering for unsupervised image classification and segmentation[J]. arXiv:1807.06653, 2018.
- [6] ZHOU Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2018: 311-312.
周志华. 机器学习[M]. 北京: 清华大学出版社, 2018: 311-312.
- [7] LIU G H, ZHANG X B. A method for personal identification of communication radiation source based on deep belief network[J]. Chinese Journal of Radio Science, 2020, 35(3): 395-403.
刘高辉, 张晓博. 一种基于深度置信网络的通信辐射源个体识别方法[J]. 电波科学学报, 2020, 35(3): 395-403.
- [8] HAN S, HAN Q H. Review of semi-supervised learning research[J]. Computer Engineering and Applications, 2020, 56(6): 19-27.
韩嵩, 韩秋弘. 半监督学习研究的述评[J]. 计算机工程与应用, 2020, 56(6): 19-27.
- [9] LI H. Statistical learning methods[M]. Beijing: Tsinghua University Press, 2019: 84-88.
李航. 统计学习方法[M]. 北京: 清华大学出版社, 2019: 84-88.
- [10] QIU X P. Neural network and deep learning[M]. Beijing: China Machine Press, 2020: 124-129.
邱锡鹏. 神经网络与深度学习[M]. 北京: 机械工业出版社, 2020: 124-129.
- [11] CASCANTE-BONILLA P, TAN F W, QI Y J, et al. Curriculum labeling: self-paced pseudo-labeling for semi-supervised learning[J]. arXiv:2001.06001, 2020.
- [12] ZHOU Z H, WANG W, GAO W, et al. Introduction to machine learning theory[M]. Beijing: China Machine Press, 2020.
周志华, 王魏, 高尉, 等. 机器学习理论导引[M]. 北京: 机械工业出版社, 2020.
- [13] GRANDVALET Y, BENGIO Y. Semi-supervised learning by entropy minimization[C]//Proceedings of the Advances in Neural Information Processing Systems, Vancouver, Dec 13-18, 2004. Red Hook: Curran Associates, 2004: 529-536.
- [14] LEE D. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks[C]//Proceedings of the Workshop on Challenges in Representation Learning, Jun 21, 2013: 8-17.
- [15] CAI L M, WANG L J. Digital image processing[M]. Beijing: Tsinghua University Press, 2019.
蔡利梅, 王利娟. 数字图像处理[M]. 北京: 清华大学出版社, 2019.
- [16] ZHANG X F, WU G. Data augmentation method based on generative adversarial network[J]. Computer Systems & Applications, 2019, 28(10): 201-206.
张晓峰, 吴刚. 基于生成对抗网络的数据增强方法[J]. 计算机系统应用, 2019, 28(10): 201-206.
- [17] CUBUK E D, ZOPH B, MANE D, et al. AutoAugment: learning augmentation strategies from data[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16- 20, 2019. Piscataway: IEEE, 2019: 113-123.
- [18] CUBUK E D, ZOPH B, SHLENS J, et al. RandAugment: practical automated data augmentation with a reduced search space[J]. arXiv:1909.13719, 2019.
- [19] ZHANG H, CISSE M, DAUPHIN Y, et al. mixup: beyond-empirical risk minimization[J]. arXiv:1710.09412, 2018.
- [20] BLUM A, MITCHELL T M. Combining labeled and unlabeled data with co-training[C]//Proceedings of the 11th Annual Conference on Computational Learning Theory, Madison, Jul 24-26, 1998. New York: ACM, 1998: 92-100.
- [21] QIAO S, SHEN W, ZHANG Z S, et al. Deep co-training for semi-supervised image recognition[J]. arXiv:1803.05984, 2018.
- [22] CHEN D D, WANG W, GAO W, et al. Tri-net for semi-supervised deep learning[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Jul 13-19, 2018: 2014-2020.
- [23] CIRESAN D, MEIER U, GAMBARDILLA L, et al. Deep, big, simple neural nets for handwritten digit recognition[J]. Neural Computation, 2010, 22(12): 3207-3220.
- [24] LAINE S, AILA T. Temporal ensembling for semisupervised learning[J]. arXiv:1610.02242, 2016.
- [25] OUALI Y, HUDELLOT C, TAMI M. An overview of deep semi-supervised learning[J]. arXiv:2006.05278, 2020.

- [26] TARVAINEN A, VALPOLA H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results[J]. arXiv:1703.01780, 2018.
- [27] XIE Q Z, DAI Z H, HOVY E H, et al. Unsupervised data augmentation for consistency training[J]. arXiv:1904.12848, 2019.
- [28] BERTHELOT D, CARLINI N, GOODFELLOW I, et al. MixMatch: a holistic approach to semi-supervised learning[J]. arXiv:1905.02249, 2019.
- [29] SOHN K, BERTHELOT D, LI C L, et al. FixMatch: simplifying semi-supervised learning with consistency and confidence[J]. arXiv:2001.07685, 2020.
- [30] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. arXiv:1406.2661, 2014.
- [31] ZOU X F, ZHU D J. Review on generative adversarial network[J]. Computer Systems & Applications, 2019, 28(11): 1-9.
邹秀芳, 朱定局. 生成对抗网络研究综述[J]. 计算机系统应用, 2019, 28(11): 1-9.
- [32] SALIMANS T, GOODFELLOW I J, ZAREMBA W, et al. Improved techniques for training GANs[J]. arXiv:1606.03498, 2016.
- [33] ODENA A. Semi-supervised learning with generative adversarial networks[J]. arXiv:1606.01583, 2016.
- [34] KHAN S. Convolution neural network and computer vision [M]. Beijing: China Machine Press, 2019.
KHAN S. 卷积神经网络与计算机视觉[M]. 黄智渊, 戴志涛, 译. 北京: 机械工业出版社, 2019.
- [35] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[D]. Toronto: University of Toronto, 2009.
- [36] COATES A, NG A, LEE H. An analysis of single-layer networks in unsupervised feature learning[C]//Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, Apr 16-18, 2011: 215-223.
- [37] LU J, GONG P H, YE J P. Learning from very few samples: a survey[J]. arXiv:2009.02653, 2020.
- [38] BACHMAN P, HJELM R D, BUCHWALTER W. Learning representations by maximizing mutual information across views[C]//Proceedings of the Annual Conference on Neural Information Processing Systems, Vancouver, Dec 8-14, 2019: 15509-15519.
- [39] HJELM R, FEDOROV A, LAVOIE-MARCHILDON S, et al. Learning deep representations by mutual information estimation[J]. arXiv:1808.06670, 2019.
- [40] ZHAI X, OLIVER A, KOLESNIKOV A, et al. S4L: self-supervised semi-supervised learning[J]. arXiv:1905.03670, 2019.
- [41] SCHMARJE L, SANTAROSSA M, KOCH K. A survey on semi, self and unsupervised learning for image classification [J]. arXiv:2002.08721, 2020.
- [42] CHEN J A, YANG Z C, YANG D Y. MixText: linguistically-informed interpolation of hidden space for semi-supervised text classification[J]. arXiv:2004.12239, 2020.
- [43] WANG P S, SONG Y, DAI L R. Fine-grained image classification with multi-channel visual attention[J]. Journal of Data Acquisition & Processing, 2019, 34(1): 157-166.
王培森, 宋彦, 戴礼荣. 基于多通道视觉注意力的细粒度图像分类[J]. 数据采集与处理, 2019, 34(1): 157-166.
- [44] CHENG W J, CHEN W Q. Hyperspectral image classification based on MCFN-attention[J]. Computer Engineering and Applications, 2020, 56(24): 201-206.
程文娟, 陈文强. 基于MCFN-Attention的高光谱图像分类[J]. 计算机工程与应用, 2020, 56(24): 201-206.



吕昊远(1997—),男,山西晋中人,硕士研究生,主要研究方向为深度学习、辐射源个体识别。

LYU Haoyuan, born in 1997, M.S. candidate. His research interests include deep learning and specific emitter identification.



俞璐(1973—),女,吉林长春人,副教授,硕士生导师,主要研究方向为多媒体信息处理、模式识别、图像处理等。

YU Lu, born in 1973, associate professor, M.S. supervisor. Her research interests include multimedia information processing, pattern recognition, image processing, etc.



周星宇(1985—),男,湖南永州人,讲师,主要研究方向为计算机视觉、人工智能安全等。

ZHOU Xingyu, born in 1985, lecturer. His research interests include computer vision, artificial intelligence security, etc.



邓祥(1994—),男,安徽六安人,硕士研究生,主要研究方向为深度学习、辐射源行为分析。

DENG Xiang, born in 1994, M.S. candidate. His research interests include deep learning and emitter behavior analysis.