计算机科学与探索
Journal of Frontiers of Computer Science and Technology
1673-9418/2022/16(06)-1279-12
doi: 10.3778/j.issn.1673-9418.2111144

深度半监督学习中伪标签方法综述

刘雅芬[1,2],郑艺峰[1,2+],江铃燚[1,2],李国和[3],张文杰[1,2]
1. 闽南师范大学 计算机学院,福建 漳州 363000
2. 数据科学与智能应用福建省高校重点实验室,福建 漳州 363000
3. 中国石油大学(北京) 信息科学与工程学院,北京 102249
+ 通信作者 E-mail: zyf@mnnu.edu.cn

摘 要:随着智能技术的发展,深度学习已成为机器学习的研究热点,在各个领域发挥着越来越重要的作用。深度学习需要大量的标签数据用于提升模型性能。为了有效解决标签问题,研究人员将半监督学习与深度学习相结合。同时利用少量的标签数据和大量的无标签数据构建模型,有利于扩大样本空间。鉴于深度半监督学习的理论意义和实际应用价值,以深度半监督学习方法中的伪标签方法作为切入点进行分析。首先,对深度半监督学习进行介绍,指出伪标签方法优势所在;其次,从自训练和多视角训练角度出发对伪标签方法进行阐述,对已有的模型进行综合性分析;接着,重点介绍基于图和伪标签的标签传播方法,并对已有伪标签方法进行实验分析;最后,从无标签数据效用性、噪声数据、合理性和伪标签方法的结合上总结伪标签方法所面临的问题和未来研究方向。

关键词:深度学习;半监督学习;伪标签;标签传播
文献标志码:A **中图分类号:**TP391.41;TP18

Survey on Pseudo-Labeling Methods in Deep Semi-supervised Learning

LIU Yafen[1,2], ZHENG Yifeng[1,2+], JIANG Lingyi[1,2], LI Guohe[3], ZHANG Wenjie[1,2]
1. College of Computer Science, Minnan Normal University, Zhangzhou, Fujian 363000, China
2. Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Zhangzhou, Fujian 363000, China
3. College of Information Science and Engineering, China University of Petroleum, Beijing 102249, China

Abstract: With the development of intelligent technology, deep learning has become a hot topic in machine learning. It is playing a more and more important role in various fields. Deep learning requires a lot of labeled data to improve model performance. Therefore, researchers effectively combine semi-supervised learning with deep learning to solve the labeled data problem. It utilizes a small amount of labeled data and a large amount of unlabeled data to build the model simultaneously. It can help to expand the sample space. In view of its theoretical significance and practical application value, this paper focuses on the pseudo-labeling methods as the starting point. Firstly, deep semi-supervised learning is introduced and the advantage of pseudo-labeling methods is pointed out. Secondly, the pseudo-labeling methods are described from self-training and multi-view training and the existing model is comprehensively analyzed. And then, the label propagation method based on graph and pseudo-labeling is introduced. Furthermore, the existing pseudo-labeling methods are analyzed and compared. Finally, the problems and future research direction

基金项目:国家自然科学基金(62141602);福建省自然科学基金(2021J011002,2021J011004);克拉玛依科技发展计划项目(2020CGZH0009)。
This work was supported by the National Natural Science Foundation of China (62141602), the Natural Science Foundation of Fujian Province (2021J011002, 2021J011004) and the Kalamay Science & Technology Research Project (2020CGZH0009).
收稿日期:2021-11-02 **修回日期:**2022-01-05

of pseudo-labeling methods are summarized from the utility of unlabeled data, noise data, rationality, and the combination of pseudo-labeling methods.

Key words: deep learning; semi-supervised learning; pseudo-labeling; label propagation

随着智能技术的发展,深度学习已得到学术界和工业界的广泛关注,尤其在计算机视觉[1]、图像处理[2-3]、自然语言处理[4-5]和语音识别[6]等领域。例如百度的无人驾驶、阿里的用户行为分析等。

深度学习以数据为驱动,其优异的性能离不开大量标签数据。然而,在现实生活中,标签数据获取代价高昂。例如:在医疗任务中,标签均由领域专家分析得出。相比于标签数据,无标签数据获取相对容易,半监督学习则将二者相结合用以训练模型。研究表明将少量有标签数据和大量无标签数据相结合有助于提高学习任务的准确率[7]。基于上述思想,研究人员将半监督学习引入到深度学习,提出深度半监督学习。根据所采用半监督损失函数和模型设计方式,深度半监督学习方法可分为:生成式方法、一致性正则化方法、基于图的方法、混合方法和伪标签方法[8]。

(1)生成式方法:生成式方法学习数据的隐式特征,假设所有数据均来自同一潜在模型,以更好地将无标签数据与学习目标关联建模,并采用最大期望进行求解[9-10]。在标签数据极少时,相比其他方法,能获得较好的性能。其关键在于与真实分布吻合程度[11-13]。

(2)一致性正则化方法:将无标签数据用以模型强化,即将一个实际的扰动应用于一个无标签的数据,亦不会使预测结果出现明显变化。对于具有不同标签的数据,在聚类假设中属于低密度区域分离,因此,数据在扰动后标签发生变化的可能性微乎其微。由此可见,可将一致性正则化项作用于损失函数,以指定假设的先验约束[14-17]。

(3)基于图的方法:在数据集上构建图,图中每个节点表示一个训练数据,每个边缘表示节点对相似性。可分为图正则化和图嵌入两种。图正则化使用Laplacian正则化,假设具有强连接边缘的节点可能共享相同的标签,例如标签传播(label propagation)[18]、高斯随机场(Gaussian random fields)[19]和局部全局一致性(local and global consistency)[20]。图嵌入则是将节点编码为向量,用于度量节点之间的相似性[21-23]。

(4)混合方法:融合伪标签、伪一致性正则化和熵最小化的思想用以提高模型性能。此外还引入一种混合物学习原理,即一种简单的、数据不可知的数据增强方法[24],一个配对的数据及其各自标签的凸组合[25-26]。

大部分深度半监督学习方法不足之处在于过分依赖特定区域的数据增强,然而在大多数应用场景下,数据增强并不容易生成,而其中伪标签方法却不受数据增强的约束。现阶段为无标签数据标注伪标签的方法则大多先利用标签数据训练模型,而后将伪标签数据与标签数据相结合扩大数据集,共同训练模型。可见,伪标签方法的性能主要依赖于所选择的模型。伪标签方法可分成自训练和多视角训练两大部分,自训练通过获得无标签数据的伪标签从而得到更多训练数据。多视角训练是通过训练多个模型,利用模型间的“分歧”给无标签数据打上伪标签。而Zhu于2002年提出的标签传播算法,无需依赖于任何的分类模型,将图和伪标签相结合,利用样本间的关系建立图模型,通过相似度给无标签节点标记标签[18]。其具备易于实现且复杂度较低的特点,已被广泛应用于虚拟社区挖掘等领域[27]。

在本文中,首先,对深度半监督学习进行分析;其次,从自训练和多视角训练两方面对伪标签方法进行详细的剖析;然后,着重阐述利用相似性且无需预训练的基于图和伪标签的标签传播方法,并讨论其优势所在;接着,对已有的伪标签方法进行实验分析的比对;最后,从无标签数据在实际应用中是否适用于所有模型、真实数据集带有噪声数据、数据采样的合理性以及伪标签方法和其他方法结合的情况总结伪标签方法所面临的问题和未来研究方向。

1 深度半监督学习

深度学习以数据为驱动,而获取大量的标签数据代价昂贵。深度半监督学习可通过少量标签数据和大量无标签数据构建模型,其无标签信息能提供更多关于数据分布的信息,从而更好地估计不同类别的决策边,有助于提高模型的性能。

近年来,随着智能信息技术的推广,机器学习方法得到广泛的研究,其主要分为:监督学习(supervised learning)、无监督学习(unsupervised learning)和半监督学习(semi-supervised learning)。

半监督学习介于监督学习和无监督学习二者之间,其基本思想是利用无标签数据提高模型的泛化

能力，以减少对外界交互的过分依赖，从而训练更好的模型。三者之间对比如图1所示。从图中可以看出，在半监督学习中，同时提供标签数据集 $D_l=\{(x_1,y_1),(x_2,y_2),\cdots,(x_l,y_l)\}$ 和无标签数据集 $D_u=\{x_{l+1},x_{l+2},\cdots,x_{l+u}\}$，且无标签数据数量远远多于标签数据，即 $l\ll u$。更具体地说，半监督学习的目标是利用无标签数据集 D_u 辅助生成预测函数 f_θ，比仅使用标签数据集 D_l 所获得的函数更准确[28]。

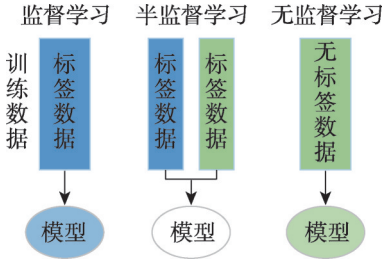

图1 监督学习、半监督学习、无监督学习结构对比

Fig.1 Structure comparison of supervised learning, semi-supervised learning and unsupervised learning

随着智能应用的普及，数据量急剧增加，数据标注标签信息代价昂贵。例如：在进行医学影像分析时，虽可获得大量的医院影像，但对影像中的病灶进行标注则需要由医学专家才能进行标注。同样，在进行商品推荐时，仅有少部分的用户愿意协助对商品进行标注。由此可见，半监督学习具有较高的应用价值。

如何有效利用无标签数据成为亟需解决的问题。无标签数据因其与标签数据从相同的数据源独立分布采样而来，虽未包含标签信息，但其分布的信息有助于模型的构建。本文给出一个直观的示例，如图2所示，图中包含一个正方形类和一个三角形类，待判别样本恰好位于两者之间，则在进行样本类别判断时仅能依靠随机猜测。倘若能观察到图中的无标签数据分布状况，则可将此待判别样本归为正方形类。由此可见，无标签数据可提供关于数据分布结构的额外信息，有助于更好地估计不同类别之间的决策边界。

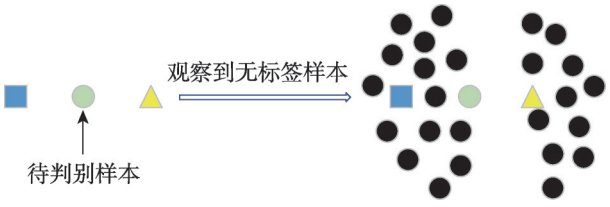

图2 无标签数据效用示例（黑点为无标签数据）

Fig.2 Unlabeled data utility example (black dots indicate unlabeled data)

最早将无标签数据应用到半监督学习中的方法是Self-training方法[29-31]，该方法使用有标签数据构建模型，进而对无标签数据进行预测，从中筛选出预测置信度高的样本加入标签数据集中，不断更新模型，直至收敛。然而要有效利用无标签数据，则必须对无标签样本所揭示的数据分布信息与类别标签之间的关系进行假设。目前，可分为聚类假设（cluster assumption）、平滑假设（smoothing assumption）和流行假设（manifold assumption）。

（1）聚类假设：当两个数据属于同一簇时，则拥有相同的类标签，即当数据 x_1 和 x_2 位于同一簇时，y_1 和 y_2 的预测结果应一致[32]。聚类假设亦称为低密度分离假设，即决策边界应位于低密度区域。

（2）平滑假设：指位于稠密数据区域的两个距离相近的数据具有相同的标签，即对于稠密区域中的两个数据，如果其存在边连接，则具有相同的标签信息[28]，反之亦然。这个假设在分类任务中很有帮助，但对回归任务没有多大的帮助。

（3）流行假设：将高维数据嵌入到低维流形中，如两个数据在低维流形中同属于一个局部邻域，则其应具有相似的类信息[33]，其着重于模型的局部特性。在该假设下，无标签数据就能使数据空间更加密集，有助于分析局部区域特征信息，从而使决策函数较好地拟合数据。

综上所述，上述三类假设虽然实现的方式不同，但其本质都是考虑样本的相似性。

近年来，深度学习在实际应用中取得优异的表现，但其以数据为驱动，需要大量标签样本用以训练模型。然而，在现实生活中，对样本进行标注代价高昂。为此，研究人员将半监督学习引入到深度学习中，提出深度半监督学习。

在早期的方法中标签数据和无标签数据分开使用，先利用无标签数据进行初始化，再利用标签数据对模型进行调整，其本质上仍是监督学习的模式。在半监督模式下，神经网络则应同时训练有标签和无标签样本，其损失函数的范式定义如下：

$$Loss=L_s+w(t)\times L_u \tag{1}$$

其中，L_s 表示为监督损失，L_u 表示为无监督损失，$w(t)$ 为权重。不同的深度半监督方法区别在于所采用的 L_u 的不同。

2 伪标签

现阶段，以一致性正规化方法为主的深度半监

督学习由于过分依赖特定区域的数据增强，不易实现。为此Lee提出伪标签方法，标签数据和无标签数据同时参与模型的训练[34]。对于无标签数据，在每次权重更新时，为每个无标签数据赋予具有最大预测概率的标签，再将标注后的无标签数据放入标签数据集用以模型训练。本章将从自训练和多视角训练两方面对伪标签方法进行详细的剖析。

2.1 自训练

自训练是基于最可信预测以此标记无标签数据[35-36]，根据模型自身生成伪标签，可分为熵最小化方法、代理标签方法、噪声学生模型方法、自半监督方法和元伪标签方法。首先，使用少量的标签数据 D_l 来训练预测模型 f_θ，再使用 f_θ 为无标签数据 $x_i \in D_u$ 分配伪标签。如果模型预测概率高于预定的阈值 τ，则将数据 $(x, \operatorname{argmax} f_\theta(x))$ 添加到标签数据集中，继续训练模型，为 $D_u - \{x_i\}$ 中的数据标记伪标签，重复上述过程，直至模型无法产生最可信预测或所有的无标签数据都标注伪标签。在实际训练过程中，可采用相对置信度决定为哪些无标签数据标记伪标签，即在每次训练后对前 n 个高置信度预测的无标签样本进行标记，并添加至标签数据集 D_l 中。Yalniz等人[37]将自训练方法用以训练ResNet-50模型，先在带有伪标签的无标签图像上进行训练，再对标签图像进行微调，实验结果表明自训练方法进一步提高训练模型的鲁棒性。

2.1.1 熵最小化方法

熵最小化方法（entropy minimization）是一种熵正则化的方法，其通过鼓励模型对无标签数据进行低熵预测，再将其应用到监督学习中以实现半监督学习。理论分析表明熵最小化有助于阻止决策边界通过高密度的数据点区域，如无法阻止则将对无标签的数据产生低置信度的预测[38]。

给定图像数据 $x \in D$，令 $f(x)$ 表示特定神经输出函数，将所有概率分布 $P_{f(x)}$ 的熵 $H(P_{f(x)})$ 最小化，上述方法仅精确神经网络的预测，无法单独使用。如果将其作为损失，则会导致预测退化。Grandvalet和Bengio考虑从标签和无标签的数据中学习决策规则，并使熵最小化方法规范化[38]。熵最小化方法可作用于任何特定的或限制最低熵规范的模型。当生成模型被错误指定时，熵最小化方法更有助于实现最低熵规范化。最新研究表明，熵最小化方法本身并不能产生有竞争力的结果，但当与不同的方法结合时，可以产生最先进的结果[39]。

2.1.2 代理标签方法

代理标签是一种估计无标签数据为伪标签的最简单方法，目标是生成代理标签用以增强学习[34]。代理标签同时将标签数据和无标签数据以监督方式进行训练，如图3所示。

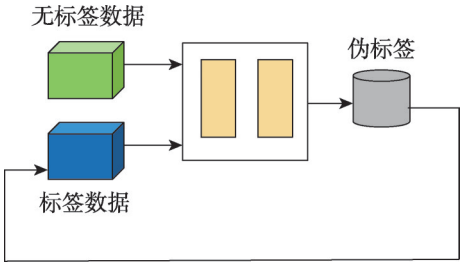

图3 代理标签模型

Fig.3 Proxy-label model

对于无标签数据，用同样的模型进行预测，筛选出最大置信度的预测作为伪标签，即具有最大的预测概率。由于标签数据和无标签数据的总数相差很大，为了保持训练平衡，将整体损失函数定义为：

$$Loss = \frac{1}{l}\sum_{m=1}^{l}\sum_{i=1}^{C} L_s(y_i^m, f_i^m) + \alpha(t)\frac{1}{u}\sum_{m=1}^{u}\sum_{i=1}^{C} L_u(y_i'^m, f_i'^m) \quad (2)$$

其中，l 是随机梯度下降中标签数据的个数，u 表示无标签数据的个数，f_i^m 为标签数据的输出类，y_i^m 为标签数据真实类，$f_i'^m$ 表示无标签数据输出类，$y_i'^m$ 为无标签数据的伪标签。$\alpha(t)$ 直接影响网络性能，如果取值过大，则标签数据会受到干扰；如果取值过小，则无法使用无标签数据。

Shi等试图确定其最优标签和最优模型参数，并通过迭代训练最小化损失函数[40]。Iscen等人将代理标签方法用于标签传播，在标签数据和伪标签数据上交替训练网络模型，同时引入两个不确定性参数，即每一个样本基于输出概率的熵（用以克服对预测的不平等置信度问题）和基于每个类得分的类种群（用以处理类的不平衡问题）[41]。Arazo等人则认为由于存在确认偏差，从而导致单纯的伪标签会过度拟合于不正确的伪标签。同时证明采用混合方式并设置每批的最少标签样本数量有助于减少上述偏差[42]。

2.1.3 噪声学生模型

噪声学生模型（noisy student）方法受知识蒸馏思想[43-44]启发，基于“教师-学生”框架，如图4所示。其具体过程：首先采用教师EfficientNet[45]模型对标签数据进行训练，为无标签数据生成伪标签，加入标签数据集；再采用规模更大的EfficientNet模型作为学生模型，在新数据集上进行训练。同时，可在学生模型

训练阶段加入 Dropout 和 Stochastic Depth 等模型噪声。经多次迭代，获更具有鲁棒性的学生模型，此时学生模型可作为教师模型，重新标注无标签数据。

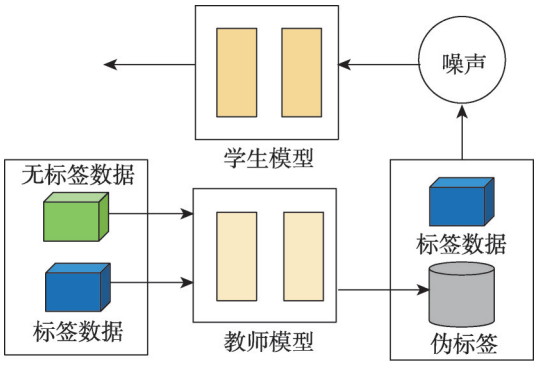

图4 噪声学生模型

Fig.4 Noisy student model

Liu 等人将噪声学生模型法用于探索药物代谢作用，可进一步加速药物发现过程，从而降低成本[46]。Kumar 等人也采用噪声学生模型方法进行面部表情的识别，模型隔离面部的不同区域，并使用多级注意机制独立进行处理。其结果表明，与其他单一模型相比，该方法更加有助于提升模型的性能[47]。

2.1.4 自半监督学习方法

自半监督学习（self-supervised semi-supervised learning）将自监督学习技术用以解决半监督图像分类的问题[48]。在自半监督学习方法中，有四个旋转度 {0°, 90°, 180°, 270°}，用以旋转输入图像，其旋转损失为旋转图像预测输出的交叉熵损失。对于无标签数据，预测其不同的旋转角度打上伪标签，后与标签数据共同训练模型，如图 5 所示。

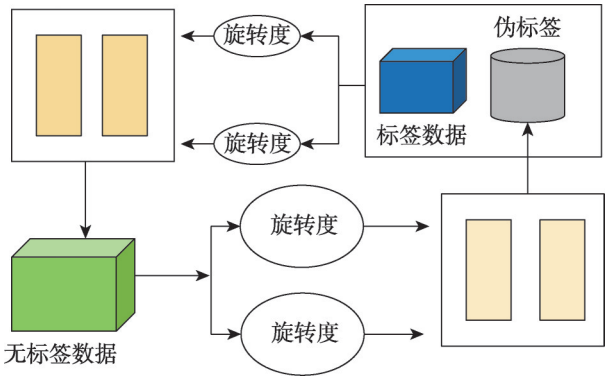

图5 自半监督学习模型

Fig.5 Self-supervised semi-supervised learning model

Beyer 等人将损失分成有监督损失和无监督损失两部分，其中监督损失为交叉熵损失，而无监督损失是基于自监督技术的旋转和样本预测的损失。同时，提出两种半监督图像分类方法，有助于解决图像分类的半监督问题[48]。

2.1.5 元伪标签方法

在半监督学习过程中，伪标签通常是由教师模型生成，不能有效适应网络训练的学习状态。为此，Pham 等人提出元伪标签（meta pseudo labels）方法，采用“学生-教师”框架[49]，如图 6 所示。在该框架中，教师模型使用元学习方法生成代理标签，并鼓励教师模型以改进学生模型学习的方式从而调整训练的目标分布，再通过评估学生模型用以更新教师模型。虽然允许教师模型调整和适应学生的学习状态，但不足以训练教师模型。为了克服上述问题，在教师模型中，还需使用验证集对标签数据进行训练。

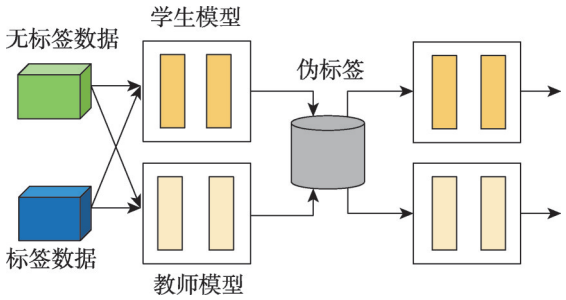

图6 元伪标签模型

Fig.6 Meta pseudo labels model

Pham 等在 CIFAR-10、SVHN 和 ImageNet 实验更进一步证明 MPL 方法的有效性。此外，在 CIFAR10 和 ImageNet 上附加额外的无标签数据，并使用 EfficientNet 进行训练。实验结果表明，采用元伪标签方法在 CIFAR-10 上获得 88.6% 的准确率，在 ImageNet 上获得 86.9% 的 top-1 准确率[49]。

自训练具备简单性和通用性，可广泛应用于各个领域。例如，图像分类、语义分割和目标对象检测等任务。但其不足之处在于，无法纠正其自身错误（即任何错误的分类都会被迅速放大）。而多视角训练在理想情况下，不同的视角可相互补充、相互协作，进而提高彼此的性能。

2.2 多视角训练

多视角训练[50-51]亦称为基于分歧的模型训练，根据不同数据视角训练的模型生成伪标签，可分为协同训练方法和三体训练方法。与自训练不同之处在于，其数据存在多个视角，例如图像的颜色信息和纹理信息。多视角训练的基本思想是同时训练多个学习模型，分别用以标记无标签的样本。

2.2.1 协同训练方法

协同训练方法(co-training)[52]是指在两个视角上训练不同的分类模型,即在标签数据上分别训练两个预测函数 f_{θ_1} 和 f_{θ_2},如图7所示。在每次迭代过程中,将 f_{θ_1} 标记的无标签数据添加到 f_{θ_2} 中,彼此交换,重复此过程,直至无标签数据耗尽或满足最大迭代次数。

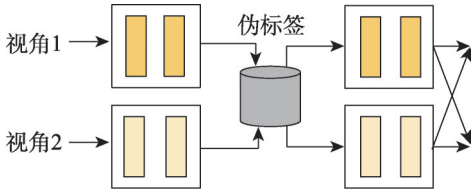

图7 协同训练模型

Fig.7 Co-training model

具体过程描述如下:令 $v_1(x)$ 和 $v_2(x)$ 表示两个不同的数据视图,使得 $x=(v_1,v_2)$。假设 C_1 为在 v_1 上训练的分类模型,C_2 表示在 v_2 上训练的分类模型,在目标函数中,协同训练方法假设定义如下:

$$L_{ct}=H\left(\frac{1}{2}(C_1(v_1)+C_2(v_2))\right)-\frac{1}{2}(H(C_1(v_1))+H(C_2(v_2))) \quad (3)$$

其中,$H(\cdot)$ 表示熵。

在标签数据集上,标准的交叉熵损失可定义为:

$$Loss=H(y,C_1(v_1))+H(y,C_2(v_2)) \quad (4)$$

其中,$H(p,q)$ 表示 p 和 q 之间的交叉熵。

对于协同训练模型,关键在于其两种视角是不同且互补的,但损失函数 L_{ct} 和 L_s 仅确保模型对于数据集上的预测趋于一致。为了解决此问题,可在协同训练模型强制引入视角差异约束。

Tran等人提出协同训练半监督回归和自适应算法,利用不同的视角增加输入数据量,并结合互相关等技术用于基于可见光的指纹技术定位。实验结果表明,随着输入数据量的增加,模型的定位精度随着增高[53]。Díaz等人提出一种使用深度神经网络的视觉对象识别的联合训练模型,通过添加多层自我监督神经网络作为视图的中间输入,视图会因其输出的交叉熵正则化而呈现多样化。该模型综合考虑输出的差异性,将协同训练和自我监督学习相结合,可称为差分自我监督共同训练(different self-supervised co-training)。结果表明,该方法虽然简单,但有助于提高模型的精度[54]。

2.2.2 三体训练方法

三体训练(tri-training)[55]试图克服多个视角存在的数据缺乏问题,从三个不同的训练集(均通过自助抽样法得到)中训练三个分类模型,有助于减少自我训练中产生的预测偏差,如图8所示。其基本思想:首先利用标签数据集训练三个预测函数 f_{θ_1}、f_{θ_2} 和 f_{θ_3}。令 x 表示无标签数据,若其在 f_{θ_1} 和 f_{θ_2} 上预测结果一致,则认为伪标签自信且稳定。此时,将标记好的 x 添加到 f_{θ_3} 的标签数据集中,再对其进行微调。如果无数据点再被添加到任何模型的训练集中,则训练停止。在整个增强过程中,三个模型会变得越来越相似。因此,需分别在训练集上进行微调,以确保模型多样性。根据所采用的框架不同,三体训练可分为多任务三体训练(multi-task tri-training)[56]和交叉视图训练(cross-view training)[57]。

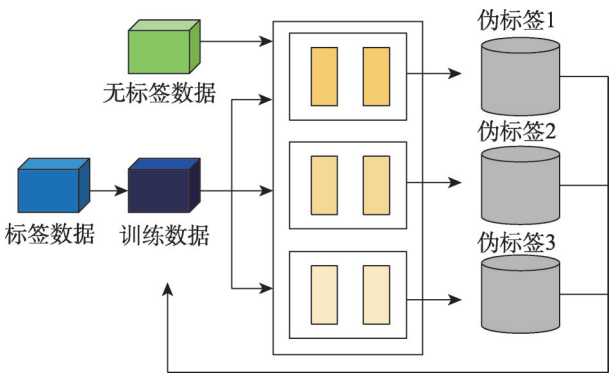

图8 三体训练模型

Fig.8 Tri-training model

(1)多任务三体训练:使用神经网络的三体训练代价昂贵,需要对三个模型中的所有无标签进行预测。为了缓解上述问题,Ruder和Plank[56]将迁移学习思想引入半监督学习中,提出多任务三体训练方法,三个模型与特定于模型的分类层共享相同的特征提取器,将模型与一个额外的正交约束联合训练,从而进一步减少时间和空间复杂度。多任务三体训练不再单独训练模型,而是采用共享参数,并用多任务学习机制进行联合训练。需要注意的是,由于模型作用相同,其属于伪多任务学习。

(2)交叉视图训练:Clark等人[58]结合多视角学习和一致性训练,提出交叉视图训练,对于不同的输入视图能获得一致的预测输出。其基本思想是采用共享编码器,再添加辅助预测模块,将编码器表示转换为预测输出。可将上述模块分为辅助学生模块和初级教师模块,二者具有一致的预测。学生预测模块可以从教师模块的预测中学习,既提高编码器产生表示的质量,也有助于改进使用相同共享表示的完整模型。

在车辆识别中，不同视点下车辆的视觉外观会发生显著变化。为此，Yang等人提出弱监督交叉视图学习模块，用于车辆的重识别，仅通过基于车辆入侵检测系统最小化交叉视角特征距离，而不使用任何视角标注来学习一致的特征表示。该模型在VeRi-776、VehicleID、VRIC和VRAI数据集上均获得显著的性能改进[59]。

3 标签传播

基于伪标签的深度半监督学习方法均需使用标签数据训练模型，继而标注无标签数据，其算法复杂度较高。而将伪标签方法与基于图的方法相结合可解决训练模型复杂度高和数据分布形状局限的问题。本章主要介绍标签传播方法，即为二者相结合的深度半监督学习方法，其满足聚类假设和流行假设，即同一簇和同一流行中的数据可能共享相同的标签，利用簇的结构和节点间的相似性，将标签数据标签传播给无标签数据，具有运算简单和复杂度小的特点。

3.1 基于图的半监督学习

基于图的半监督学习基本假设是，对于每个数据样本 x_i（既包括标签数据又包括无标签数据），都可表示为图的一个节点，每条边上的权值表示节点对的相似性。根据上述内容，可将图表示为 $G(V,E)$，$V=\{x_1,x_2,\cdots,x_n\}$ 表示节点集合，$E=\{e_{ij}\}_{i,j=1}^{n}$ 表示边集合。此外，可用图 G 的邻接矩阵 $\boldsymbol{A}$ 描述图的结构，每个元素为对应两点间的非负权重（可用相似性度量进行推导），若两个节点之间不连接，则 $A_{ij}=0$。

周志华认为基于图形的半监督学习概念清晰，且易通过对所涉矩阵运算的分析来阐述其性质[60]。其不足之处在于存储开销成本较大。此外，在图构建过程仅依赖于训练样本集，对于新数据样本，难以判断其在图中的位置。Yi等人建立了一种自适应的基于图的标签传播模型，解决了非负矩阵分解不能充分利用标签信息的弱点，采用局部约束来反映数据的局部结构，迭代优化算法求解目标函数。实验结果表明，该框架具有优异的性能[61]。

3.2 基于图和伪标签的标签传播

标签传播主要假设是流行假设，即属于同一流形中的数据样本很可能共享相同的语义标签。为此，标签传播根据数据流形结构和中间节点相似性，将标签数据的标签传播给无标签数据[62]。

首先，根据给定数据构建图，若假设图为完全图，则节点 x_i 和 x_j 边的权重可表示为：

$$w_{ij}=\exp\left(-\frac{\|x_i-x_j\|^2}{\alpha^2}\right) \tag{5}$$

其中，α 是超参数。

标签传播算法通过相邻节点之间传播标签，若节点间的权重越大，则表示其相似程度越高，标签越容易传播。为此，概率转移矩阵 $\boldsymbol{P}$ 可定义为：

$$\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} \\ p_{21} & p_{22} & \cdots & p_{2j} \\ \vdots & \vdots & & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} \end{bmatrix} \tag{6}$$

其中，p_{ij} 表示从节点 x_i 转移到节点 x_j 的概率。

$$p_{ij}=p(i\rightarrow j)=\frac{w_{ij}}{\sum_{k=1}^{n}w_{ik}} \tag{7}$$

假设数据集中有 Y 个类和 l 个标签样本，则定义一个 $l\times Y$ 的标签数据矩阵 $\boldsymbol{F}^l$：

$$\boldsymbol{F}^l=\begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1j} \\ f_{21} & f_{22} & \cdots & f_{2j} \\ \vdots & \vdots & & \vdots \\ f_{i1} & f_{i2} & \cdots & f_{ij} \end{bmatrix} \tag{8}$$

其第 i 行表示第 i 个样本的标签指示向量，即若第 i 个样本的类别为 Y_k，则第 k 个元素为1，其他为0。

为了便于说明，将上述标签数据矩阵表示为 $\boldsymbol{F}^l=[f_1,f_2,\cdots,f_l]^{\mathrm{T}}$。

同样对于 u 个无标签样本定义一个 $u\times Y$ 无标签数据矩阵 $\boldsymbol{F}^u$：

$$\boldsymbol{F}^u=\begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1j} \\ f_{21} & f_{22} & \cdots & f_{2j} \\ \vdots & \vdots & & \vdots \\ f_{u1} & f_{u2} & \cdots & f_{uj} \end{bmatrix} \tag{9}$$

值得注意，其数值初始可进行[0,1]之间随机初始化。为了便于说明，将上述无标签数据矩阵表示为 $\boldsymbol{F}^u=[f_{l+1},f_{l+2},\cdots,f_{l+u}]^{\mathrm{T}}$。

将 $\boldsymbol{F}^l$ 和 $\boldsymbol{F}^u$ 合并得到标签向量矩阵 $\boldsymbol{F}=[\boldsymbol{F}^l:\boldsymbol{F}^u]$。

标签传播算法的具体过程如下：

（1）执行传播 $\boldsymbol{F}=\boldsymbol{PF}$；

（2）重置 $\boldsymbol{F}$ 中前 l 行标签样本的标签 $\boldsymbol{F}_l=\boldsymbol{F}^l$；

（3）重复步骤（1）、（2）直至 $\boldsymbol{F}$ 收敛。

上述过程中，步骤（1）表示将矩阵 $\boldsymbol{P}$ 和矩阵 $\boldsymbol{F}$ 相乘，即对于每个节点按传播概率将其周围节点传播的标注值按权重相加，并更新自身的概率分布。两个节点越相似（在欧式空间中距离越近），则对方的伪标签会越容易受影响。对于步骤（2），由于标签数据的标签是事先确定的，在每次传播后，需要回归其

初始标签。随着标签数据不断将其标签传播出去，最后的类边界会穿越高密度区域，而停留在低密度的间隔中。

在每次迭代过程中，需对 $\boldsymbol{F}=[\boldsymbol{F}^{l}:\boldsymbol{F}^{u}]$ 进行计算，由于 $\boldsymbol{F}^{l}$ 已知，且需要重新恢复初始值，$\boldsymbol{F}^{u}$ 是最终结果，于是可将矩阵 $\boldsymbol{P}$ 表示如下：

$$\boldsymbol{P}=\begin{bmatrix}P_{ll} & P_{lu}\\ P_{ul} & P_{uu}\end{bmatrix} \tag{10}$$

$\boldsymbol{F}^{u}$ 计算方式可表示为：

$$\boldsymbol{F}^{u}\leftarrow P_{uu}\boldsymbol{F}^{u}+P_{ul}\boldsymbol{F}^{l} \tag{11}$$

重复此步骤直至收敛。

近年来，社交媒体已广泛应用于各个领域，影响最大化（influence maximization，IM）已成为社会网络分析研究的热点问题之一。Kumar等人提出一种基于节点播种、标签传播和社团检测的影响最大化算法，其使用扩展h指数中心性检测种子节点，再使用标签传播技术检测群落[63]。经典的标签传播方法不足之处在于无法有效地联合节点属性和标签，且在大规模图上收敛速度较慢。为解决上述问题，Xie等提出一种基于图结构数据的可伸缩半监督节点分类方法（简称为GraphHop）[64]，其使用适当的初始标签嵌入向量。模型主要包括：标签聚合和标签更新。在标签聚合过程中，每个节点将前一次迭代得到的相邻节点的标签向量进行聚合；在标签更新过程中，利用邻域信息，根据节点本身的标签和其所得到的聚合标签信息，为每个节点预测新的标签向量。实验结果表明，GraphHop在各种规模的图表中均能取得较好的结果。王俊斌对标签传播算法进行扩展，提出基于成对约束的标签传播算法，将先验信息保存到成对关系矩阵中，并采用成对关系与聚类结果之间的差异来代替划分矩阵之间的差异。同时，通过构建一种新的最优化模型，将标签传播算法的最优化问题转化为谱聚类问题，并通过特征值分解方法进行求解[65]。

4 实验分析

本章将介绍各类半监督伪标签方法所采用的数据集，同时对各种伪标签方法进行实验分析对比。

4.1 实验数据集介绍

在实验分析过程中，本文主要采用UCI（University of California，Irvine）数据集和图像数据集进行实验比对。UCI数据集主要包括Iris、Cmc（contraceptive method choice）和Iono（Ionosphere），数据集信息如表1所示。在实验过程中，为了保证实验结果的有效性，需对每个数据集进行归一化处理，并划分训练集和验证集。在进行半监督训练时，标签数据占训练集的10%，采取分层抽样的方式对每个类别进行采样。

表1 实验中使用的UCI数据集

Table 1 UCI datasets used in experiment

数据集	节点（样本）	特征	类别	类别分布
Iris	150	4	3	50,50,50
Cmc	1 473	9	3	629,333,511
Iono	351	34	2	225,126

图像数据集主要包括ILSVRC-2012[66]（多用于自训练）、CIFAR-10（多用于多视角训练）和CIFAR-100[67]（多用于多视角训练）。

ILSVRC-2012是ImageNet的子集，包含1 000个图像类别，其中训练集包含120万张图像，验证集和测试集共包含15万张图像。由于类别的数量较多，通常会将精确度设置为Top-1和Top-5。Top-1准确度是指一个预测与一个真实标签相比较的经典准确度，而Top-5准确性则是检查一个基本真实标签是否在一组最多5个预测中。本文实验过程中所给出的结果为仅使用10%的标签进行训练的Top-1准确度。

CIFAR-10和CIFAR-100是大小为32×32的彩色自然图像大型数据集，其中CIFAR-10包含10个类别，CIFAR-100包含100个类别。均使用5万张图像用于训练，1万张图像用于测试。本文实验过程中，对于CIFAR-10，使用从训练集中随机选择的4 000张图像作为标签数据，其余的图像作为无标签数据；对于CIFAR-100，则是随机挑选10 000张图像作为标签数据，其余的图像作为无标签数据。

4.2 实验结果分析

为了对已有的伪标签方法进行分析，本文分别在图像数据集和UCI数据集上进行实验，具体结果分别如表2和表3所示。其中图像数据集CIARF-10和CIFAR-100还未在自训练模型实验中大规模投入使用。因此，为保证实验的公平性，自训练模型仍然以ILSVRC-2012为主。

表2主要描述图像数据集中不同方法的实验结果，其中自半监督模型在不同数据集上均取得最高的准确率。自半监督模型为混合模型，将自监督旋转预测、VAT（virtual adversarial training）、交叉熵损失和fine-tuning结合到一个具有多个训练步骤的单一模型中[48]。此外，其将损失函数分为有监督和无监督的部分，其监督损失为交叉熵损失，而无监督损失则采用旋转和范例的自监督技术。由此可见，基于伪

表2 伪标签方法在不同图像数据集上实验结果

Table 2 Experimental results of pseudo-labeling method on different image datasets %

方法	CIFAR-10	CIFAR-100	ILSVRC-2012
熵最小化	86.41	—	83.39
代理标签	—	—	82.41
噪声学生	—	—	88.39
元伪标签	88.62	—	90.20
自半监督	—	—	91.23
协同训练	90.97	65.37	—
三体训练	91.55	70.26	—

标签半监督学习方法仍然有着很大的进步空间。此外，从实验结果不难发现，随着数据样本的类别增多，模型的不确定程度逐渐增大，精确率随之下降。在相同的数据集上，三体训练方法效果也都优于协同训练方法，因三体训练方法同时使用半监督学习和集成学习机制，进一步提升学习性能。综上所述，随着基于伪标签半监督学习方法的发展，模型的识别准确率逐渐提高。而随着所使用的架构复杂程度增加，可以预测模型精度亦会随着时间的推移而提高。

表3主要描述在3个不同的UCI数据集中，协同训练、三体训练和标签传播方法在kNN（k nearest neighborhood）上的实验效果（$k=10$）。为了更好地挑拣出结果的差异，采用十折交叉验证方式。从结果可以看出，标签传播方法优于前两者。模型的训练与数据的分布情况直接有关，标签传播主要假设是流行假设（即属于同一流形中的数据样本很可能共享相同的语义标签），可获得更好的实验结果。协同训练要求数据能够从不同的角度提取出两份不同的数据，即使用同一份数据构造出两个分类器，然而现实的数据大多缺乏多个视角。而三体训练能有效解决协同训练缺乏视角的问题，相比协同训练，其在UCI数据集和图像数据集上均表现出更好的性能。但需要注意的是，基于图的标签传播无法有效地联合节点属性并且具有很强的随机性从而导致结果不稳定。在后续的工作中，可对此进行研究。

表3 伪标签方法在不同UCI数据集上实验结果

Table 3 Experimental results of pseudo-labeling method on different UCI datasets %

方法	Iris	Cmc	Ionosphere
协同训练	75.21	32.33	63.82
三体训练	80.03	35.92	64.13
标签传播	85.02	40.95	67.65

5 问题与挑战

尽管基于伪标签的深度半监督学习已取得有效的进展，但仍存在有待研究的开放研究问题。

（1）无标签数据效用性：在半监督学习中，人们普遍认为无标签数据可以提高学习性能，特别是在标签数据稀缺的情况下。值得注意的是，无标签数据可以提高学习性能是在适当的假设或条件下，一些研究表明[68-69]，使用无标签的数据可能导致性能退化。现有的基于伪标签的深度半监督方法主要使用无标签数据来生成约束，然后与标签数据共同更新模型。一般情况下，使用权衡因子用于平衡监督和无监督的损失，即所有无标签数据等权。然而，并非所有的无标签数据在实际应用中都同样适用于该模型。此时，需考虑无标签数据的权重问题。

（2）噪声数据：本文所提到的标签数据均认为是准确的，从而可以学习标准的交叉熵损失函数。然而现实生活中得到的标签数据可能带有噪声，在训练时只能训练带有噪声的数据集。在基于图的半监督学习中，为增强数据预测的一致性，引入一种由稀疏编码实现的L_1范数形式的Laplacian正则化[70]。从记忆效应的角度提出了一种协同训练和平均教师相结合的学习范式。还可对数据进行预处理，降低噪声数据带来的损失[71]。

（3）合理性：在标签传播方法中，目前大多采用有放回的取样方式，使得样本在下次采样时仍然有可能被抽取到，这面临的问题是有时取到的样本集不能代表整体，从而降低其合理性，通过计算可得约有36.8%的样本未出现在采集数据集中[63]。在之后的工作中，可对群优化进行研究，群优化的核心价值在于研究和探索“个体与总体之间的冲突和求得一致结果的条件”，进而提升数据采样的合理性。

（4）方法的结合：在调查过程中发现，一些平常的方法与伪标签方法结合在一起会显示出超乎预期的效果，第3章有相应的介绍。然而，目前只有少数方法与伪标签方法相结合，而合理的组合策略有助于进一步提高模型的性能，因此，不同思想的相结合的融合策略是一个值得探索的未来研究领域。

6 结束语

本文首先介绍深度半监督学习，可根据半监督损失函数和模型中最显著的特征，将其分为生成式方法、一致性正则化方法、基于图的方法、伪标签方法和混合方法。本文以伪标签方法作为切入点展开

详细的叙述,该方法旨在标签数据上训练模型,用以预测无标签数据的类别(即伪标签),再将新生成的伪标签数据扩充训练数据。针对伪标签方法需预训练模型这一问题展开讨论,引入基于图的标签传播方法,即无需经过预训练模型就可得到伪标签。此外,本文进一步阐述标签传播方法的基本思想,其利用数据的分布及其内在关系(即样本间的相似关系),用以标记无标签数据。最后,本文对伪标签学习研究过程中所存在的问题进行总结,并提出未来的研究方向。

参考文献:

[1] SZELISKI R. Computer vision[M]. Berlin, Heidelberg: Springer, 2011.

[2] WANG X, SOUMITRA G, SUN W G. Quantitative quality control in microarray image processing and data acquisition [J]. Nucleic Acids Research, 2019, 29(15): 75-80.

[3] BILLINGSLEY F C. Applications of digital image processing [J]. Applied Optics, 1970, 9(2): 101-106.

[4] TSURUOKA Y. Deep Learning and natural language processing [J]. Brain and Nerve, 2019, 71(1): 45-55.

[5] DOWNS J M. Identifying suicidal adolescents from mental health records using natural language processing[J]. Studies in Health Technology and Informatics, 2019, 264: 413-417.

[6] 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展, 2013, 50(9): 1799-1804.
YU K, JIA L, CHEN Y Q, et al. Deep learning yesterday, today and tomorrow[J]. Journal of Computer Research and Development, 2013, 50(9): 1799-1804.

[7] OLIVIER C, BERNHARD S, ALEXANDER Z. Semi-supervised learning[J]. IEEE Transactions on Neural Networks, 2009, 20(3): 542-543.

[8] YANG X L, SONG Z X, IRWIN K, et al. A survey on deep semi-supervised learning[J]. arXiv:2103.00550, 2021.

[9] MILLER D J, UYAR H S. A mixture of experts classifier with learning based on both labelled and unlabelled data[C]//Advances in Neural Information Processing Systems 9, Denver, Dec 2-5, 1996. Cambridge: MIT Press, 1997: 571-577.

[10] NIGAM K, MCCALLUM A, THRUN S, et al. Text classification from labeled and unlabeled documents using EM[J]. Machine Learning, 2000, 39(2/3): 103-134.

[11] SPRINGENBERG J T. Unsupervised and semi-supervised learning with categorical generative adversarial networks[J]. Computer Science, 2015, 42(7): 70-85.

[12] ABBASNEJAD M E, DICK A R, VAN DEN HENGEL A. Infinite variational autoencoder for semi-supervised learning [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 781-790.

[13] JOY T, SCHMON S M, TORR P H, et al. Rethinking semi-supervised learning in VAEs[J]. arXiv:2006.10102, 2020.

[14] ZHU X J. Semi-supervised learning literature survey[D]. Madison: University of Wisconsin, 2005.

[15] SAJJADI M, JAVANMARDI M, TASDIZEN T. Regularization with stochastic transformations and perturbations for deep semi supervised learning[C]//Advances in Neural Information Processing Systems 29, Barcelona, Dec 5-10, 2016. Red Hook: Curran Associates, 2016: 1163-1171.

[16] IZMAILOV P, PODOPRIKHIN D, GARIPOV T, et al. Averaging weights leads to wider optima and better generalization [C]//Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence, Monterey, Aug 6-10, 2018: 876-885.

[17] PARK S, PARK J, SHIN S, et al. Adversarial dropout for supervised and semi-supervised learning[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, Feb 2-7, 2018. Menlo Park: AAAI, 2018: 3917-3924.

[18] ZHU X J. Learning from labeled and unlabeled data with label propagation[R]. Carnegie Mellon University, 2002: 1-8.

[19] ZHU X J, GHAHRAMANI Z, LAFFERTY J D. Semi-supervised learning using Gaussian fields and harmonic functions[C]//Proceedings of the 20th International Conference on Machine Learning, Washington, Aug 21-24, 2003. Menlo Park: AAAI, 2003: 912-919.

[20] ZHOU D Y, BOUSQUET O, LAL T N, et al. Learning with local and global consistency[C]//Advances in Neural Information Processing Systems 16, Vancouver, Dec 8-13, 2003. Cambridge: MIT Press, 2003: 321-328.

[21] WANG D X, CUI P, ZHU W W. Structural deep network embedding[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, Aug 13-17, 2016. New York: ACM, 2016: 1225-1234.

[22] CAO S, LU W, XU Q. Deep neural networks for learning graph representations[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence, Phoenix, Feb 2-17, 2016. Menlo Park: AAAI, 2016: 1145-1152.

[23] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.

[24] ZHANG H, CISS M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization[J]. arXiv:1710.09412, 2017.

[25] VERMA V, LAMB A, KANNALA J, et al. Interpolation consistency training for semi-supervised learning[J]. Neural Networks, 2019, 145: 90-106.

[26] BERTHELOT D, CARLINI N, GOODFELLOW I J, et al. Mixmatch: a holistic approach to semi-supervised learning [C]//Advances in Neural Information Processing Systems 32, Vancouver, Dec 8-14, 2019: 5050-5060.

[27] XIE J R, SZYMANSKI B K. LabelRank: a stabilized label propagation algorithm for community detection in networks [C]//Proceedings of the 2nd IEEE Network Science Work-

shop, Thayer Hotel, Apr 29-May 1, 2013. Washington: IEEE Computer Society, 2013: 138-143.

[28] OLIVIER C, BERNHARD S, ALEXANDER Z. Semi-supervised learning[J]. IEEE Transactions on Neural Networks, 2009, 20(3): 542.

[29] DAVID Y. Unsupervised word sense disambiguation rivaling supervised methods[C]//Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, Cambridge, Jun 26-30, 1995. Stroudsburg: ACL, 1995: 189-196.

[30] SCUDDER H. Probability of error of some adaptive pattern-recognition machines[J]. IEEE Transactions on Information Theory, 1965, 11(3): 363-371.

[31] RILOFF E. Automatically generating extraction patterns from untagged text[C]//Proceedings of the 13th National Conference on Artificial Intelligence and 8th Innovative Applications of Artificial Intelligence Conference, Portland, Aug 4-8, 1996. Menlo Park: AAAI, 1996: 1044-1049.

[32] 李南. 基于聚类假设的数据流分类算法[J]. 模式识别与人工智能, 2017, 30(1): 1-10.
LI N. Clustering assumption based classification algorithm for stream data[J]. Pattern Recognition and Artificial Intelligence, 2017, 30(1): 1-10.

[33] XI W, JANG U, CHEN L, et al. Manifold assumption and defenses against adversarial perturbations[J]. arXiv: 1711.08001, 2017.

[34] LEE D H. Pseudo-Label: the simple and efficient semisupervised learning method for deep neural networks[C]//Proceedings of the Workshop: Challenges in Representation Learning, Atlanta, Jun 16-21, 2013. New York: ACM, 2013: 1-6.

[35] YAROWSKY D. Unsupervised word sense disambiguation rivaling supervised methods[C]//Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, Cambridge, Jun 26-30, 1995. Stroudsburg: ACL, 1995: 189-196.

[36] SCUDDER H. Probability of error of some adaptive pattern-recognition machines[J]. IEEE Transactions on Information Theory, 1965, 11(3): 363-371.

[37] YALNIZ I Z, JÉGOU H, CHEN K, et al. Billion-scale semi-supervised learning for image classification[J]. arXiv: 1905.00546, 2019.

[38] GRANDVALET Y, BENGIO Y. Semi-supervised learning by entropy minimization[C]//Proceedings of the Conférence Francophone sur l'apprentissage Automatique, Nice, 2005: 281-296.

[39] OLIVER A, ODENA A, RAFFEL C, et al. Realistic evaluation of deep semi-supervised learning algorithms[C]//Advances in Neural Information Processing Systems 31, Montréal, Dec 3-8, 2018: 3239-3250.

[40] SHI W W, GONG Y H, DING C, et al. Transductive semi-supervised deep learning using min-max features[C]//LNCS 11209: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 311-327.

[41] ISCEN A, TOLIAS G, AVRITHIS Y, et al. Label propagation for deep semi-supervised learning[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 5070-5079.

[42] ARAZO E, DIEGO O, PAUL A, et al. Pseudo-labeling and confirmation bias in deep semi-supervised learning[J]. arXiv: 1908.02983, 2019.

[43] XIE Q, LU M, HOVY E H, et al. Self-training with noisy student improves ImageNet classification[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Jun 13-19, 2020. Piscataway: IEEE, 2020: 10684-10695.

[44] HINTON G E, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. arXiv:1503.02531, 2015.

[45] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[J]. arXiv:1905.11946v5, 2019.

[46] LIU Y, LIM H, XIE L. Exploration of chemical space with partial labeled noisy student self-training for improving deep learning: application to drug metabolism[EB/OL]. [2021-08-23]. https://doi.org/10.1101/2020.08.06.239988.

[47] KUMAR V, RAO S, YU L. Noisy student training using body language dataset improves facial expression recognition [C]//LNCS 12535: Proceedings of the 16th ECCV Workshops on Computer Vision, Glasgow, Aug 23-28, 2020. Cham: Springer, 2020: 756-773.

[48] BEYER L, ZHAI X H, OLIVER A, et al. S4L: self-supervised semi-supervised learning[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Oct 27-Nov 2, 2019. Piscataway: IEEE, 2019: 1476-1485.

[49] PHAM H, DAI Z H, XIE Q Z, et al. Meta pseudo labels[J]. arXiv:2003.10580, 2020.

[50] KUMAR A, DAUMÉ Ⅲ H. A co-training approach for multi-view spectral clustering[C]//Proceedings of the 28th International Conference on Machine Learning, Bellevue, Jun 28-Jul 2, 2011. Madison: Omni Press, 2011: 393-400.

[51] ZHAO J, XIE X J, XU X, et al. Multi-view learning overview: recent progress and new challenges[J]. Information Fusion, 2017, 38: 43-54.

[52] 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.

[53] TRAN H Q, HA C. Reducing the burden of data collection in a fingerprinting-based VLP system using a hybrid of improved co-training semi-supervised regression and adaptive boosting algorithms[J]. Optics Communications, 2021, 488: 126857.

[54] DÍAZ G, PERALTA B, CARO L A, et al. Co-training for visual object recognition based on self-supervised models using a cross-entropy regularization[J]. Entropy, 2021, 23(4): 423.

[55] 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.

[56] RUDER S, PLANK B. Strong baselines for neural semi-supervised learning under domain shift[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Jul 15-20, 2018. Stroudsburg: ACL, 2018: 76-85.
[57] CLARK K, LUONG M T, MANNING C D, et al. Semi-supervised sequence modeling with cross-view training[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Oct 31-Nov 4, 2018. Stroudsburg: ACL, 2018: 1914-1925.
[58] CLARK K, LUONG T, LE Q V. Cross-view training for semi-supervised learning[C]//Proceedings of the 2018 Conference Acceptance Decision, Vancouver, Apr 30-May 3, 2018: 1-14.
[59] YANG L, LIU H B, ZHOU J H, et al. Pluggable weakly-supervised cross-view learning for accurate vehicle reidentification[J]. arXiv:2103.05376, 2021.
[60] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
ZHOU Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016.
[61] YI Y, CHEN Y Q, WANG J Z, et al. Joint feature representation and classification via adaptive graph semi-supervised nonnegative matrix factorization[J]. Signal Processing: Image Communication, 2020, 89: 115984.
[62] OUALI Y, HUDELOT C, TAMI M. An overview of deep semi-supervised learning[J]. arXiv:2006.05278, 2020.
[63] KUMAR S, SINGHLA L, JINDAL K, et al. IM-ELPR: influence maximization in social networks using label propagation based community structure[J]. Applied Intelligence, 2021, 51(11): 7647-7665.
[64] XIE T, WANG B, KUO C C J. GraphHop: an enhanced label propagation method for node classification[J]. arXiv: 2101.02326, 2021.
[65] 王俊斌. 基于标签传播的半监督聚类算法研究[D]. 太原: 山西大学, 2020.
WANG J B. Research on semi-supervised clustering algorithm based on label propagation[D]. Taiyuan: Shanxi University, 2020.
[66] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems 25, Lake Tahoe, Dec 3-6, 2012. Red Hook: Curran Associates, 2012: 1106-1114.
[67] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[J]. Handbook of Systemic Autoimmune Diseases, 2009, 1(4): 1-60.
[68] SINGH A, NOWAK R D, ZHU X J. Unlabeled data: now it helps, now it doesn't[C]//Proceedings of the 22nd Annual Conference on Neural Information Processing Systems, Vancouver, Dec 8-11, 2008. Red Hook: Curran Associates, 2008: 1513-1520.
[69] YANG T, PRIEBE C E. The effect of model misspecification on semi-supervised classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(10): 2093-2103.
[70] LU Z W, WANG L W. Noise-robust semi-supervised learning via fast sparse coding[J]. Pattern Recognition, 2015, 48(2): 605-612.
[71] HAN B, YAO Q M, YU X R, et al. Co-teaching: robust training of deep neural networks with extremely noisy labels[C]//Advances in Neural Information Processing Systems 31, Montréal, Dec 3-8, 2018: 8536-8546.

刘雅芬(1999—),女,福建南平人,硕士研究生,CCF会员,主要研究方向为机器学习、深度学习。
LIU Yafen, born in 1999, M.S. candidate, member of CCF. Her research interests include machine learning and deep learning.

郑艺峰(1980—),男,福建漳州人,博士,讲师,硕士生导师,CCF会员,CAAI会员,主要研究方向为人工智能、机器学习、深度学习。
ZHENG Yifeng, born in 1980, Ph.D., lecturer, M.S. supervisor, member of CCF and CAAI. His research interests include artificial intelligence, machine learning and deep learning.

江铃燚(1998—),女,福建福州人,硕士研究生,CCF会员,主要研究方向为机器学习、深度学习。
JIANG Lingyi, born in 1998, M.S. candidate, member of CCF. Her research interests include machine learning and deep learning.

李国和(1965—),男,福建平和人,博士,教授,博士生导师,主要研究方向为人工智能、机器学习、知识发现。
LI Guohe, born in 1965, Ph.D., professor, Ph.D. supervisor. His research interests include artificial intelligence, machine learning and knowledge discovery.

张文杰(1984—),男,福建漳州人,博士,教授,硕士生导师,主要研究方向为移动边缘计算、物联网。
ZHANG Wenjie, born in 1984, Ph.D., professor, M.S. supervisor. His research interests include mobile edge computing and Internet of things.