

文章编号: 2095-2163(2022)12-0030-06

中图分类号: TP391

文献标志码: A

基于实例相关标签噪声的消除算法综述

蔡宇佳¹, 陈旋¹, 覃芹¹, 陈丽绵¹, 张利^{1,2}

(1 贵州大学 大数据与信息工程学院, 贵阳 550025; 2 贵州大学 省部共建公共大数据国家重点实验室, 贵阳 550025)

摘要: 实例相关的标签噪声广泛存在于现实世界的数据集中, 尽管近年来国内外许多学者针对实例相关的标签噪声消除算法进行研究, 但缺乏对其的详细分析。基于此, 本文对实例相关标签噪声的消除算法进行检索, 综述了近年来各研究文献中的主要技术, 通过显式和隐式的方法对其进行分类研究和系统总结, 选取部分算法进行实验对比, 并做出展望。

关键词: 标签噪声; 实例相关; 显式方法; 隐式方法

A survey of algorithms based on instance correlation label noise

CAI Yujia¹, CHEN Xuan¹, QIN Qin¹, CHEN Limian¹, ZHANG Li^{1,2}

(1 College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China;

2 State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China)

[Abstract] Instance-related label noise widely exists in real-world datasets. Although instance-related label noise elimination algorithms have been studied by many researchers at home and abroad in recent years, detailed analysis of this subject is lacking. Based on this, this paper searches for the elimination algorithms of case-related label noise, reviews the main techniques in the literatures in recent years, analyzes and summarizes their classification by explicit and implicit methods, selects some algorithms for experimental comparison, and makes a prospect.

[Key words] label noise; instance correlation; explicit method; implicit method

0 引言

目前, 在机器学习领域, 例如图像分类等各种视觉问题在深度学习方面相继涌现出一批科研成果, 尽管计算机硬件系统在图像处理等方面得到提升, 网络训练方面的技术也在不断取得突破, 但要获得良好的性能却需要大量的数据作为支撑。随着大数据技术的蓬勃发展, 各种分类系统也日趋成为研究热点, 这些系统也需要大量的标注数据才能得到充分的训练, 但标注过程困难、且昂贵。在一些分类模型中, 若使用标注大量错误信息的数据集, 则会导致其结果准确性并不高。而诸如众包^[1]等情况在现实世界中将会产生大量的标签噪声, 尽管在数十年前就已开始对其进行研究^[2-7], 但迄今为止却依然存在各种各样的问题。

在文献[4-5]中, 将标签噪声分成2种类型: 特征噪声和标签噪声。其中, 特征噪声影响该特征的

观测值, 标签噪声则改变分配给实例的观测标签。有研究表明^[5], 标签噪声比特征噪声更为复杂。并且, 标签噪声问题也是当前学界面临的重大挑战, 比如在医疗图像领域中, 几乎通过医学设备诊断测试出的数据都不是完全正确的^[3-4]; 在射电天文图像处理模型的训练过程中^[8-9], 通常会依赖正确的标签, 从而对模型训练带来影响。在标签噪声分类问题中, 基于实例相关的标签噪声更接近于真实世界的情况, 因此基于实例相关的标签噪声学习算法的研究具有重要意义。

现如今, 基于实例标签噪声消除问题在深度学习领域已取得可观研究进展, 而现有的方法通常在许多实际应用中存在不实用的先验条件, 如需要干净标签的辅助设备^[1]或大量的先验信息^[10], 这些方法耗时、耗力, 较为麻烦。一般情况下, 使用显式和隐式两种处理方式对实例标签噪声结构模型进行分类研究。在显式处理方式中, 主要有2种经典方法:

基金项目: 国家自然科学基金(11963003, 12273007); SKA 专项资助(2020SKA0110300); 贵州省科技厅(黔科合基础-ZK[2022]一般143); 国家重点研发计划(2018YFA0404602, 2018YFA0404603); 贵州省教育厅青年科技人才成长项目(黔教合 KY 字[2018]119); 贵州大学引进人才科研基金(贵大人基合字(2018)60号); 贵州大学培育项目(贵大培育[2020]76号)。

作者简介: 蔡宇佳(1996-), 女, 硕士研究生, 主要研究方向: 深度学习、医学图像处理; 陈旋(1997-), 女, 硕士研究生, 主要研究方向: 深度学习、医学图像处理; 覃芹(1997-), 女, 硕士研究生, 主要研究方向: 深度学习、医学图像处理; 陈丽绵(1996-), 女, 硕士研究生, 主要研究方向: 天文技术与方法、深度学习与图像处理; 张利(1987-), 男, 博士, 教授, 主要研究方向: 人工智能、大数据技术与应用。

通讯作者: 张利 Email: lizhang.science@gmail.com

收稿日期: 2022-09-06

基于标签分布来监督网络学习方法^[11-13]和通过处理小损失情况^[14-17]。同样,在隐式处理方式中也有 2 种经典方法:基于损失函数方法^[18-19]和基于图论的方法^[20-21]。本文的工作则拟对实例相关标签噪声的消除算法展开研究综述。

研究中仅考虑实例相关标签噪声对目标分类器的影响,并对其进行综述。针对实例相关标签噪声消除算法加以研究,通过显式和隐式的方法进行分析和总结,选取部分算法对其进行实验对比,并做出展望。

1 实例相关标签噪声的产生及影响

在现实世界的图像分类处理、目标识别等各类应用中广泛存在着基于实例相关的标签噪声。产生标签噪声的因素有很多,例如数据获取过程中通信设备本身带有错误标签^[22];成像本身质量的变化也会导致标注错误^[2];在射电天文领域,干涉仪等设备获取天文信息数据时,传输过程中会不可避免地带有嘈杂的便签;在一些标注中,信息不足也会导致标注错误等^[23]。

近年来,基于实例相关标签噪声问题在深度学习领域引发了广泛关注^[24-27]。Heskes^[24]、Lachenbruch^[25]证明了分类器受标签噪声的影响。Arpit 等人^[28]提出了深度模型的泛化用于处理标签噪声带来的影响。Zhang 等人^[29]能够拟合随机标签。Angluin 等人^[30]和 Wu 等人^[31]证明了标签噪声不仅降低了分类精度,且对算法模型产生误导性。此外,Frénay 等人^[32]、Shanab 等人^[33]则指出标签噪声影响特征选择及排序。

Frénay 等人^[32]提出了概率模型概念的方法去捕获图像,简称有向概率图,把标签噪声、地面实况标签和噪声类型之间的关系有效联系起来,将标签噪声分为随机标签噪声(Random Classification label Noise, RCN)、类相关标签噪声(Class-Conditional label Noise, CCN)和实例相关标签噪声(Instance-Dependent label Noise, IDN),如图 1 所示。其中, x, y, \hat{y} 分别为实例特征、真实标签、相应的标签噪声, e 为随机变量^[34-35]。

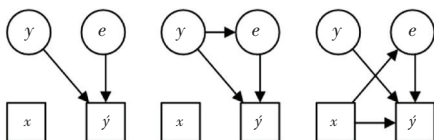


图 1 标签噪声类型图

Fig. 1 Label noise type diagram

2 相关算法

在实例标签噪声学习算法的国内外的各项成果中,皆是对不同的分类情况进行研究^[34-36]。在一些研究中,会根据实例数据去创建基础模型,或基于实例数据制定相关决策,本文主要围绕通过显式处理和隐式处理对实例相关噪声结构进行分类探讨,有关实例相关标签噪声处理算法如图 2 所示。

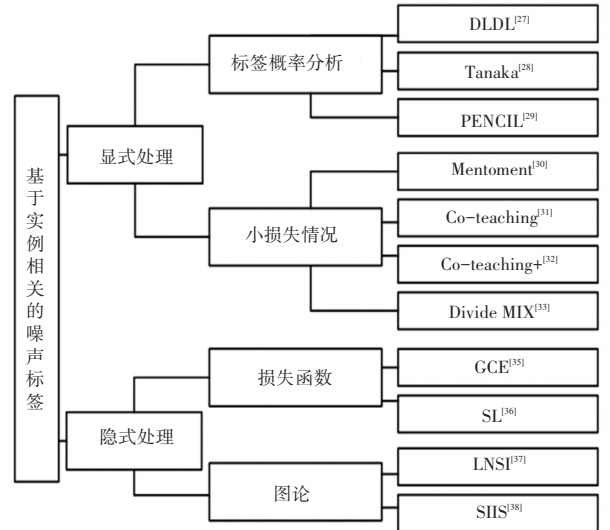


图 2 实例相关标签噪声处理算法

Fig. 2 Instance-related label noise processing algorithm

2.1 显式处理

在实例相关标签噪声中,显式处理通过对实例相关标签噪声进行建模,通常在学习的过程中,主模型对 IDN 进行清理,从而消除训练数据中错误的标签。表 1 展现了显式处理算法所存在的一些缺陷。在显式处理方式中,主要有 2 种方法对实例标签噪声算法进行综述,分别是:标签概率分布的方法和处理小损失情况的方法。对此可做阐释分述如下。

表 1 显式处理算法的缺陷总结

Tab. 1 A summary of explicit processing algorithms defects

去噪方法	存在缺陷
DLDL ^[11]	DLDL 的标签分布稳定,需要大量的先验信息,泛化性不强
Tanaka ^[12]	学习率较低时,准确率不高
PENCIL ^[13]	很难处理非平衡样本,噪声率较高时容易失效
Mentor Net ^[14]	训练过程中容易积累错误信息
Co-teaching ^[15]	训练时间长,容易产生泛化
Co-teaching+ ^[16]	噪声率极高时会导致小批量训练困难
Divide MIX ^[17]	不适用于噪声率极高的情况

2.1.1 标签概率分布的方法

文献[11]提出了一种深度标签分布学习方法(Deep Label Distribution Learning, DLDL),通过有效地利用了特征学习和分类器学习中的标签模糊性,防止在训练集很小的情况下产生过拟合,由于DLDL的标签分布较为稳定,需要大量的先验信息,导致无法更新的情况,且对不同的应用场景设计不同,泛化性不强。文献[12]提出网络参数和标签联合优化的框架,通过交替更新网络参数和标签来纠正训练中的标签,对网络噪声数据的训练改写损失值,通过修改损失函数,对网络进行更新。但在学习率较低时,得到的准确率不高。该损失函数可由如下公式进行描述:

$$L(\theta, Y | X) = L_c(\theta, Y | X) + \alpha l_p(\theta | X) + \beta l_e(\theta | X) \quad (1)$$

其中, L_c 为分类损失,是损失函数 L 的主要组成部分; l_e 为一个正则化项; l_p 为另一个正则化项; α 和 β 为超参数。

文献[13]提出一种端到端的框架(Probabilistic End-to-end Noise Correction for Learning with noisy labels, PENCIL),可以同时更新网络参数和数据标签,是DLDL方法的变体。与文献[12]相似,PENCIL同样不需要清洁数据集或有关噪声的先验信息,直接利用反向传播来概率地更新和校正图像标签,在训练过程中,PENCIL引入了损失函数(2),其中 l_e 沿用了KL-divergence的形式,并将其改为对称形式,取得不错的性能。该函数的数学公式见如下:

$$L = \frac{1}{C} l_c(f(x; \theta), Y^d) + \alpha l_o(\hat{Y}, Y^d) + \frac{\beta}{c} l_e(f(x; \theta)) \quad (2)$$

PENCIL的框架虽然能独立于任何模型进行训练,但对非平衡样本的处理较为困难,在实际情况中既存在噪声数据、又存在不均衡类别的情况也十分常见。

2.1.2 处理小损失情况的方法

文献[14]提出一种学习数据驱动课程的新方法(M-Net),利用小批量随机梯度下降法进行深度网络课程学习,通过学习另一个神经网络(Mentor Net)来监督基础深度网络(Student Net)训练的新技术,用来解决过拟合的问题,但在训练过程中容易积累错误信息。文献[15]提出了一种简单有效的学习范式(Co-teaching),通过同时训练2个深度神经

网络,并令其在每一个小批中相互教学,则能很好地解决积累的错误信息。Co-teaching 通过在小批量数据中过滤噪声样本传递给对等网络来更新参数,训练时间较长,可能会导致神经网络记忆泛化产生误差,使得网络之间达成收敛,就会得到较差的训练结果。针对 Co-teaching 的问题,文献[16]对其进行了改进,提出了一种新策略来训练网络(Co-teaching+),能够使得神经网络抵御标签噪声的鲁棒性,但该方法只适用于数据集噪声率较低的情况,当噪声率极高时会导致小批量训练变得困难。而文献[17]是在训练之前通过利用GMM模型将训练数据分为有标签数据和无标签数据,对其进行训练,使2个网络彼此互斥,即解决了 Co-teaching 的共识问题,在噪声率较高的情况下,训练效果较好,但相较而言还是不够理想。

在每个小批量处理下的相应网络训练结构(MentorNet (M-Net)^[14], Coteaching^[15], Coteaching⁺^[16]和 Divide MIX^[17])如图3所示。假设错误流来自训练实例的偏选,网络A和网络B的错误流分别用黑色实线箭头和虚线箭头表示。左面板中,M-Net只训练一个网络(A);中间偏左面板中,Coteaching同时训练2个网络(A和B);中间偏右面板中,Coteaching⁺训练2个网络(A和B),当2个网络的预测不一致(!=),更新这2个网络的参数;右面板中,Divide MIX同时训练2个网络(A和B),每个网络使用从另一个网络的数据集划分,以半监督的方式进行训练。

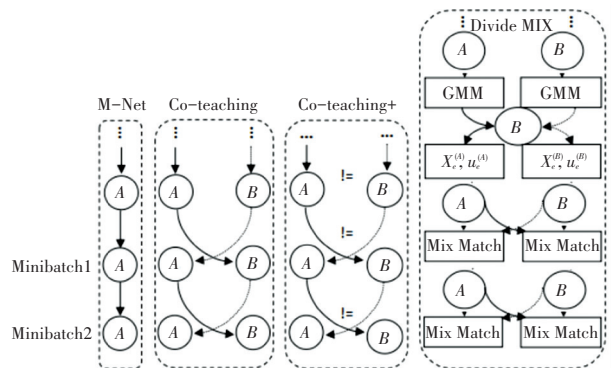


图3 Mini-batch 相关网络

Fig. 3 Mini-batch related network

2.2 隐式处理

隐式处理方法利用更加通用的算法来消除噪声的影响。该处理方法主要通过损失函数去处理标签噪声,使标签噪声训练过程中有较好的鲁棒性。由于IDN建模困难,许多研究者们对其进行假设,但在一些实际应用中,假设的方法会产生许多限制。

因此,一些研究者^[18-21,37]提出了隐式地处理实例相关的标签噪声的方法,该方法不需要对标签噪声的生成过程做出假设。

在实例相关标签噪声中,由于一些隐式的清洁处理方法会一起处理原有的干净标签,最终导致分类器性能变差,图像质量下降。为此,学者们提出从图论方向出发,通过探索矩阵所表示的实例相邻关系来设计标签噪声的修正处理方法。Wei 等人^[20]提出了一种通过边缘信息的标签噪声处理方法(LNSI),将观察到的标签矩阵分解为 2 部分。其中,一部分揭示真实的标签,一部分采用行稀疏矩阵对不正确的标签进行建模,这适用于二值分类,但需要很充分的先验知识。在隐式处理方式中,同样有 2 种方法对实例标签噪声算法进行综述,分别是基于损失函数的方法和基于图论的方法。对此可给出重点论述如下。

2.2.1 基于损失函数的方法

文献[18]提出了一个理论上的噪声鲁棒损失函数集(GCE),一种基于理论和易于使用的噪声鲁棒损失函数类,可以很容易地应用于任何现有的 DNN 体系结构和算法,同时在广泛的标签噪声场景中产生良好的性能。但在噪声率较高时,测试精度较低。

文献[19]提出了对称交叉熵(Symmetric cross entropy Learning, SL)的方法,该方法是反向交叉熵(Reverse Cross Entropy, RCE)和交叉熵(Cross Entropy, CE)结合体,不仅充分利用了两者的优点,还弥补了彼此的不足,利用噪声鲁棒对应的反向交叉熵(RCE)对称增强 CE,避免 CE 存在标签噪声的欠学习和过度拟合问题,但在实际应用中容易产生混淆。

2.2.2 基于图论的方法

实际数据集往往包含各种人为因素或测量误差引起的标签噪声,导致训练过程中可能产生错误标注,从而误导分类器的训练,严重降低分类性能。现有的隐式处理的方法中,通常是在经验风险最小化框架下通过各种替代损失函数去解决,但需要充分的先验信息。

文献[20]提出一种新的半监督学习范式、一种基于图的 SSL 算法,称为不充分和不正确监督下的半监督学习(SIIS),通过一个图来链接数据点,使标签信息可以沿着图的边缘从缺乏标签的实例传播到没有标签的实例。SIIS 采用图趋势滤波(GTF)和平滑特征基追踪(SEP)对初始的带噪标签进行过滤,

可以同时处理标签不足和标签不准确的问题。适用于图像、文本和音频等实例,当有限标记实例被错误标记过多时,训练时则稍显困难。

将实例特征视为边信息,并将标签噪声去除问题定义为矩阵恢复问题,方法称为通过侧信息处理标签噪声。具体地,将观察到的标签矩阵分解为 2 两部分之和。其中,第一部分揭示了真实的标签,可以通过对边信息进行低秩映射得到;第二部分采用行稀疏矩阵对不正确的标签进行建模。分析可知,该方法的优点体现在 3 个方面:

- (1) 该策略具有较强的恢复能力,并通过大量的边信息使理论工作得到了充分的论证。
- (2) 借助学习到的投影矩阵,可以直接处理多类情况。
- (3) 模型设计只需要非常弱的假设,使得 LNSI 适用于广泛的实际问题。

此外,本次研究从理论上推导了 LNSI 的泛化界,并证明了 LNSI 的期望分类误差是上界的。在多种数据集(包括 UCI 基准数据集和实际数据集)上的实验结果证实了 LNSI 在标签噪声处理方面的优越性。

文献[21]提出一类损失函数的若干充分条件,使多类别分类问题在该损失函数下的风险最小化,能够内在容忍标记噪声(LNSI),该方法通过研究深度网络中广泛使用的损失函数,证明其基于误差的平均绝对值的损失函数对标记噪声具有鲁棒性,适用于任何多类分类器学习且风险最小化。但当数据集过于复杂会直接导致分类器的性能变差。表 2 即展示了隐式学习的去噪方法中存在的缺陷。

表 2 隐式处理算法缺陷总结

Tab. 2 Summary of implicit processing algorithms defects

去噪方法	存在缺陷
GCE ^[18]	不适用于噪声率较高的情况
SL ^[19]	相似的类型容易产生混淆,分类精度不高
SIIS ^[20]	不适用于标注大量错误信息进行训练
LNSI ^[21]	不适用于庞大的数据量

综上所述,在实例相关标签噪声算法研究中,显式和隐式方法对其处理都有各自的优点。显式处理会对噪声本身进行建模,并在训练过程中使用建模后的信息来获得更好的性能;而隐式处理则通过图论或鲁棒性等方式对其进行研究。通过显式和隐式处理方法可以看出,基于实例相关标签噪声算法的研究中还存在许多不足,在未来,对于大噪声率的处理还需要做更进一步的探讨研究。

3 不同算法的对比实验

(1)数据集及参数设置。通过使用2种基准数据集对实例相关标签噪声的算法进行对比验证,文中使用CIFAR10和CIFAR100,见表3。这些数据集在文献中被广泛用于标签噪声的评估。为了更好地分析和比较算法的性能,对于所有实验,选取其动量为0.9,初始学习率为0.001,批处理大小为128,运行200 epoch。由于所有数据集都是干净的,所以本文通过手动标注错误标签,噪声率选为20%、50%和80%。

表3 数据集参数

Tab. 3 Data set parameters

数据集	训练	测试	类	图像大小
CIFAR10	50 000	10 000	10	32×32
CIFAR100	50 000	10 000	100	32×32

(2)性能测试。多是使用2种方式:测试精确

表4 基于实例相关标签噪声算法测试精度

Tab. 4 Test accuracy of the algorithm based on instance-related label noise

方法	CIFAR10			CIFAR100			%
	20	50	80	20	50	80	
PENCIL	91.32(±0.58)	84.87(±0.26)	21.64(±0.55)	64.45(±0.21)	33.81(±0.61)	24.78(±0.81)	
Divide MIX	94.53(±0.24)	92.04(±0.58)	91.71(±0.65)	76.65(±0.43)	72.48(±0.23)	50.49(±0.18)	
GCE	87.34(±0.16)	76.79(±0.62)	14.77(±0.24)	61.67(±0.29)	31.07(±0.27)	2.87(±0.92)	
SL	87.88(±0.23)	81.50(±0.53)	-	70.64(±0.18)	46.30(±0.33)	26.11(±0.68)	

4 结束语

本文主要讨论基于实例相关标签噪声处理算法问题,通过显式处理和隐式处理方法系统性分析可知,基于实例相关标签噪声问题更接近于实际应用,但也存在一些不足。现阶段,基于实例相关标签噪声的消除算法都不适用于数据集噪声率较大的情况,当噪声率极大时可能会导致无法工作。尽管标签噪声接近实际问题的处理,但并不能泛化为实际数据的各类情况。虽然能够处理错误的标签,但一些相邻的正确标签也会一并加以处理,若想得到较为干净的标签,过度清洁还会使得分类器性能下降。

标签噪声消除问题在大数据人工智能领域一直都是研究热点,然而标签噪声消除研究旨在将其应用于现实生活中,基于实例相关标签噪声消除问题的研究能够很好地链接实际应用,解决现实世界中存在的噪声问题。现阶段,对于实例相关标签噪声消除的研究仍在继续,且具有广泛性、实用性、可实现性等特点,未来有关基实例相关标签噪声的各类

度和标签精确度。其中,测试精确度是正确预测数和测试数据集数之比,而标签精确度则是干净标签数和所有选定标签数之比。本文只使用测试精确度对所有算法进行测试,选取5种算法进行实验,并对比各类算法的最后50个 epoch 的精确度,见表4。

表4中展示了CIFAR-10和CIFAR-100基线方法的实验对比结果。在噪声率为0.2、0.5和0.8的情况下,验证了每种方法的性能。其中,显式方法有PENCIL、Divide MIX,隐式方法包含GCE、SL。即如在2个数据集上看到的,所有方法都不适用于高噪声的情况,总地来看,显式方法中Divide MIX方法相较于显式方法中其他2个方法的测试精度较高,在数据集CIFAR-10噪声率为20%上测试精度达到94.53%。在隐式方法中,SL方法比GCE方法较好,但在CIFAR-10噪声率为80%的训练上失效,在CIFAR-100噪声率为80%上的测试精度较为理想。

研究也会越来越多。

参考文献

- [1] ZHANG L, MI L, ZHANG M, et al. Adaptive-scale wide-field reconstruction for radio synthesis imaging [J]. *Astronomy & Astrophysics*, 2020, 640: A80.
- [2] QUINLAN J R. Induction of decision trees [J]. *Machine Learning*, 1986, 1(1): 81-106.
- [3] ANGLUIN D, LAIRD P. Learning from noisy examples [J]. *Machine Learning*, 1988, 2(4): 343-370.
- [4] WU X. Knowledge acquisition from databases [M]. Greenwich, CT, USA: Ablex Publishing Corp, 1995.
- [5] WANG W, GUO B, SHEN Y, et al. Robust supervised topic models under label noise [J]. *Machine Learning*, 2021, 110(5): 907-931.
- [6] JANG R, KIM N, JANG M, et al. Assessment of the robustness of convolutional neural networks in labeling noise by using chest X-ray images from multiple centers [J]. *JMIR Medical Informatics*, 2020, 8(8), e18089.
- [7] HAUWERMEIREN V W, FILIPAN K, BOTTELDOOREN D, et al. Opportunistic monitoring of pavements for noise labeling and mitigation with machine learning [J]. *Transportation Research Part D: Transport and Environment*, 2021, 90: 102636.
- [8] BRODLEY C E, FRIEDL M A. Identifying mislabeled training

- data[J]. *Journal of Artificial Intelligence Research*, 1999, 11: 131–167.
- [9] ZHANG L, XU L, ZHANG M. Parameterized CLEAN deconvolution in radio synthesis imaging[J]. *Publications of the Astronomical Society of the Pacific*, 2020, 132(1010): 041001.
- [10] BRODLEY C E, FRIEDL M A. Identifying and eliminating mislabeled training instances [C]// *Proceedings of the National Conference on Artificial Intelligence*. Portland; AAAI, 1996: 799–805.
- [11] GAO Binbin, XING Chao, XIE Chenwei, et al. Deep label distribution learning with label ambiguity[J]. *IEEE Transactions on Image Processing*, 2017, 26(6): 2825–2838.
- [12] TANAKA D, IKAMI D, YAMASAKI T, et al. Joint optimization framework for learning with noisy labels [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA; IEEE 2018: 5552–5560.
- [13] YI Kun, WU Jianxin. Probabilistic end-to-end noise correction for learning with noisy labels [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA; IEEE, 2019: 7017–7025.
- [14] JIANG Lu, ZHOU Zhengyuan, LEUNG T, et al. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels [C]// *International Conference on Machine Learning*. Hanoi, Vietnam; ACM, 2018: 2304–2313.
- [15] HAN Bo, YAO Quanming, YU Xingrui, et al. Co-teaching: Robust training of deep neural networks with extremely noisy labels [C]// *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018*. Montréal, Canada; NIPS Foundation, 2018, 31: 8536–8546.
- [16] YU Xingrui, HAN Bo, YAO Jiangchao, et al. How does disagreement help generalization against label corruption? [C]// *International Conference on Machine Learning*. Shanxi, China; ACM, 2019: 7164–7173.
- [17] LI Junnan, SOCHER R, HOI S C. Dividemix: Learning with noisy labels as semi-supervised learning[J]. *arXiv preprint arXiv: 2002.07394*, 2020.
- [18] THULASIDASAN S, BHATTACHARYA T, BILMES J, et al. Combating label noise in deep learning using abstention[J]. *arXiv preprint arXiv: 1905.10964*, 2019.
- [19] WANG Yisen, MA Xingjun, CHEN Zaiyi, et al. Symmetric cross entropy for robust learning with noisy labels [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, South Korea; IEEE, 2019: 322–330.
- [20] WEI Yang, GONG Chen, CHEN Shuo, et al. Harnessing side information for classification under label noise [J]. *IEEE transactions on neural networks and learning systems*, 2019, 31(9): 3178–3192.
- [21] GONG Chen, ZHANG Hengming, YANG Jian, et al. Learning with inadequate and incorrect supervision [C]// *2017 IEEE International Conference on Data Mining (ICDM)*. New York, USA; IEEE, 2017: 889–894.
- [22] SINGLA M, GHOSH D, SHUKLA K K. Improved Sparsity of Support Vector Machine with Robustness Towards Label Noise Based on Rescaled α -Hinge Loss with Non-smooth Regularizer[J]. *Neural Processing Letters*, 2020, 52(3): 2211–2239.
- [23] NETTLETON D F, ORRIOLS-PUIG A, FORNELLS A. A study of the effect of different types of noise on the precision of supervised learning techniques[J]. *Artificial Intelligence Review*, 2010, 33(4): 275–306.
- [24] HESKES T. The use of being stubborn and introspective [C]// *Prerational Intelligence: Adaptive Behavior and Intelligent Systems Without Symbols and Logic, Volume 1, Volume 2 Prerational Intelligence: Interdisciplinary Perspectives on the Behavior of Natural and Artificial Systems*. Dordrecht; Springer, 2000, 3: 1184–1200.
- [25] LACHENBRUCH P A. Note on initial misclassification effects on the quadratic discriminant function[J]. *Technometrics*, 1979, 21(1): 129–132.
- [26] DIETTERICH T G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization[J]. *Machine learning*, 2000, 40(2): 139–157.
- [27] JOACHIMS T. SvmLight: Support vector machine [J]. *SVM-Light Support Vector Machine*, 1999, 19(4): 25.
- [28] ARPIT D, JASTRZEBSKI S, BALLAS N, et al. A closer look at memorization in deep networks [C]// *International Conference on Machine Learning*. Sydney, NSW, Australia; dblp, 2017: 233–242.
- [29] ZHANG Chiyuan, BENGIO S, HARDT M, et al. Understanding deep learning (still) requires rethinking generalization [J]. *Communications of the ACM*, 2021, 64(3): 107–115.
- [30] ANGLUIN D, LAIRD P. Learning from noisy examples [J]. *Machine Learning*, 1988, 2(4): 343–370.
- [31] WU Songhua, GONG Mingming, HAN Bo, et al. Fair classification with instance-dependent label noise [C]// *Conference on Causal Learning and Reasoning*. Eureka, CA, USA; PMLR, 2022: 927–943.
- [32] FRÉNAY B, DOQUIRE G, VERLEYSSEN M. Feature selection with imprecise labels: Estimating mutual information in the presence of label noise [J]. *Computational Statistics & Data Analysis*, 2014, 71: 832–848.
- [33] SHANAB A A, KHOSHGOFTAAR T M, WALD R. Robustness of threshold-based feature rankers with data sampling on noisy and imbalanced data [C]// *Proceedings of Twenty-Fifth International Florida Artificial Intelligence Research Society Conference*. Florida, USA; dblp, 2012: 92–97.
- [34] 王晓莉, 薛丽. 标签噪声学习算法综述 [J]. *计算机系统应用*, 2020, 30(01): 10–18.
- [35] 官辰, 张闯, 王启舟. 标签噪声鲁棒学习算法研究综述 [J]. *航空兵器*, 2020, 27(03): 20–26.
- [36] FENG Wei, QUAN Yinghui, DAUPHIN G. Label noise cleaning with an adaptive ensemble method based on noise detection metric [J]. *Sensors*, 2020, 20(23): 6718.
- [37] CHENG Jiacheng, LIU Tongliang, RAMAMOCHANARAO K, et al. Learning with bounded instance and label-dependent label noise [C]// *International Conference on Machine Learning*. Florida; IEEE, 2020: 1789–1799.
- [38] FRÉNAY B, VERLEYSSEN M. Classification in the presence of label noise: A survey [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, 25(5): 845–869.