

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2024)12-3684-15

论文引用格式: Dong Y F, Li J W, Wang Z and Jia W Y. 2024. Co-history: learning with noisy labels by co-teaching with history losses. Journal of Image and Graphics, 29(12):3684-3698(董永峰, 李佳伟, 王振, 贾文玉. 2024. Co-history: 协同学习中考虑历史信息的标签噪声鲁棒学习方法. 中国图象图形学报, 29(12):3684-3698)[DOI: 10. 11834/jig. 230541]

Co-history: 协同学习中考虑历史信息的标签噪声鲁棒学习方法

董永峰, 李佳伟, 王振*, 贾文玉

1. 河北工业大学人工智能与数据科学学院, 天津 300401; 2. 河北省大数据计算重点实验室(河北工业大学), 天津 300401;
3. 河北省数据驱动工业智能工程研究中心(河北工业大学), 天津 300401

摘要: 目的 深度神经网络在计算机视觉分类任务上表现出优秀的性能, 然而, 在标签噪声环境下, 深度学习模型面临着严峻的考验。基于协同学习(co-teaching)的学习算法能够有效缓解神经网络对噪声标签数据的学习问题, 但仍然存在许多不足之处。为此, 提出了一种协同学习中考虑历史信息的标签噪声鲁棒学习方法(Co-history)。方法 首先, 针对在噪声标签环境下使用交叉熵损失函数(cross entropy, CE)存在的过拟合问题, 通过分析样本损失的历史规律, 提出了修正损失函数, 在模型训练时减弱CE损失带来的过拟合带来的影响。其次, 针对co-teaching算法中两个网络存在过早收敛的问题, 提出差异损失函数, 在训练过程中保持两个网络的差异性。最后, 遵循小损失选择策略, 通过结合样本历史损失, 提出了新的样本选择方法, 可以更加精准地选择干净样本。结果 在4个模拟噪声数据集F-MNIST(Fashion-mixed National Institute of Standards and Technology)、SVHN(street view house number)、CIFAR-10(Canadian Institute for Advanced Research-10)和CIFAR-100和一个真实数据集Clothing1M上进行对比实验。其中, 在F-MNIST、SVHN、CIFAR-10、CIFAR-100, 对称噪声(symmetrical)40%噪声率下, 对比co-teaching算法, 本文方法分别提高了3.52%、4.77%、6.16%和6.96%; 在真实数据集Clothing1M下, 对比co-teaching算法, 本文方法的最佳准确率和最后准确率分别提高了0.94%和1.2%。结论 本文提出的协同学习下考虑历史损失的带噪声标签鲁棒分类算法, 经过大量实验论证, 可以有效降低噪声标签带来的影响, 提高模型分类准确率。

关键词: 深度神经网络(DNN); 分类; 噪声标签; 协同学习; 历史损失

Co-history: learning with noisy labels by co-teaching with history losses

Dong Yongfeng, Li Jiawei, Wang Zhen*, Jia Wenyu

1. School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China;
2. Hebei Province Key Laboratory of Big Data Computing (Hebei University of Technology), Tianjin 300401, China;
3. Hebei Engineering Research Center of Data-Driven Industrial Intelligent (Hebei University of Technology), Tianjin 300401, China

Abstract: Objective Deep neural networks (DNNs) have been successfully applied in many fields, especially in computer vision, which cannot be achieved without large-scale labeled datasets. However, collecting large-scale datasets with accurate labels is difficult in practice, especially in some professional fields. The labeling of these datasets requires the

收稿日期: 2023-08-14; 修回日期: 2024-03-08; 预印本日期: 2024-03-15

* 通信作者: 王振 wangzhen@hebut.edu.cn

基金项目: 国家自然科学基金项目(62306103, 62376194); 河北省高等学校自然科学基金项目(QN2023262)

Supported by: National Natural Science Foundation of China (62306103, 62376194); Higher Education Science and Technology Research Project of Hebei Province (QN2023262)

involvement of relevant experts, thus increasing manpower and financial resources. To cut costs, researchers have started using datasets built by crowdsourcing annotations, search engine queries, and web crawling, among others. However, these datasets inevitably contain noisy labels that seriously affect the generalization of DNNs because DNNs memorize these noise labels during training. Learning algorithms based on co-teaching methods, including Co-teaching+, JoCoR, and CoDis, can effectively alleviate the learning problem of neural networks on noisy label data. Scholars have put forward different opinions regarding the use of two networks to solve noisy labels. However, in a noisy label environment, the deep learning model based on CE loss is very sensitive to the noisy label, thus making the model easily fit the noisy label sample and unable to learn the real pattern of the data. With the progress of training, Co-teaching causes the parameters of the two networks to gradually become consistent and prematurely converge to the same network, thus stopping the learning process. As the iteration progresses, the network inevitably remembers some of the noisy label samples and thus failing to distinguish the noisy from the clean samples accurately based on the cross entropy (CE) loss value. In this case, relying solely on CE loss as a small loss selection strategy is not reliable. To solve these problems, this paper proposes learning with noisy labels by co-teaching with history losses (Co-history) that considers historical information in collaborative learning.

Method First, to solve the overfitting problem of cross entropy loss (CE) in a noisy label environment, a correction loss is proposed by analyzing the history of sample loss. The revised loss function adjusts the weight of the CE loss in the current iteration in order for the CE loss of the sample to remain stable in the historical iteration as a whole, hence conforming to the law that the classifier should be maintained after separating the noisy from the clean samples so as to reduce the influence of overfitting caused by CE loss. Second, the difference loss is proposed to address the problem of premature convergence of two networks in the co-teaching algorithm. Inspired by contrast loss, the difference loss makes the two networks maintain a certain distance from the feature representation of the same sample so as to maintain the difference between these networks in the training process and to avoid their degradation into a single network. Given the differences in the network parameters, various decision boundaries are generated, and different types of errors are filtered. Therefore, maintaining such difference can benefit collaborative learning. Finally, due to the existence of overfitting, those samples with noisy labels tend to have larger loss fluctuations than those samples with clean labels. By combining the historical loss information of these samples and following the small loss selection strategy, a new sample selection method is proposed to select clean samples accurately. Specifically, those samples with low classification losses and low fluctuations in historical losses are selected as clean samples for training. **Result** Several experiments are conducted to demonstrate the effectiveness of the Co-history algorithm, including comparison experiments on four standard datasets (F-MNIST, SVHN, CIFAR-10, and CIFAR-100) and one real dataset (Clothing1M). Four categories of artificially simulated noise are added to the standard dataset, including symmetric, asymmetric, pairflip, and tri-diagonal noise types, with 20% and 40% noise rates for each noise type. In the real dataset, the labels are generated by the text around the image, which contains the noise label itself, thus generating no additional label noise. At the symmetric noise type with 20% noise rate, the co-history algorithm demonstrates 2.05%, 2.19%, 3.06%, and 2.58% improvements over the co-teaching algorithm in the F-MNIST, SVHN, CIFAR-10, and CIFAR-100 datasets, respectively. With 40% noise rate, the corresponding improvements are 3.52%, 4.77%, 6.16%, and 6.96%. In the real Clothing1M dataset, the best and lowest accuracies of co-history have improved by 0.94% and 1.2%, respectively, compared with the co-teaching algorithm. The effectiveness of the proposed loss is proven by ablation experiments. **Conclusion** A correction loss is proposed in this paper to address the overfitting problem of CE loss training and the historical law of sample loss, and a difference loss function is introduced to solve the premature convergence of two networks in Co-teaching. In view of the traditional small-loss sample selection strategy, the historical law of sample loss is fully considered in this paper, and a highly accurate sample selection strategy is developed. The proposed Co-history algorithm demonstrates its superiority over the existing co-teaching strategies in a large number of experiments. This algorithm also shows strong robustness in datasets with noisy labels and is particularly suitable for noisy label scenarios. The various improvements in this algorithm are also clearly demonstrated in ablation experiments. Given that this algorithm needs to analyze the historical loss information of each sample, the historical loss value of each sample should be saved. Increasing the number of training samples would occupy more memory space, thus increasing computing and storage costs. In addition, with a large number of sample categories, the performance of the proposed algorithm

becomes suboptimal under some noisy environments (e. g., asymmetric noise type with 40% noise rate and the CIFAR-100 dataset with 20% noise rate). Future work will focus on the development of high-performance solutions under the premise of guaranteed accuracy and excellent robust classification algorithms for learning with noisy labels.

Key words: deep neural network(DNN); classification; noisy labels; co-teaching; historical loss

0 引言

深度神经网络(deep neural network, DNN)已经在许多领域取得了显著效果,尤其是在计算机视觉领域(Diaz, 2009; Krizhevsky 等, 2017; 张珂 等, 2021; 李雅馨 等, 2023),这离不开大规模的有标注数据集。然而,实际中收集大规模带有精准标签的数据集是非常困难的,尤其是某些专业领域的数据,需要相关专家进行标注,其会耗费大量的人力和财力。因此,为降低成本,研究人员转而使用众包标注(Blum 等, 2003)、搜索引擎查询(Yan 等, 2014)以及网络爬取(Mahajan 等, 2018)等方法构建的数据集。而这样获得的数据集会不可避免地引入标签噪声,严重影响DNN的泛化性,因为DNN会在训练过程中记忆这些噪声标签。

目前标签噪声问题已经成为研究领域的热点之一(Song 等, 2023),如何减轻噪声标签对网络的影响是解决问题的关键。现有技术主要包括设计鲁棒损失函数、标签矫正和小损失样本选择等。小损失样本选择的合理性在于DNN的先学习模式(Arpit 等, 2017),即使存在噪声标签,DNN也会在初始阶段学习到干净、简单的样本。因此,DNN有能力在训练开始时利用损失值过滤掉噪声样本。协同学习(co-teaching)(Han 等, 2018)算法在噪声标签学习任务上取得了不错的结果。具体来说,co-teaching采用交叉熵损失函数(cross entropy, CE)同时训练两个网络,基于小损失选择策略,每个网络都将其小损失样本视为干净样本,并将这些样本传递给对等网络进行学习。两个网络的初始随机参数不同,具有不同的学习能力,可以产生不同的决策边界,过滤不同类型的错误。两个网络交替学习,会自适应地纠正对等网络的训练错误。

然而,在噪声标签环境下,基于CE损失函数的深度学习模型对噪声标签十分敏感,使得模型极易对噪声标签样本拟合,无法学习数据的真实模式。针对这一问题,本文通过分析样本历史分类损失规

律,提出了修正损失函数,减轻过拟合给模型带来的影响。其次,随着训练的进行,co-teaching学习模式导致两个网络的参数逐渐变得一致,过早收敛为相同的网络,导致学习停止。针对这一问题,本文从样本的特征表示出发,借鉴对比学习的思想,引入新的差异损失函数使网络保持最佳的“分歧”,从而增强协同学习中两个网络过滤不同错误的能力,提高模型的泛化性。最后,由于DNN具有强大的学习能力,网络模型会不可避免地记忆部分噪声标签样本,导致仅仅根据CE的损失值无法正确区分噪声和干净样本。因此,单纯依靠CE损失使用小损失选择策略是不可靠的。针对这一问题,本文通过分析干净样本所具备的历史损失规律,提出了一种更为准确的样本选择策略,从而使干净样本的选择更为精准。

综上,本文的贡献主要分为以下3点:1)针对仅使用CE损失会导致过拟合的问题,分析样本损失的历史规律,提出了修正损失函数,减轻过拟合给模型带来的影响;2)针对co-teaching中两个网络过早收敛为相同网络的问题,提出了差异损失函数,增强网络的学习性能;3)提出一种基于样本损失历史规律的样本选择策略,从而提高选择样本的纯净率。

1 相关工作

本节简单总结了现有关于噪声标签分类方法,主要分为3大类:鲁棒损失函数、标签矫正和小损失样本选择。

1.1 鲁棒损失函数

CE损失已经被证明在噪声标签学习中存在过拟合现象(Ghosh 等, 2017)。因此,许多学者开展了一系列针对标签噪声的鲁棒损失函数的研究。Ghosh 等人(2015)设计了平均绝对误差(mean absolute error, MAE),并在理论上证明了对称损失函数对于标签噪声具有的鲁棒性。Wang 等人(2019)提出的逆交叉熵(reverse cross entropy, RCE),在多分类标签任务下有着不错的表现。但是,在很多情况下,这种抗噪声损失函数存在欠拟合的现象。因此,

Zhang 和 Sabuncu(2018)将 MAE 和 CE 进行结合,设计了广义交叉熵(generalized cross entropy, GCE)。Ma 等人(2020)使用归一化技术,提出更为鲁棒的损失函数。但是这种损失函数依然存在欠拟合的问题,之后又设计了主动被动损失(active passive loss, APL),APL 结合归一化损失函数和鲁棒损失函数用于提高收敛性。之后, Kim 等人(2021)为进一步解决欠拟合问题,通过分析损失函数梯度,提出了 NL+(negative learning+)和 PL+(positive learning+)损失。

1.2 标签矫正

标签矫正方法通过对噪声标签进行矫正,使标签噪声样本也可以参与训练,充分利用所有的样本信息,提高模型性能。标签平滑(label smooth)是标签矫正中常用到的方法,其相当于一种正则化方法(Lukasik 等, 2020)。Yi 和 Wu(2019)通过先使用原始标签在网络中学习,当网络测试准确率有一定提高时,通过网络预测结果不断更新纠正噪声标签,进而提升模型的预测能力。DivideMix(Li 等, 2020)对噪声样本采用联合预测(co-guess)的方式,使用多个网络预测均值来矫正噪声标签。Huang 等人(2023)使用样本在高斯混合模型(Gaussian mixture model, GMM)中的后验概率,将给定样本的标签和预测结果进行重新加权得到修正标签,指导模型训练。Meta pseudo labels(Pham 等, 2021)通过使用蒸馏方法,同时更新教师和学生网络来优化伪标签的生成。Wei 等人(2022)在其方法中通过有选择地矫正噪声样本,使其具有更高的矫正精度。

但是伪标签矫正的准确程度依赖于模型的性能,当模型的分类准确率较低时,无法将噪声标签纠正为正确的标签,甚至可能会造成错误积累,给模型带来不良影响。

1.3 小损失样本选择

得益于深度网络的先学习模式(Arpit 等, 2017),尽管存在噪声标签,模型也会在初始阶段学习到干净、简单的样本。因此, DNN 在训练初期可以根据损失的大小选择出干净的样本。Ding 等人(2018)提出了基于预测可能性的干净样本的选择方法。根据样本损失大小将数据集划分为干净子集与噪声子集,随后采取不同方式进行学习。Jiang 等人(2018)通过预训练教师网络选择小损失样本作为干净样本指导学生网络学习。Han 等人(2018)提出了协同学习(co-teaching),通过两个网络交替学习,避

免错误积累。然而交替学习会使两个网络参数过早收敛,从而导致学习停止。Yu 等人(2019)为解决这一问题,提出了 Co-teaching+,其认为仅在两个网络之间存在预测分歧的样本中选择小损失样本进行更新。但是在某些噪声类型的情况下,Co-teaching+ 可能无法识别出足够多的干净样本,导致性能下降。JoCoR(Wei 等, 2020)则建议使用两个网络之间具有预测一致性的样本联合训练两个网络。与其相反,CoDis(Xia 等, 2023)认为选择预测方差大的样本加入模型训练,可以增加两个网络的差异性,提高模型鲁棒性。Liu 等人(2022)也在其工作中详细分析了 co-teaching 中两个网络是否应该保持分歧,其通过实验论证,两个网络存在适当的差异,有益于模型鲁棒性的提升。此外, CNLCU (combats noisy labels by concerning uncertainty)(Xia 等, 2022)在其方法中,通过区间估计的方法确定置信下界,筛选干净样本,其本质仍采用了小损失样本选择的策略。

2 方法

本文方法主要包含 3 部分:通过分析样本的历史损失信息提出的修正损失函数、改善 co-teaching 中两网络过早收敛问题的差异损失函数,以及结合样本历史损失规律的样本选择策略。

2.1 问题描述

考虑一个 C 类分类任务,存在噪声标签的数据集 $D = \{(x_i, y_i)\}_{i=1}^n$,其中, n 为样本容量, x_i 为第 i 个样本, $y_i = [C]$ 为含有噪声的标签。对于 y_i ,其不可观测的真实标签用 \tilde{y}_i 来表示。目标为学习一个鲁棒的分类器,该分类器可以通过在含有噪声标签的数据集上训练,仍然可以得到正确的预测结果。

2.2 基于历史信息的鲁棒学习

本文通过分析样本损失的历史规律,提出修正损失(correction loss)和差异损失(difference loss),从而提高模型的鲁棒性。

2.2.1 修正损失函数(correction loss)

CE 损失已经被证明在噪声标签学习中存在过拟合现象(Ghosh 等, 2017)。由于小损失样本选择策略并不能完全正确地分离出干净样本,随着迭代进行,CE 损失会使网络模型不可避免地拟合一些噪声标签,影响模型泛化性。为了解决这一问题,本文

设计修正损失函数,其通过结合样本损失的历史规律,优化每个样本CE损失的权重,减弱拟合噪声的情况。具体来说,当模型分离出干净样本和噪声样本后,每个样本的分类损失数值应当维持在其原有的数值,不应随着学习的进行而有较大的波动。由于过拟合问题,原本具有较大损失的噪声样本,损失可能逐渐减小,造成分类器无法正确区分干净和噪声样本。

因此,本文应用样本历史损失对分类损失函数进行修正,使其在分离出干净和噪声样本之后趋于平稳。具体地,定义每个样本的历史损失信息为一组长度为 s 的滑动窗口序列,记为 $L_s = \{CE_{t-1}, CE_{t-2}, \dots, CE_{t-s}\}$,用于记录历史CE值, t 表示当前训练迭代次数。定义 L_{history} 为

$$L_{\text{history}} = \frac{1}{s} \sum_{i=t-s}^{t-1} CE_i \quad (1)$$

式中, t 表示当前训练迭代次数, CE 为交叉熵损失,具体为

$$CE = - \sum_{j=1}^c y_i^j \log p^j(x_i) \quad (2)$$

式中, p 为模型输出的样本预测概率。使用 $|CE - L_{\text{history}}|$ 来衡量模型对样本训练的稳定性,为了使样本获得稳定的训练过程,避免模型错误地拟合噪声样本,本文使用 $L_{\text{correction}}$ 作为修正分类损失函数对模型进行训练。最小化 $L_{\text{correction}}$,其意义为调整本轮迭代中CE损失的权重,使样本的CE损失在历史迭代中整体保持平稳,这符合分类器在分离噪声和干净样本后所应保持的规律。具体为

$$L_{\text{correction}} = |CE - L_{\text{history}}| \quad (3)$$

2.2.2 差异损失函数(difference loss)

co-teaching同时训练两个网络,使两个网络交替学习。两个网络的初始随机参数不同,可以产生不同的决策边界,具有不同的学习能力,过滤不同类型的错误。因此,两个网络交替学习,会自适应地纠正对等网络的训练错误,提升模型鲁棒性。然而,在相互学习的过程中,两个网络会存在过早收敛的问题,导致学习提前停止。因此设计差异损失函数,用于维持两个网络之间的差异。

对比学习是一种无监督学习的方法,通过对比相似和不相似的样本来学习样本的特征表示。对比损失为

$$L_{\text{contrast}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp \frac{\langle z_i, z'_i \rangle}{K}}{\sum_{j=1}^N M_{j \neq i} \exp \frac{\langle z_i, z_j \rangle}{K}} \quad (4)$$

式中, z_i 为网络对第 i 个样本的特征表示, z'_i 为其经过随机数据增强后的特征表示, $\langle *, * \rangle$ 为余弦相似度, K 为温度系数, $M_{j \neq i}$ 为取值为1的指示函数,当 $j = i$ 时为0,当 $j \neq i$ 时为1。该损失函数的目的在于使第 i 个样本和其随机增强的样本(正样本)之间的距离尽可能近,而与其他样本(负样本)之间的距离尽可能远,使模型可以学习到具有判别性的特征表达。受对比损失启发,本文将这种思想应用到co-teaching中,目的为让两个网络对同样样本的特征表示保持距离,以此来控制两个网络的差异。为此,本文设计了一种差异损失函数 L_{unco} ,具体为

$$L_{\text{unco}} = \frac{1}{N} \sum_{i=1}^N \log \frac{\exp \frac{\langle z_{1i}, z_{2i} \rangle}{K}}{\sum_{j=1}^N M_{j \neq i} \exp \frac{\langle z_{1i}, z_{1j} \rangle + \langle z_{2i}, z_{2j} \rangle}{K}} \quad (5)$$

式中, z_{1i} 和 z_{2i} 分别为网络1和网络2对第 i 个样本的特征表示。在模型训练中最小化式(5),其意义为最大化两个模型对同样样本的特征表示距离,控制两个网络的发散程度,防止过早收敛。使用超参数 γ 来控制差异损失函数的影响程度,避免网络无意义的发散。

最终的损失函数为

$$L = L_{\text{correction}} + \gamma L_{\text{unco}} \quad (6)$$

由于差异损失使两个网络具有明显不同的分类边界,在模型预测时,使用两个网络的预测平均作为最终预测结果,可以充分考虑两个网络的分类边界,提高整体的预测准确率。

2.3 样本选择策略

在早期训练过程中,DNN倾向于先学习干净的样本,只选择CE损失小的样本,一定程度上可以避免噪声样本对网络的干扰。但是,随着迭代的进行,模型最终可能会偏向学习噪声样本,进而产生过拟合现象。因此,噪声样本相比于干净样本,往往会产生较大的损失波动。本文认为仅仅考虑CE损失小的样本作为干净样本是不够精准的,尽管使用了修正分类损失函数进行训练,但是噪声样本损失的历史规律对比干净样本往往有着更大的波动。因此应用 $L_{\text{selection}}$ 作为样本选择函数,旨在选择分类损失小且

历史损失波动小的样本作为干净样本训练, 解决传统小损失选择策略在训练后期无法正确区分噪声和干净样本的问题。

$$L_{\text{selection}} = \alpha CE + \beta L_{\text{history}} \quad (7)$$

式中, α 和 β 为超参数, 用于控制 CE 和 L_{history} 的影响占比。算法伪码见算法 1, 整体流程图如图 1 所示。

算法 1 Co-history

输入: 两个网络参数 w_1 和 w_2 , 学习率 τ , 噪声率 η , 预热轮数 T_{warmup} , 训练总轮数 T_{max} , 每轮迭代次数 t_{max} 。

输出: 网络参数 w_1 和 w_2 。

1) FOR $T = 1 \rightarrow T_{\text{max}}$ DO;

- 2) 随机化数据集 \tilde{D} ;
- 3) IF $T < T_{\text{warmup}}$ THEN;
- 4) 通过标准交叉熵(CE)训练模型;
- 5) ELSE;
- 6) FOR $T = 1 \rightarrow T_{\text{max}}$ DO;
- 7) 获取小批量样本 \bar{D} ;
- 8) $\bar{D}_1 = \arg \min_{D': |D'| \geq (1-\eta)|\bar{D}|} L_{\text{selection}}(w_1, D')$;
- 9) $\bar{D}_2 = \arg \min_{D': |D'| \geq (1-\eta)|\bar{D}|} L_{\text{selection}}(w_2, D')$;
- 10) 更新网络 $w_1 = w_1 - \tau \nabla L(w_1, \bar{D}_2)$;
- 11) 更新网络 $w_2 = w_2 - \tau \nabla L(w_2, \bar{D}_1)$;
- 12) END FOR;
- 13) END FOR。

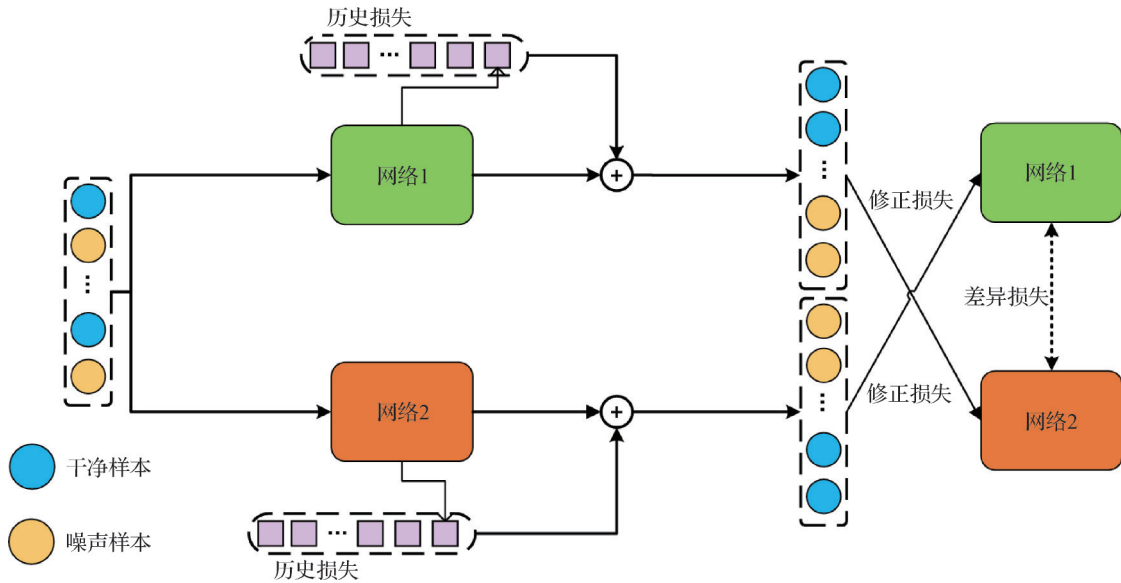


图 1 算法整体过程

Fig. 1 Overall process of the algorithm

3 实验与分析

3.1 实验细节

3.1.1 实验数据集

本文在 4 个标准数据集 F-MNIST(Fashion-mixed National Institute of Standards and Technology)(Xiao 等, 2017)、SVHN(street view house number)(Netzer 等, 2011)、CIFAR-10(Canadian Institute for Advanced Research-10)、CIFAR-100(Krizhevsky, 2009) 和一个真实数据集 Clothing1M(Xiao 等, 2015) 上验证方法的有效性。F-MNIST 由 60 000 个训练样本和 10 000

个测试样本组成, 每个样本都是一个 28×28 像素的灰度图像, 包含 10 个类别。SVHN 数据集来源于谷歌街景门牌号码, 由 73 257 个训练样本和 26 032 个测试样本组成, 包含 10 个类别, 每个样本为 32×32 像素的 RGB 图像; CIFAR-10 由 60 000 幅 32×32 像素的 RGB 图像组成, 包含 10 个类别, CIFAR-100 和 CIFAR-10 相似, 包含 20 个超类, 每个超类又包含 5 个子类, 总共 100 个类, 每个样本为 32×32 像素的 RGB 图像。Clothing1M 由 100 万幅从在线购物网站上收集的训练图像组成, 包含 18 个类别, 其标签是用周围的文字生成的。

3.1.2 噪声类型

Clothing1M为真实数据集,本身含有噪声标签。由于F-MNIST、SVHN、CIFAR-10和CIFAR-100本身为干净数据集,所以需要设置标签转换矩阵(Patrini等,2017),通过人为破坏数据集来模拟噪声。实验遵循Zhang等人(2021)的设置,进行多种类型的模拟噪声标签实验,包括:对称噪声、成对反转噪声、三角对称噪声和非对称噪声。每种类型的噪声分别设置20%和40%的噪声率,其目的是保证干净样本的数量要多于噪声样本(Ma等,2020),否则会无法分辨样本真正所属的类别。每种噪声类型的构造如下:

1)对称噪声(symmetric)。将每个类中干净的标签统一翻转为其他错误类别的标签。

2)非对称噪声(asymmetric)。它考虑了翻转过程中的视觉相似性,更接近于真实世界的噪声,例如:会将猫和狗的标签反转,飞机和鸟的标签反转。非对称噪声类型是一种不平衡的噪声。

3)成对反转噪声(pairflip)。将每个类中的干净标签翻转到相邻类别中。

4)三角对称噪声(tridiagonal)。通过在相反方向上的类进行两次连续的成对反转实现。

4种标签转换矩阵如图2所示,以6类、40%噪声率为例。

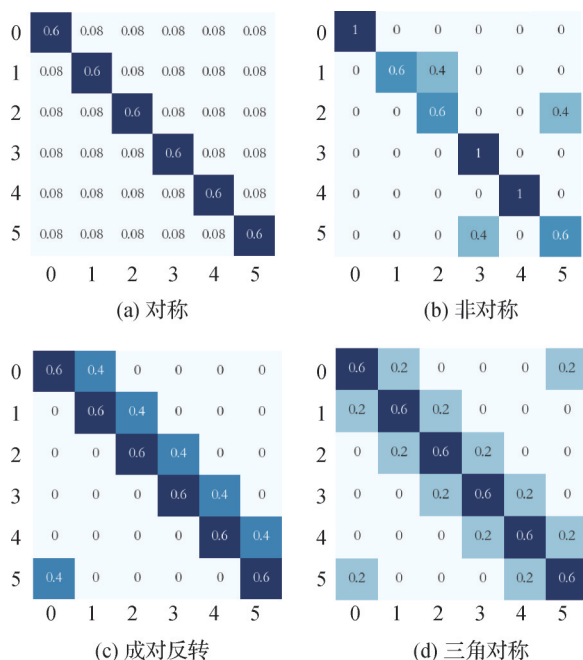


图2 标签转换矩阵

Fig. 2 Label transition matrix ((a) symmetric; (b) asymmetric; (c) pairflip; (d) tridiagonal)

3.1.3 网络结构和优化器

本文在F-MNIST、SVHN、CIFAR-10数据集中使用co-teaching中的9层CNN网络进行训练。在CIFAR-100数据集中使用7层CNN网络进行训练。在Clothing1M数据集中,使用在ImageNet数据集上预训练18层ResNet网络进行训练。

本文在F-MNIST、SVHN、CIFAR-10和CIFAR-100数据集的实验中,使用Adam优化器(动量=0.9),初始学习率为 1×10^{-3} ,批量(mini-batch)大小设置为128。总共训练200个周期,从第80到第200周期,学习率线性衰减为零。对于Clothing1M数据集,首先对数据集进行预处理,将图像的尺寸调整为 256×256 像素,裁剪中心 224×224 像素作为输入,并执行归一化,遵循Wei等人(2020)的设置,只使用含有噪声标签的训练集训练。对于真实数据集Clothing1M同样使用Adam优化器,批量(mini-batch)大小设置为64。在训练过程中,总共训练15轮,学习率分别设置为 8×10^{-4} 、 5×10^{-4} 和 5×10^{-5} ,每5轮改变一次学习率。

3.1.4 对比环境

本文算法以及所有对比实验算法,均通过PyTorch实现默认参数,并在NVIDIA RXT3090 GPU上进行实验。所有的对比算法,均严格按照原始论文或者公开代码进行复现,为保证公平性,所有的实验均采用相同的实验设置且在相同的实验环境中进行。

3.1.5 评价指标和样本选择设置

本文采用模型在测试集上的准确率作为评价指标。

计算实验最后10轮的平均值和标准差(%),分别反映模型的分类准确率以及模型最终的收敛情况。所有实验重复3次,记录3次实验分类准确率平均值和标准差。同时,为了更直观地分析实验结果,使用3次实验的平均值绘制了测试准确率曲线,标准差使用阴影突出显示。由于Clothing1M数据集计算量大,所以只进行一次实验。记录训练周期中最佳测试准确率(best)和最终测试准确率(last)。

假设数据集中的噪声率 τ 是已知的。本文在标准数据集的每个小批量(mini-batch)中选择前 $1-\tau$ 的小损失样本作为干净样本。对于真实数据集Clothing1M。如果对 τ 的值不清楚,则可以使用验证集对 τ 进行推断(Liu和Tao,2016;Yu等,2018)。

3.1.6 对比算法

本文算法将与以下先进的算法进行对比:

- 1) Standard: 采用标准交叉熵进行训练;
- 2) APL(Ma等, 2020): 构造鲁棒损失函数训练模型;
- 3) co-teaching(Han等, 2018): 同时训练两个网络, 两个网络相互指导学习;
- 4) co-teaching+(Yu等, 2019): 同时训练两个网络, 选择两个网络分歧样本中的小损失样本;
- 5) CNLCU-S(Xia等, 2022): 对损失采用区间估计的方式进行样本选择;
- 6) CoDis(Xia等, 2023): 同时训练两个网络, 应用协方差正则化方法保持两个网络之间的发散性。

3.2 实验结果分析

本文进行了大量的实验来表明本文方法的有效性。包括在模拟噪声数据集(F-MNIST、SVHN、CIFAR-10、CIFAR-100)和真实世界噪声数据集(Clothing1M)进行的实验。

3.2.1 消融实验结果分析

消融实验在模拟噪声数据集 CIFAR-10 的 2 种具有代表性的噪声类型下进行实验, 分别为对称噪声(symmetric)和成对翻转噪声(pairflip), 并在 20% 和 40% 噪声率下分别进行实验, 以验证本文方法各项改进的有效性。实验结果如表 1 所示。消融实验使用的评价指标以及实验设置同对比实验保持一致。

表 1 在 CIFAR-10 数据集上进行的消融实验
Table 1 Ablation experiment on the CIFAR-10 dataset

基线	样本选择	修正损失	差异损失	/%			
				symmetric		pairflip	
				20%	40%	20%	40%
√	-	-	-	75.74±0.33	58.64±0.81	76.38±0.35	55.03±0.27
√	√	-	-	82.54±0.28	77.36±0.41	82.75±0.21	75.94±0.14
√	√	√	-	85.66±0.12	81.63±0.33	86.45±0.14	82.21±0.76
√	√	-	√	86.04±0.20	80.74±0.24	86.49±0.18	80.77±0.21
√	√	√	√	87.47±0.05	83.32±0.14	87.16±0.16	85.60±0.13

注: 加粗字体表示最优准确率, “√”表示采用, “-”表示未采用。

首先, 在基线模型上添加样本选择策略进行实验, 在对称噪声(symmetric) 20% 和 40% 噪声率下测试准确率分别提高了 6.8% 和 18.72%, 在成对噪声(pairflip) 20% 和 40% 噪声率下测试准确率分别提高了 6.37% 和 20.91%。结果表明了本文方法中考虑历史损失的样本选择策略可以有效筛选出干净样本, 避免模型拟合噪声样本。

其次, 在其基础上分别添加修正损失函和差异损失函数进行实验。添加修正损失函数的实验中, 在对称噪声(symmetric) 20% 和 40% 噪声率下测试准确率分别提高了 3.12% 和 4.27%, 在成对噪声(pairflip) 20% 和 40% 噪声率下测试准确率分别提高了 3.7% 和 6.27%。添加差异损失函数的实验中, 在对称噪声(symmetric) 20% 和 40% 噪声率下测试准确率分别提高了 3.50% 和 3.38%, 在成对噪声(pairflip) 20% 和 40% 噪声率下测试准确率分别提高了

3.74% 和 4.83%。测试准确率的进一步提高充分说明修正损失函数有效地减弱了噪声标签给模型带来的影响, 并且差异损失函数可以改善协同学习模式下两个网络过早收敛导致学习停止的问题。

最终, 添加本文提出的所有函数进行实验, 测试准确率达到最高, 同时后 10 轮准确率标准差相比于基线均有所降低。说明本文方法不仅可以提升模型性能, 也可以使模型的收敛更为稳定。

3.2.2 修正损失函数分析

对修正损失函数所产生的作用进行了实验证明。选取准确率和纯净率两个指标, 其中, 纯净率为模型选取样本中干净样本数量除以模型选取样本总数的百分比。

由于 CE 存在过拟合问题, 模型可能会拟合噪声样本, 一方面破坏了模型泛化性, 导致测试准确率降低, 另一方面由于模型记忆了噪声样本, 在训练过程

中其区分干净和噪声样本的能力会减弱,导致模型选择样本的纯净率降低。因此,CE过拟合的问题具体会反映在测试准确率和模型筛选样本纯净率的降低。选取 co-teaching 为基线(其训练损失函数为

CE),在其基础上,仅将其训练损失 CE 替换为本文所提出的修正损失函数进行对比实验(CIFAR-10数据集,symmetric 噪声类型,40%噪声率)。实验结果如图3所示。

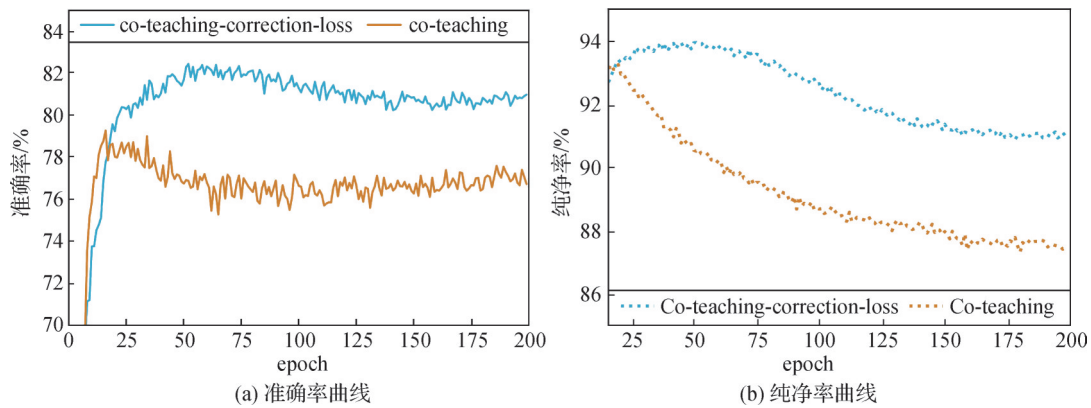


图3 修正损失函数实验

Fig. 3 Correction loss experiment ((a) test accuracy; (b) pure rate)

实验结果表明,在修正损失函数的作用下,co-teaching 学习模式的性能有了一定的提升,并且随着迭代轮次的增加,选择样本的纯净率仍保持较高水平。这些情况证明了修正损失函数可以缓解 CE 带来的过拟合影响,减弱模型拟合噪声样本的情况。

3.2.3 模拟噪声数据集实验结果分析

表2展示了F-MNIST数据集下本文方法及所有对比算法的实验结果。测试准确率曲线如图4所示,标准差用阴影表示。

在F-MNIST数据集上,Co-history 在所有噪声类

型和噪声率下均达到了最优测试准确率,相比于 co-teaching, Co-history 在对称噪声类型(symmetric)20%和40%噪声率下测试准确率分别提高了2.05%和3.52%,在非对称噪声类型(asymmetric)20%和40%噪声率下测试准确率分别提高了2.19%和3.60%,在成对噪声类型(pairflip)20%和40%噪声率下测试准确率分别提高了3.06%和4.06%,在三角对称噪声类型(tridiagonal)20%和40%噪声率下测试准确率分别提高了2.58%和4.11%。测试准确率的提高可以证明本文提出方法的有效性,特别是在高噪声

表2 不同方法在F-MNIST数据集上的对比分析

Table 2 Comparisons with different methods on F-MNIST dataset

方法	/%							
	symmetric		asymmetric		pairflip		tridiagonal	
	20%	40%	20%	40%	20%	40%	20%	40%
Standard	87.46±0.09	72.46±0.68	89.34±0.15	77.73±0.19	84.23±0.23	61.46±0.32	85.86±0.4	66.44±0.28
APL	93.08±0.02	90.08±0.08	92.91±0.05	78.35±0.33	91.98±0.12	65.25±0.52	92.87±0.03	83.56±0.31
co-teaching	91.20±0.03	88.06±0.06	91.29±0.16	86.32±0.09	89.97±0.09	84.99±0.11	90.54±0.05	86.29±0.09
co-teaching+	92.85±0.03	91.42±0.03	93.42±0.02	89.38±0.17	92.90±0.03	88.38±0.16	92.75±0.08	90.40±0.08
CNLCU-S	92.36±0.12	91.32±0.07	92.06±0.04	81.78±0.09	91.69±0.08	87.74±0.10	92.20±0.08	89.74±0.07
CoDis	90.97±0.05	87.92±0.10	91.55±0.08	85.77±0.35	90.23±0.04	83.92±0.08	90.33±0.02	86.10±0.09
Co-history (本文)	93.25±0.02	91.58±0.03	93.48±0.02	89.92±0.28	93.03±0.04	89.05±0.21	93.12±0.03	90.40±0.07

注:加粗字体表示各列最优结果。

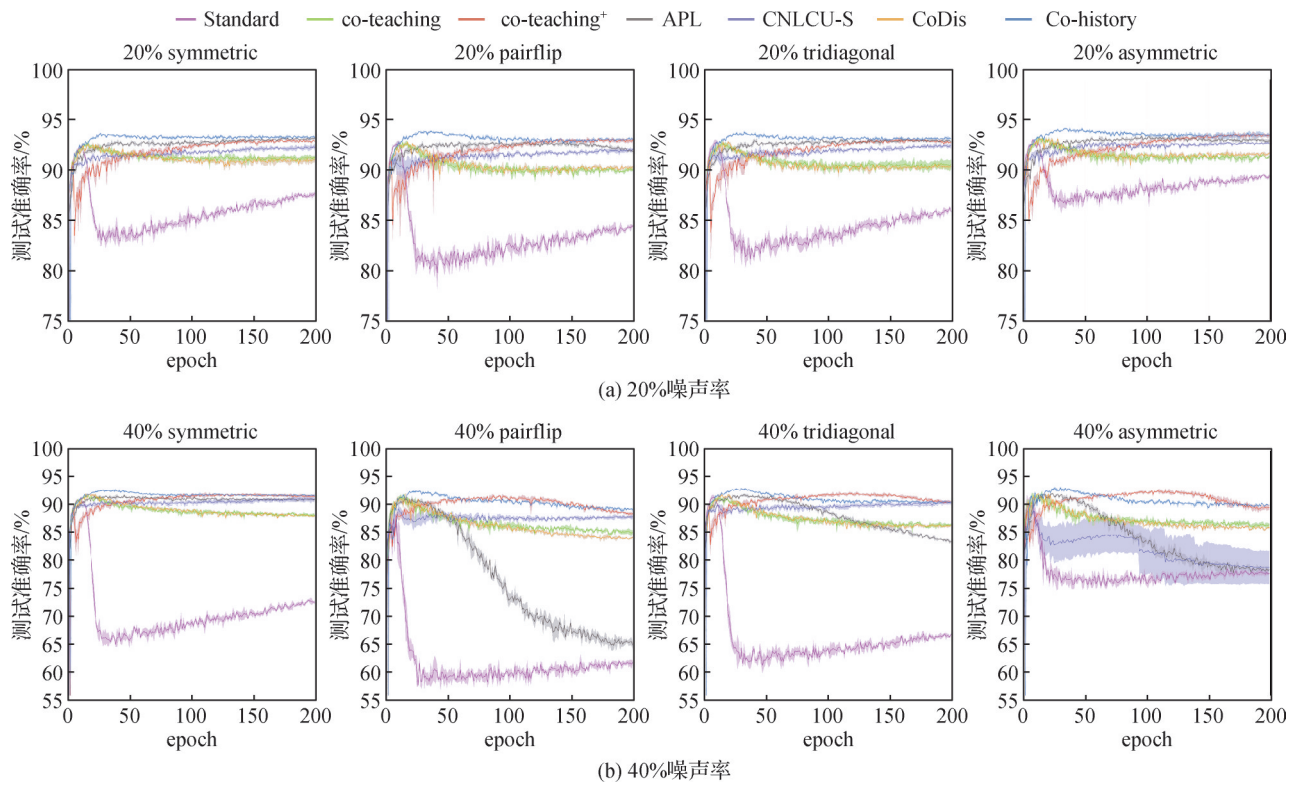


图4 F-MNIST数据集测试准确率

Fig. 4 Test accuracy on F-MNIST dataset((a) 20% noise rate; (b) 40% noise rate)

率环境下, 本文的实验结果相较于 co-teaching 有着更为明显的提升, 进一步说明了本文提出的方法更具优势。

表3展示了 SVHN数据集下 Co-history 及所有对算法的最后10轮的平均测试准确率和标准差。测试准确率曲线如图5所示。

表3 不同方法在SVHN数据集上的对比分析

Table 3 Comparisons with different methods on SVHN dataset

方法	symmetric		asymmetric		pairflip		tridiagonal	
	20%	40%	20%	40%	20%	40%	20%	40%
	/%							
Standard	87.21±0.1	70.97±0.34	89.55±0.05	77.79±0.21	84.84±0.1	61.39±0.35	86.37±0.2	67.25±0.44
APL	95.12±0.01	92.71±0.02	94.22±0.04	79.04±0.6	93.37±0.01	65.42±0.17	94.92±0.02	85.24±0.32
co-teaching	91.98±0.04	89.21±0.23	91.91±0.35	88.0±0.43	91.24±0.27	85.94±0.06	92.03±0.04	87.94±0.04
co-teaching+	94.71±0.01	92.86±0.03	94.37±0.03	88.14±0.06	94.15±0.02	88.68±0.14	94.74±0.01	91.88±0.06
CNLCU-S	93.16±0.10	91.91±0.06	94.67±0.04	92.84±0.06	93.52±0.04	89.99±0.10	93.32±0.05	90.76±0.08
CoDis	91.67±0.04	89.19±0.08	92.10±0.06	87.27±0.18	91.25±0.07	85.09±0.11	91.83±0.03	88.34±0.01
Co-history (本文)	95.43±0.01	93.98±0.01	94.80±0.04	91.48±0.07	94.60±0.04	90.95±0.15	95.3±0.01	92.98±0.03

注: 加粗字体表示各列最优结果。

Co-history 在绝大部分噪声类型和噪声率下达到了最优测试准确率。可以证明本文针对 co-teaching 中存在问题的改进是有效的。同时, 对比同

样采用分歧策略的 co-teaching+ 和 CoDis, 本文方法在所有噪声类型下的测试准确率均优于 co-teaching+ 和 CoDis, 特别是在噪声类型相对复杂、不易学习的

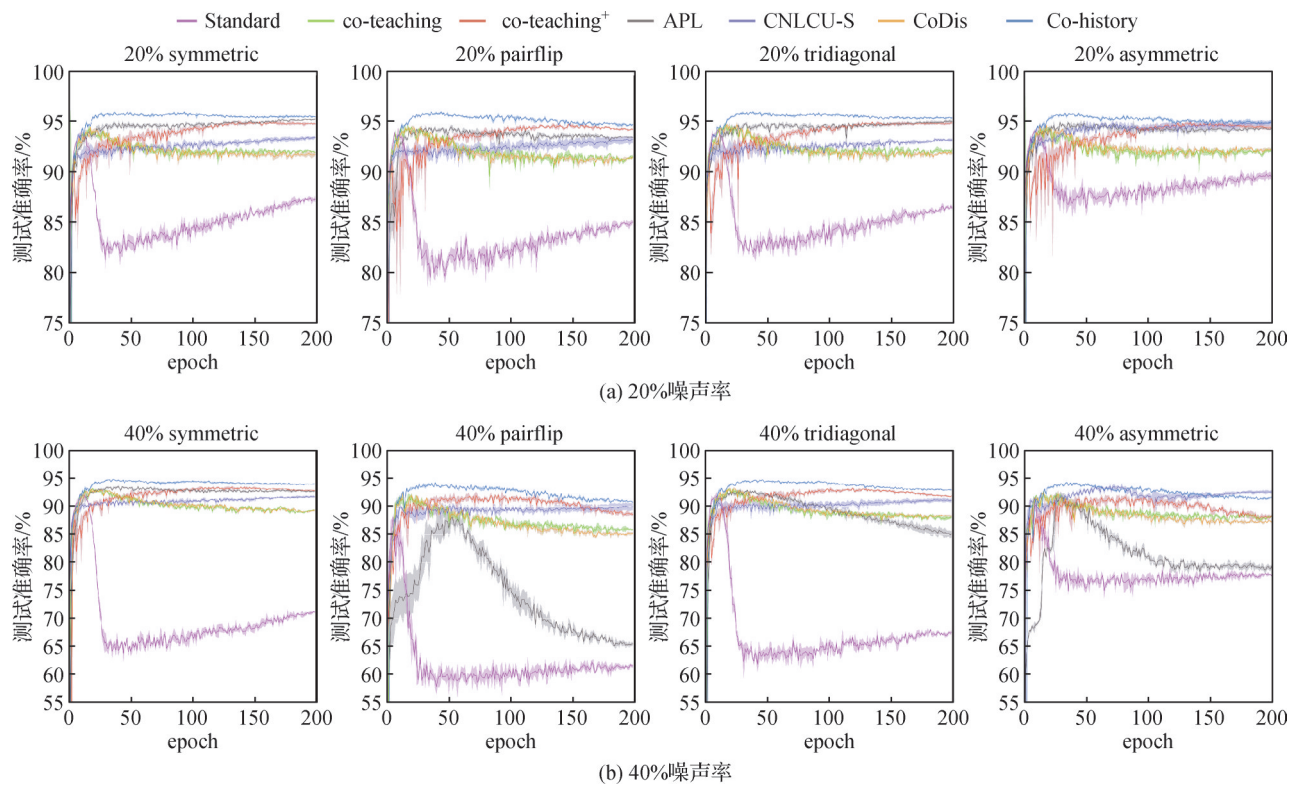


图5 SVHN数据集测试准确率

Fig. 5 Test accuracy on SVHN dataset((a) 20% noise rate (b) 40% noise rate)

成对噪声类型(pairflip)40%噪声率下,Co-history相较于co-teaching+和CoDis分别高出2.27%和5.86%,进一步说明本文方法提出维持两个网络差异的方法更为合理。

表4展示了CIFAR-10数据集下Co-history及所有对比算法的最后10轮的平均测试准确率和标准差。测试准确率曲线如图6所示。

CIFAR-10数据集的识别难度更大,对模型的训练更具挑战性。Co-history方法在此数据集上表现出优异的性能,在所有噪声类型下均达到了最佳准确率,尤其是在高噪声率下,在对称噪声类型(symmetric)40%噪声率下,Co-history相比Co-teaching提升6.16%,在非对称噪声类型(asymmetric)40%噪声率下提高了9.59%,在成对噪声类型(pairflip)40%

表4 不同方法在CIFAR-10数据集上的对比分析

Table 4 Comparisons with different methods on CIFAR-10 dataset

方法	/%							
	symmetric		asymmetric		pairflip		tridiagonal	
	20%	40%	20%	40%	20%	40%	20%	40%
Standard	75.74±0.33	58.64±0.81	82.32±0.14	72.19±0.32	76.38±0.35	55.03±0.27	76.26±0.26	58.72±0.26
APL	84.20±0.11	76.19±0.20	83.36±0.21	71.14±0.15	81.48±0.08	52.98±0.18	83.29±0.09	65.80±0.62
co-teaching	82.24±0.18	77.16±0.10	80.76±0.11	72.85±0.11	82.55±0.10	75.74±0.14	82.50±0.12	76.28±0.12
co-teaching+	81.96±0.12	71.49±0.33	79.68±0.13	70.96±0.69	79.71±0.14	58.39±0.76	81.15±0.05	64.79±0.47
CNLCU-S	83.31±0.12	78.60±0.17	85.24±0.15	73.96±0.10	82.55±0.39	73.35±0.16	83.31±0.19	76.11±0.13
CoDis	82.36±0.24	77.04±0.09	84.58±0.05	75.30±0.32	82.53±0.23	70.86±0.22	82.69±0.07	74.59±0.05
Co-history (本文)	87.47±0.05	83.32±0.14	87.95±0.12	82.44±0.09	87.16±0.16	85.60±0.13	86.98±0.05	82.20±0.16

注:加粗字体表示各列最优结果。

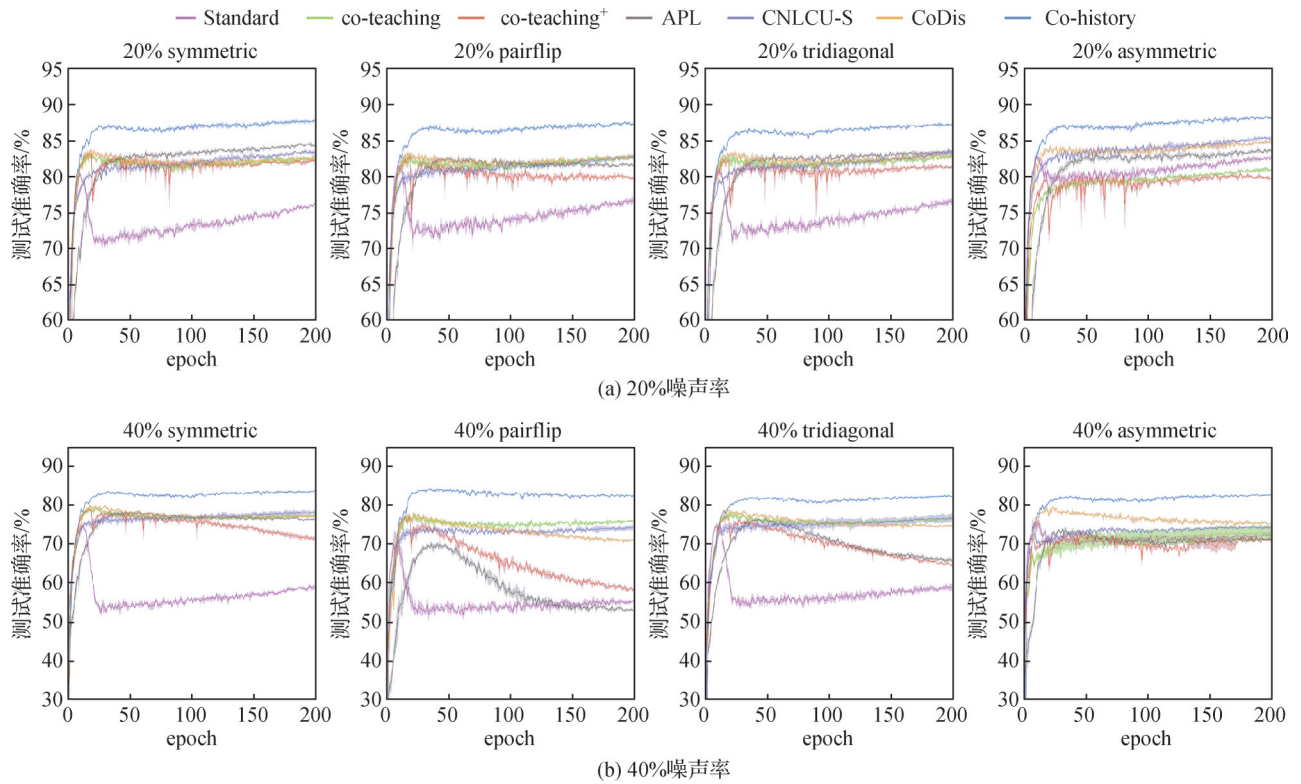


图6 CIFAR-10数据集测试准确率

Fig. 6 Test accuracy on CIFAR-10 dataset((a) 20% noise rate;(b) 40% noise rate)

噪声率下提高了9.86%,在三角对称噪声类型(tridiagonal)40%噪声率下提高了5.92%。实验结果充分说明本文提出的方法在识别难度更大的CIFAR-10数据集中依然可以有优秀的表现,相比于其他方法,

Co-history方法更具鲁棒性。

表5展示了CIFAR-100数据集下Co-history及所有对比算法的最后10轮的平均测试准确率和标准差。测试准确率曲线如图7所示。

表5 不同方法在CIFAR-100数据集中的对比分析

Table 5 Comparisons with different methods on CIFAR-100 dataset

方法	/%							
	symmetric		asymmetric		pairflip		tridiagonal	
	20%	40%	20%	40%	20%	40%	20%	40%
Standard	31.98±0.30	20.36±0.60	39.21±0.06	27.80±0.09	39.30±0.03	27.65±0.02	38.73±0.07	27.81±0.05
APL	44.66±0.04	37.83±0.04	45.39±0.07	32.9±0.03	44.81±0.04	31.31±0.08	45.48±0.03	38.07±0.04
co-teaching	43.61±0.09	37.96±0.14	43.37±0.11	31.84±0.1	43.37±0.11	31.84±0.10	44.46±0.08	36.11±0.06
co-teaching+	49.41±0.02	43.89±0.07	48.78±0.09	35.94±0.03	48.33±0.02	36.24±0.09	48.95±0.05	42.99±0.05
CNLCU-S	47.17±0.02	41.91±0.08	42.96±0.02	29.97±0.05	43.16±0.01	30.25±0.02	46.00±0.01	35.70±0.01
CoDis	43.16±0.08	38.01±0.06	39.42±0.01	28.51±0.06	43.31±0.07	31.37±0.14	44.48±0.17	37.40±0.09
Co-history(本文)	49.60±0.07	44.92±0.04	45.07±0.11	32.52±0.04	48.42±0.04	34.55±0.05	49.48±0.05	40.03±0.16

注:加粗字体表示各列最优结果

尽管Co-history在此数据集中的表现不能全部达到最优,但是仍然具有一定的竞争力,在部分噪声类型和噪声率下依旧可以达到最优测试准确率。并

且Co-history方法在所有噪声类型的所有噪声率下均优于co-teaching算法,对比co-teaching,在对称噪声(symmetric)20%和40%噪声率下,分别提高了

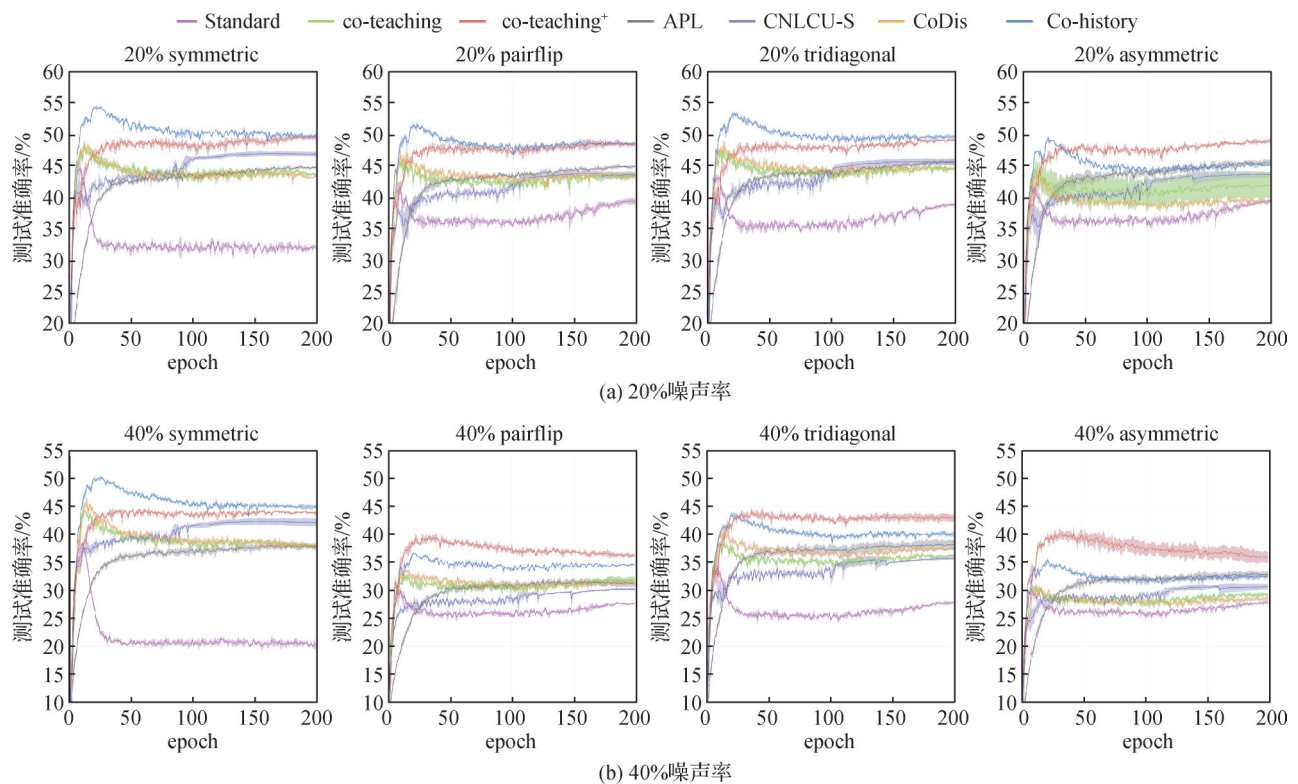


图7 CIFAR-100数据集测试准确率

Fig. 7 Test accuracy on CIFAR-100 dataset((a) 20% noise rate; (b) 40% noise rate)

5.99%和6.96%,在非对称噪声(asymmetric)20%和40%噪声率下,分别提高了1.7%和0.68%。在成对噪声(pairflip)20%和40%噪声率下,分别提高了5.05%和2.71%。在三角对称噪声(tridiagonal)20%和40%噪声率下分别提高了5.02%和3.92%,说明本文方法在多分类数据集下,也可以抵抗两个网络发生收敛的情况,改善模型使用CE损失训练出现的过拟合问题。

3.2.4 真实世界噪声数据集实验结果分析

为了验证本文方法在实际应用中的效果,本文进一步在真实数据集Clothing1M上进行实验,测试准确率如表6所示,Best表示训练过程中的最高测试精度,Last表示训练结束时的测试精度。

从表6可以看出,在Clothing1M数据集上,Co-history相比Co-teaching算法,Best提高了0.94%,Last提高了1.2%,同时可以观察到co-teaching+在真实数据集中表现出了较差的效果,进一步印证了本文所提出的差异损失函数更为合理。Co-history在与其他先进的方法进行比较时也表现出一定的竞争力,验证了本文方法在真实世界数据集上的有效性。

表6 不同方法在Clothing1M数据集上的对比分析

Table 6 Comparisons with different methods on Clothing1M dataset

方法	/%	
	Best	Last
Standard	67.22	64.68
APL	56.01	55.84
co-teaching	69.21	68.51
co-teaching+	59.32	58.79
CNLCU-S	71.37	70.76
CoDis	70.48	69.95
Co-history(本文)	70.15	69.81

注:加粗字体表示各列最优结果。

4 结论

本文针对模型使用CE损失训练存在的过拟合现象,结合样本损失的历史规律,提出了修正损失函数。同时,针对Co-teaching中存在的两个网络过早收敛导致学习提前停止的问题,提出了差异损失函

数。最后,针对传统小损失样本选择策略,本文充分考虑样本损失的历史规律,提出了更为精准的样本选择策略。不同于以往 co-teaching 中的学习策略,通过加入新的视角,提升模型性能。并且,经过大量实验论证,本文方法优于基线方法,在带有噪声标签的数据集中,有着更强的鲁棒性,更加适用于噪声标签场景中。此外,消融实验清楚地证明了本文方法各项改进的有效性。

因为本文方法需要分析每个样本的历史损失信息,所以需要保存每个样本的历史损失值。当训练样本数量规模增加时,所占用的内存空间会随之增加,从而导致计算和存储成本增加。此外,在样本类别数量较多时,本文方法的性能在部分噪声环境下不能达到最优(如 CIFAR-100 数据集 asymmetric 噪声类型 20%、40% 噪声率等)。因此,在未来的工作中,将在保证精度的前提下寻找更加高性能的解决方案,进一步探索更加优秀的噪声标签鲁棒分类算法。

参考文献 (References)

- Arpit D, Jastrzebski S, Ballas N, Krueger D, Bengio E, Kanwal M S, Maharaj T, Fischer A, Courville A, Bengio Y and Lacoste-Julien S. 2017. A closer look at memorization in deep networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org: 233-242
- Blum A, Kalai A and Wasserman H. 2003. Noise-tolerant learning, the parity problem, and the statistical query model. *Journal of the ACM (JACM)*, 50(4): 506-519 [DOI: 10.1145/792538.792543]
- Diaz F. 2009. Integration of news content into web results//Proceedings of the 2nd ACM International Conference on Web Search and Data Mining. Barcelona, Spain: ACM: 182-191 [DOI: 10.1145/1498759.1498825]
- Ding Y F, Wang L Q, Fan D L and Gong B Q. 2018. A semi-supervised two-stage approach to learning from noisy labels//Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe, USA: IEEE: 1215-1224 [DOI: 10.1109/wacv.2018.00138]
- Ghosh A, Kumar H and Sastry P S. 2017. Robust loss functions under label noise for deep neural networks//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI: 1919-1925
- Ghosh A, Manwani N and Sastry P S. 2015. Making risk minimization tolerant to label noise. *Neurocomputing*, 160: 93-107 [DOI: 10.1016/j.neucom.2014.09.081]
- Han B, Yao Q M, Yu X R, Niu G, Xu M, Hu W H, Tsang I W and Sugiyama M. 2018. Co-teaching: robust training of deep neural networks with extremely noisy labels//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc.: 8536-8546
- Huang Z, Zhang J and Shan H. 2023. Twin contrastive learning with noisy labels//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 11661-11670 [DOI: 10.1109/cvpr52729.2023.01122]
- Jiang L, Zhou Z Y, Leung T, Li L J and Li F F. 2018. MentorNet: learning data-driven curriculum for very deep neural networks on corrupted labels//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR: 2304-2313
- Kim Y, Yun J, Shon H and Kim J. 2021. Joint negative and positive learning for noisy labels//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 9437-9446 [DOI: 10.1109/cvpr46437.2021.00932]
- Krizhevsky A. 2009. Learning multiple layers of features from tiny images [EB/OL]. [2023-07-02]. <http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf>
- Krizhevsky A, Sutskever I and Hinton G E H. 2017. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90 [DOI: 10.1145/3065386]
- Li J N, Socher R and Hoi S C H. 2020. DivideMix: learning with noisy labels as semi-supervised learning//Proceedings of the 8th International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview.net: 1-14
- Li Y X, Shen J R and Xu Q. 2023. A summary of image recognition-relevant multi-layer spiking neural networks learning algorithms. *Journal of Image and Graphics*, 28(2): 385-400 (李雅馨, 申江荣, 徐齐. 2023. 面向图像识别的多层脉冲神经网络学习算法综述. *中国图象图形学报*, 28(2): 385-400) [DOI: 10.11834/jig.220452]
- Liu J R, Jiang D G, Yang Y K and Li R R. 2022. Agreement or disagreement in noise-tolerant mutual learning?//Proceedings of the 26th International Conference on Pattern Recognition. Montréal, Canada: IEEE: 4801-4807 [DOI: 10.1109/icpr56361.2022.9956595]
- Liu T L and Tao D C. 2016. Classification with noisy labels by importance reweighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3): 447-461 [DOI: 10.1109/tpami.2015.2456899]
- Lukasik M, Bhojanapalli S, Menon A K and Kumar S. 2020. Does label smoothing mitigate label noise?//Proceedings of the 37th International Conference on Machine Learning. Vienna, Austria: PMLR: 6448-6458
- Ma X J, Huang H X, Wang Y S, Sarah Erfani S R and Bailey J. 2020. Normalized loss functions for deep learning with noisy labels//Proceedings of the 37th International Conference on Machine Learning. Vienna, Austria: JMLR.org: 6543-6553
- Mahajan D, Girshick R, Ramanathan V, He K M, Paluri M, Li Y X,

- Bharambe A and van der Maaten L. 2018. Exploring the limits of weakly supervised pretraining//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 185-201 [DOI: 10.1007/978-3-030-01216-8_12]
- Netzer Y, Wang T, Coates A, Bissacco A, Wu B and Ng A Y. 2011. Reading digits in natural images with unsupervised feature learning//Proceedings of NIPS Workshop on Deep Learning and Unsupervised Feature Learning. Granada, Spain: NIPS: 462-471
- Patrini G, Rozza A, Menon A K, Nock R and Qu L Z. 2017. Making deep neural networks robust to label noise: a loss correction approach//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2233-2241 [DOI: 10.1109/cvpr.2017.240]
- Pham H, Dai Z H, Xie Q Z and Le Q V. 2021. Meta pseudo labels//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11552-11563 [DOI: 10.1109/CVPR46437.2021.01139]
- Song H, Kim M, Park D, Shin Y and Lee J G. 2023. Learning from noisy labels with deep neural networks: a survey. IEEE Transactions on Neural Networks and Learning Systems, 34(11): 8135-8153 [DOI: 10.1109/tnnls.2022.3152527]
- Wang Y S, Ma X J, Chen Z Y, Luo Y, Yi J F and Bailey J. 2019. Symmetric cross entropy for robust learning with noisy labels//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 322-330 [DOI: 10.1109/iccv.2019.00041]
- Wei H X, Feng L, Chen X Y and An B. 2020. Combating noisy labels by agreement: a joint training method with co-regularization//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 13723-13732 [DOI: 10.1109/cvpr42600.2020.01374]
- Wei Q, Sun H L, Lu X K and Yin Y L. 2022. Self-filtering: a noise-aware sample selection for label noise with confidence penalization//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 516-532 [DOI: 10.1007/978-3-031-20056-4_30]
- Xia X B, Han B, Zhan Y B, Yu J, Gong M M, Gong C and Liu T L. 2023. Combating noisy labels with sample selection by mining high-discrepancy examples//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 1833-1843 [DOI: 10.1109/iccv51070.2023.00176]
- Xia X B, Liu T L, Han B, Gong M M, Yu J, Niu G and Sugiyama M. 2022. Sample selection with uncertainty of losses for learning with noisy labels//Proceedings of the 10th International Conference on Learning Representations. Virtual Place: IEEE: 1-23
- Xiao H, Rasul K and Vollgraf R. 2017. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms [EB/OL]. [2023-07-02]. <https://arxiv.org/pdf/1708.07747.pdf>
- Xiao T, Xia T, Yang Y, Huang C and Wang X G. 2015. Learning from massive noisy labeled data for image classification//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 2691-2699 [DOI: 10.1109/cvpr.2015.7298885]
- Yan Y, Rosales R, Fung G, Subramanian R and Dy J. 2014. Learning from multiple annotators with varying expertise. Machine Learning, 95(3): 291-327 [DOI: 10.1007/s10994-013-5412-1]
- Yi K and Wu J X. 2019. Probabilistic end-to-end noise correction for learning with noisy labels//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 7010-7018 [DOI: 10.1109/cvpr.2019.00718]
- Yu X R, Han B, Yao J C, Niu G, Tsang I and Sugiyama M. 2019. How does disagreement help generalization against label corruption?//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA: PMLR: 7164-7173
- Yu X Y, Liu T L, Gong M M, Batmanghelich K and Tao D C. 2018. An efficient and provable approach for mixture proportion estimation using linear independence assumption//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4480-4489 [DOI: 10.1109/cvpr.2018.00471]
- Zhang K, Feng X H, Guo Y R, Su Y K, Zhao K, Zhao Z B, Ma Z Y and Ding Q L. 2021. Overview of deep convolutional neural networks for image classification. Journal of Image and Graphics, 26(10): 2305-2325 (张珂, 冯晓晗, 郭玉荣, 苏昱坤, 赵凯, 赵振兵, 马占宇, 丁巧林. 2021. 图像分类的深度卷积神经网络模型综述. 中国图象图形学报, 26(10): 2305-2325) [DOI: 10.11834/jig.200302]
- Zhang Y, Niu G and Sugiyama M. 2021. Learning noise transition matrix from only noisy labels via total variation regularization//Proceedings of the 38th International Conference on Machine Learning. [s.l.]: PMLR: 12501-12512
- Zhang Z L and Sabuncu M R. 2018. Generalized cross entropy loss for training deep neural networks with noisy labels//Proceedings of the 32nd Conference on Neural Information Processing Systems. Montréal, Canada: NeurIPS: 8792-8802

作者简介

董永峰,男,教授,硕士生导师,主要研究方向为机器学习、知识工程、计算机视觉和智能信息处理。

E-mail: dongyf@hebut.edu.cn

王振,通信作者,男,副教授,硕士生导师,主要研究方向为机器学习、计算机视觉和可信学习。

E-mail: wangzhen@hebut.edu.cn

李佳伟,男,硕士研究生,主要研究方向为图像处理 and 计算机视觉。E-mail: lijiawei1024@foxmail.com

贾文玉,女,硕士研究生,主要研究方向为图像处理和可信学习。E-mail: jiawenyu2021@163.com