

基于样本选择的标签含噪图像分类^①

闻 铮, 曹 国

(南京理工大学 计算机科学与工程学院, 南京 210094)

通信作者: 曹 国, E-mail: caoguo@njust.edu.cn



摘 要: 标签噪声广泛存在、无法避免且影响深度学习模型的性能. 利用神经网络的“记忆效应”, 基于小损失原则的样本选择方法能简单有效地处理标签噪声. 本文基于特征空间中样本距离越近越相似的原则, 结合样本的高低置信度假设, 提出了新的样本选择原则以及二阶段加权样本选择重标签方法 (WSSR-2s). (1) 在训练前期阶段, 对于高置信度样本, 在特征空间中对其票权进行加权, 更好地引导训练; (2) 在训练中后期阶段, 对于低置信度样本, 将其票权转移给其最相似的特征样本, 以更正确地训练. 在合成噪声数据集 CIFAR-10、CIFAR-100 以及真实噪声数据集 ANIMAL-10N、WebVision 的实验结果表明, 本文提出的方法取得更高的精度, 能够更好地处理标签噪声问题.

关键词: 标签噪声; 样本选择; 置信度假设; 样本票权; 样本加权

引用格式: 闻铮, 曹国. 基于样本选择的标签含噪图像分类. 计算机系统应用, 2024, 33(2): 54-61. <http://www.c-s-a.org.cn/1003-3254/9396.html>

Label Noisy Image Classification Based on Sample Selection

WEN Zheng, CAO Guo

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: Label noise is widely present and unavoidable, and it affects the performance of deep network models. Sample selection methods based on the principle of small loss can easily and effectively handle label noise by the “memory effect” of neural networks. This study proposes a new sample selection principle and a two-stage weighted sample selection and relabeling method (WSSR-2s) based on the principle that a closer sample distance in the feature space results in more similarity, combined with the assumption of high and low confidence of the samples. In the early training stage, for high-confidence samples, their voting rights are weighted in the feature space to better guide training. In the middle and later stages of training, for low-confidence samples, their voting rights are transferred to their most similar feature samples for more accurate training. The experimental results on synthetic noise datasets CIFAR-10 and CIFAR-100, as well as real noise datasets ANIMAL-10N and WebVision, show that the proposed method achieves higher accuracy and can better handle label noise problems.

Key words: label noise; sample selection; confidence hypothesis; sample voting right; sample weighting

1 概述

深度神经网络 (DNN) 在处理不同的计算机视觉任务中 (如图像分类、分割和目标检测) 均表现了出色的性能, 适用于不同的应用邻域. 提高深度学习模型性能的因素之一是大规模数据集的使用, 如 ImageNet. 然

而, 大规模数据集的标注工作是昂贵且耗时的^[1], 研究人员有时采取更便宜的替代方案, 如网络搜索和众包^[2], 但这可能会产生带有错误或噪声标签的数据集. 不正确的标签也可能存在于小规模数据集中, 因其标注目标困难或标注者之间存在分歧, 例如医学图像的

^① 基金项目: 国家自然科学基金 (62201282); 江苏省自然科学基金 (BK20231456)

收稿时间: 2023-08-08; 修改时间: 2023-09-09; 采用时间: 2023-09-26; csa 在线出版时间: 2023-12-18

CNKI 网络首发时间: 2023-12-20

标注。

大多数基于 DNN 的工作都假设标签是由专家标注的,或者是经过设计的,因此会被认为是完美标注。然而,这个假设往往不符合现实,标注的正确性会受到数据来源,数据获取数据量以及人为因素等影响^[3]。McNicol^[4]指出,当涉及人类标注时,噪声标签的产生可能无法避免。因此,使用噪声标签训练的 DNN 可能会降低准确性并需要更大的训练集。Zhang 等人^[5]已经表明,卷积神经网络(CNN)可以很容易地适应任何比率的噪声标签,导致泛化性能变差。然而,研究表明,模型会首先学习简单模式,对应于具有干净标签的简单样本,而更接近于噪声标签的困难模式则会在之后学习甚至“强行记忆”。这就是神经网络的“记忆效应”。

基于这一特性,用样本选择方法解决标签噪声问题无疑是简单直接的思路,而这其中又多以小损失原则为干净样本假设。这些工作试图从嘈杂的实例中选择干净的实例,然后使用它们来更新网络。直观地,随着训练数据的标签噪声更少,模型可以获得更好的性能。其中,代表性的方法是 MentorNet 和 Decoupling。Jiang 等人^[6]提出了 MentorNet,它预先训练一个额外的网络,然后使用该网络来选择干净的实例来指导训练。当没有干净的验证数据供使用时,MentorNet 必须使用预定义的课程。同时它也继承了样本选择偏差导致误差累积的劣势。Malach 等人^[7]提出了 Decoupling 同时训练两个网络,然后仅使用来自这两个网络的具有不同预测的实例来更新模型。尽管如此,噪声标签均匀地分布在样本空间中。因此,不一致区域包括多个噪声标签,而 Decoupling 不能显式地处理噪声标签。Han 等人^[8]提出 Co-teaching,它不是训练单个深度神经网络,而是同时训练两个深度神经网络,并让它们在每一个 mini-batch 中相互教学,根据其对等网络选择的数据更新自己。由于两个网络具有不同的学习能力,它们可以过滤由噪声标签引入的不同类型的错误。然而,随着训练时间的增加,两个网络将趋于一致,Co-teaching 在功能上将退化为自我训练的 MentorNet。Yu 等人^[9]引入了“不一致更新”策略^[7]来保持 Co-teaching 网络的差异,并将其方法命名为 Co-teaching+。一般来说,Co-teaching+ 包括不一致更新步骤和协同更新步骤。在不一致更新步骤中,两个网络首先对所有数据进行前馈和预测,只保留预测不一致的数据。这一步骤确保两个网络保持差异。在协同更新步骤中沿用了 Co-teaching

的更新方法。Wei 等人^[10]使用了与 Co-teaching+ 相同的思想,但它使用的是“联合一致策略”而非“不一致更新”策略,并使用 Co-regularization 联合损失作计算。

可以总结这类方法的 3 个关键因素:(1) 使用小损失技巧;(2) 两个网络的协同更新参数;(3) 保持两个网络的差异。

除了基于小损失的样本选择方法外,鲁棒损失^[11,12],样本加权^[13,14],元学习^[15,16],组合方法^[17,18]也是常见的策略。鲁棒损失即损失校正法,通过添加正则项或修改网络概率,从而对可能与噪声样本相关的低置信度预测减轻惩罚;样本加权,即通过不同样本的权重改变,降低噪声样本的影响;将元学习方法用来重加权或者过滤噪声样本;组合方法则是结合各种方法的优点来获取更好的效果。本文对样本选择方法展开深入研究。

目前提出的方法虽然性能有所提升,但也付出了相应的代价。一是方法越来越复杂,各种方法的组合使网络模型变得庞大甚至臃肿,训练成本也大大提升。另一个方面,各种方法为了提升性能也需要额外的信息,包括需要为模型提供干净的训练子集,需要假设噪声的比率,甚至是模拟噪声的分布。这些条件都让标签噪声问题变得复杂且特殊,与噪声标签的普遍存在现实相背。于是 Feng 等人^[19]考虑了一个新的问题设置,学习未知标签噪声(LULN),即模型学习时噪声的程度和类型是未知的。为了追求简单性和鲁棒性,他们提出了一种简单的 LUNL 方法,即样本选择和重标签(SSR),其中两个组件相互独立:选择机制,识别具有正确标签的干净样本,以及重标签机制,旨在恢复错误标记的噪声样本的正确标签。这两个主要组成部分基于 LULN 的两个简单而必要的假设,即具有与其邻域样本高度一致标注的样本通常是干净的,并且高置信的模型预测通常是值得信赖的。一旦以这种方式构建了良好标记的子集,就可以使用最基本的监督训练方案,并使用交叉熵损失作计算。

SSR 方法简单高效,既不需要结合半监督学习、自监督模型的预训练以及模型协同训练,又不需要额外的干净子集辅助训练。然而,SSR 对特征空间的样本信息处理简单,本文认为其仍具有进一步利用的空间,在保有简洁高效特点的同时,进一步提升对噪声样本的处理效果。

本文的主要工作如下。

(1) 本文基于样本选择重标签方法, 提出了样本置信度特性假设, 即高置信度样本为干净样本概率大, 低置信度样本为干净样本概率小. 并提出了一种新的对标签噪声的图像分类方法, 即二阶段加权样本选择重标签方法 (WSSR-2s).

(2) 针对高低置信度样本, 设计了不同的样本选择优化策略. 从样本票权角度, 对于高置信度样本, 进行票权加权; 对于低置信度样本, 进行票权转移.

(3) 本文进行了大量实验证明了本文方法的有效性.

2 相关工作

目前很多方法^[6,14,17]都受限于其模型的复杂程度或是额外的附加条件. SSR 方法从简单适用的角度出发, 沿用了样本选择方法, 这也是最直观的解决标签噪声问题的思路. 但是其假设不同于 MentorNet、Decoupling、Co-teaching 等基于小损失方法, 而是基于邻域样本高度一致标注的样本通常是干净的这一假设. 即在特征表示层面上, 对样本作 K 近邻算法 (KNN), 用余弦相似度度量, 以统计其邻域样本对样本各类别的票权. 当最多票权类别与标签相一致时, 认为该样本是干净样本, 进而选择其加入干净子集进行训练. 同时, SSR 方法还提出样本重标签机制, 基于假设样本的高置信度预测认为是正确的. 即在预测层设定高阈值, 对预测高于该阈值的样本重标签. 可以看出, 该方法简洁清晰, 同时也表现了不错的性能. 同时, 样本的选择与重标签机制是相对独立的部分, 本文认为样本置信度的潜在价值未被完全利用. 基于此提出了高低置信度假设实现进一步改进.

算法 1. SSR

输入: 数据集 (X, Y) , 样本选择阈值 θ_s , 样本重标签阈值 θ_r , 最大训练周期 T

1 当 $0 \leq i < T$ 时执行

2 生成数据集 (X, Y') ; // 根据样本重标签策略

3 生成数据集 (X_c, Y'_c) ; // 根据样本选择策略

4 训练模型;

5 结束

注意算法 1 中 Y' 表示原标签集 Y 经过样本重标签策略新生成的标签集; (X_c, Y'_c) 表示经过样本选择策略新生成的数据集. 训练模型指的是用选择出的干净数据集进行模型训练, 进而得到模型输出.

3 提出的 WSSR-2s 方法

3.1 问题定义

定义式 (1):

$$X = \{x_i\}_{i=1}^N, x_i \in R^d \quad (1)$$

为训练集 (d 表示数据映射维度), 其对应的独热向量标签为:

$$Y = \{y_i\}_{i=1}^N, y_i \in \{0, 1\}^M \quad (2)$$

其中, M 是类别数量, N 是样本数量. 方便起见, 定义对应于独热向量标签 y_i 的每个样本 x_i 的标签表示为:

$$l_i = \arg_j [y_i(j) = 1] \in \{1, \dots, M\} \quad (3)$$

其中, \arg_j 表示取满足条件的参数 j .

最后, 定义真实标签为:

$$Y' = \{y'_i\}_{i=1}^N \quad (4)$$

显然, 对于开集噪声标签来说 $y'_i \neq y_i, y'_i \notin \{0, 1\}^M$, 对于闭集噪声标签则 $y'_i \neq y_i, y'_i \in \{0, 1\}^M$.

本文将用于提取特征表示的分类网络视为编码器 f , 定义一个参数模型分类器 (PMC) g_p 用来处理分类问题, 同时定义一个基于编码器 f 所提取特征表示的非参数 K 近邻分类器 (NPK) g_q . 简单起见, 本文定义样本 x_i 的特征表示为:

$$f_i \triangleq f(x_i) \quad (5)$$

并将经 PMC 和 NPK 的预测向量分别定义为:

$$p_i \triangleq g_p(f_i) \quad (6)$$

$$q_i \triangleq g_q(f_i) \quad (7)$$

本文采取分阶段训练方案: (1) 高置信度样本票权加权 (训练初期), 以及 (2) 低置信度样本票权转移 (训练中后期).

3.2 根据样本置信度改进样本票权

本文的样本置信度 θ_h, θ_l 于样本重标签方法的 PMC 分类器 g_p 设置, 样本票权则是基于样本选择方法的特征邻域样本的标签 y'_i 设置. 根据高低置信度样本对特征邻域的样本票权改进, 以下是具体的方法.

(1) 高置信度样本票权加权

首先, 高置信度样本基于 PMC 分类器 g_p , 通过设定高阈值 θ_h 得到. 具体地, 获取高置信度样本的索引集, 即用 H 表示样本集 X 中 PMC 分类器 g_p 预测 p_i , 其最大元素超过阈值 θ_h 的所有样本 x_i 的索引集. 其定义为:

$$H = \left\{ i \mid \arg_l \left[\max_l p_i(l) > \theta_h \right] \right\} \quad (8)$$

此集合表达式中 \arg_i 表示取满足条件的 i , \max 表示对应下标参数 l 使 p_i 取得最大值。

接着, 根据得到的高置信度样本索引集 H , 将对应的高置信度样本的票权进行加权。具体地, 对于包含在高置信度样本索引集中的样本索引 i , 其对应样本 x_i 的样本票权 y_i^r 乘权重系数 w 以加权; 对于不包含在高置信度样本索引集中的样本索引 i , 其对应样本 x_i 的样本票权 y_i^r 保持不变。形式上:

$$y_i^h = \begin{cases} w \cdot y_i^r, & i \in H \\ y_i^r, & i \notin H \end{cases} \quad (9)$$

这样, 加权过票权的高置信度样本票权可以修正 NPK 分类器 g_q 的预测向量 q_i , 进一步帮助在模型训练初期更正确地选择出干净样本。

(2) 低置信度样本票权转移

同样地, 低置信度样本基于 PMC 分类器 g_p , 通过设定低阈值 θ_l 得到。具体地, 获取低置信度样本的索引集, 即用 L 表示样本集 X 中 PMC 分类器 g_p 的预测 p_i , 其最大元素不及阈值 θ_l 的所有样本 x_i 的索引集。其定义为:

$$L = \left\{ i \mid \arg_i \left[\max_l p_i(l) < \theta_l \right] \right\} \quad (10)$$

此集合表达式中 \arg_i 表示取满足条件的 i , \max 表示对应下标参数 l 使 p_i 取得最大值。

接着, 根据得到的低置信度样本索引集 L , 将对应的低置信度样本的票权进行转移。具体地, 对于包含在低置信度样本索引集中的样本索引 i , 其对应样本 x_i 的样本票权 y_i^r 转移给在特征表示中与 x_i 最相似的样本 x_s 。具体来说, 本文根据前文 SSR 方法, 用 s_{ij} , $i, j = 1, \dots, N$ 表示任何两个样本 x_i 和 x_j 的特征表示 f_i 和 f_j 之间的相似性。默认情况下, 本文使用余弦相似度, 即:

$$s_{ij} \triangleq \frac{f_i^T f_j}{\|f_i\|_2 \|f_j\|_2} \quad (11)$$

同时基于所计算的相似性用 S_i 表示 X 中的样本 x_i 的最近邻样本的索引。接着, 将原本的低置信度样本 x_i 的票权值置 0, 其对应的样本 x_{S_i} 的票权值加 1; 对于不包含在低置信度样本索引集中的样本索引 i , 其对应的样本票权 y_i^r 保持不变。形式上:

$$y_i^l = \begin{cases} y_{S_i}^r, & i \in L \\ y_i^r, & i \notin L \end{cases} \quad (12)$$

注意定义:

$$y_{S_i}^r = \sum y_i^r \quad i \in S_i \quad (13)$$

另需将被转移的票权置零, 这样, 转移过票权的低置信度样本票权可以修正 NPK 分类器 g_q 的预测向量 q_i , 进一步帮助模型训练在中后期更好地处理噪声样本。

至此算法 WSSR-2s 的流程总结如算法 2。

算法 2. WSSR-2s

输入: 数据集 (X, Y) , 高置信度样本阈值 θ_h , 低置信度样本阈值 θ_l , 最大训练周期 T , 训练转折期 T'

- 1 当 $0 \leq i < T$ 时执行
- 2 根据 SSR 进行样本重标签和选择;
- 3 当 $i < T'$ 时执行
- 4 根据式 (9) 得到高置信度样本票权 y_i^h ; // 高置信度样本票权加权
- 5 当 $i \geq T'$ 时执行
- 6 根据式 (12) 得到低置信度样本票权 y_i^l ; // 低置信度样本票权转移
- 7 根据干净样本训练模型;
- 8 结束

(3) 对分阶段不同置信度票权改进的解释

本文将训练阶段分为初期和中后期。在训练初期, 对高置信度样本票权加权; 在训练中后期, 对低置信度样本票权转移。这里, 本文讨论这两种方法的有效性。

在训练初期, 大多数样本的置信度都处在较低水平。对于那些能给出高置信度的样本, 本文假设其标签正确的概率大。如图 1 所示, 竖线圆表示为当前样本, 虚线圈出的则是样本的特征邻域。实心圆 (或三角)、空心圆、斜线圆代表 3 种样本类别。在不考虑高置信度样本的情况下, 特征邻域里 3 种类别样本数量相同, 对应的样本票权也相同。这样, 样本无法由票权确定所属类别。现在, 考虑高置信度样本, 对应于图 1 中的实心三角形样本。当对这些样本票权进行加权后, 样本类别可以由票权确定, 并且因为高置信度特性, 所确定的样本类型大概率是正确的, 与之前的假设保持一致。另一种形象的解释是, 在一场选举初期, 那些有想法有意见的人往往有更高的话语权。需要说明的是, 此方法看起来和样本重标签方法有相似之处, 但是进一步利用了高置信度样本的价值。将 PMC 分类器和 NPK 分类器通过高置信度样本联系起来, 利用高置信度样本的特性提升样本选择的正确率。

随着训练的进行, 越来越多的样本可以给出可靠的高置信度预测, 因此对高置信度样本票权加权的方法效果会退化。所以, 在模型训练的中后期, 本文对低置信度样本作假设, 认为其标签正确概率小。如图 2 所示, 空心三角形样本表示低置信度样本。这些样本若符合假设, 其本身的标签意义就不大。同时, 箭头指向的

是与其特征表示最相似的样本. 将低置信度样本的票权转移给最相似的样本, 让这些样本代表原样本的票权, 可以在训练中后期降低噪声样本带来的干扰. 形象的解释是, 在选举的中后期, 中间派的意见往往是找最符合自身利益的派别. 需要说明的是, 该方法在高噪声多类别的情况下, 提升样本选择的正确率是更为有效的. 给特征相似的样本进行票权转移是符合假设的改进思路.

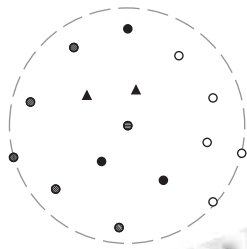


图1 高置信度样本的特征邻域

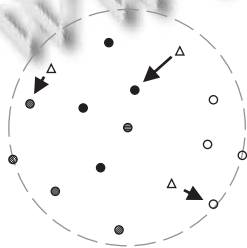


图2 低置信度样本的特征邻域

4 实验

4.1 概述

本文在基于合成噪声的标准数据集 CIFAR-10 和 CIFAR-100 以及两个真实世界噪声数据集 ANIMAL-10N 和 WebVision 进行了广泛实验. 简单起见, 本文为相应的噪声设置定义了缩写名称, 例如“Sym 50”表示 50% 对称噪声, “Asym 40”表示 40% 不对称噪声. 在第 4.2 节和第 4.3 节中, 本文将所提出的方法与合成噪声数据集和真实世界噪声数据集的最新技术进行了比较. 在第 4.4 节中, 本文进行了消融实验, 使所提出分阶段样本票权改进方法的有效性得到验证. 最后第 4.5 节为计算开销分析.

4.2 合成噪声数据集

遵循先前工作, 本文使用 PresActResNet-18 作为所有的 CIFAR-10/CIFAR-100 实验的主干. 与之前使用特定预热设置的 CIFAR-10/CIFAR-100 方法不同, 本文在所有实验中设定 θ_s 为 1.0 并从头开始训练模型.

同时对于较高的噪声比 Sym 50、Sym 80 和 Sym 90 设置 θ_r 为 0.8, 之外的所有 CIFAR 实验中的设置 θ_r 为 0.9. 本文使用相同的 SGD 优化器并训练 300 个 epoch, 动量设为 0.9, 权重衰减为 $5E-4$. 初始学习率为 0.02, 并由余弦退火调度器控制. 批大小固定为 128. 对于新参数设置 $\theta_h = 0.85$, $\theta_l = 0.15$, $w = 2$, T' 设置为 epoch 的 1/6 为 50.

本文将 WSSR-2s 与最近最具竞争力的工作进行比较. 表 1、表 2 显示了 CIFAR-10 和 CIFAR-100 的结果. 再次说明, 本文方法没有使用模型协同训练或预训练. 很明显, 本文方法远远优于对比方法 (例如在 90% 对称噪声的 CIFAR-100 上的准确度为 63.6%), 不仅在对称噪声的情况下, 而且在更现实的非对称合成噪声设置中. 同时对于原方法, 在相同实验环境下, 本文方法实现了全方面改进.

表 1 CIFAR-10 数据集的实验结果 (%)

方法	Sym 20	Sym 50	Sym 80	Sym 90	Asym 40
Cross-entropy	86.8	79.4	62.9	42.7	85.0
Co-teaching ^[9]	89.5	85.7	67.4	47.9	—
F-correction ^[20]	86.8	79.8	63.3	42.9	87.2
PENCIL ^[21]	92.4	89.1	77.5	58.9	88.5
DivideMix ^[17]	96.1	94.6	93.2	76.0	93.4
ELR+ ^[22]	95.8	94.8	93.3	78.7	93.0
RRL ^[23]	95.8	94.3	92.4	75.0	91.9
NGC ^[24]	95.9	94.5	91.6	80.5	90.6
AugDesc ^[25]	96.3	95.4	93.8	91.6	94.6
C2D ^[26]	96.4	95.3	94.4	93.6	93.5
SSR ^[19]	96.4	96.1	95.3	94.0	95.5
WSSR-2s	96.8	96.3	95.6	94.1	95.8

表 2 CIFAR-100 数据集的实验结果 (%)

方法	Sym 20	Sym 50	Sym 80	Sym 90
Cross-entropy	62.0	46.7	19.9	10.1
Co-teaching ^[9]	65.6	51.8	27.9	13.7
F-correction ^[20]	61.5	46.6	19.9	10.2
PENCIL ^[21]	69.4	57.5	31.1	15.3
DivideMix ^[17]	77.3	74.6	60.2	31.5
ELR+ ^[22]	77.6	73.6	60.8	33.4
RRL ^[23]	79.1	74.8	57.7	29.3
NGC ^[24]	79.3	75.9	62.7	29.8
AugDesc ^[25]	79.5	77.2	66.4	41.2
C2D ^[26]	78.7	76.4	67.8	58.7
SSR ^[19]	79.7	76.2	71.6	63.0
WSSR-2s	80.1	76.6	71.9	63.6

表 3 显示出了本文的方法在更复杂的组合噪声场景中的性能 (组合噪声指的是噪声由闭集噪声和开集

噪声组合成的噪声,例如 NR=0.3、OR=0.5 指的是总噪声率 (noise ratio) 为 30%,其中开集噪声 (open ratio) 占总噪声 50% 的噪声设置). 当开集噪声比从 1 降低到 0.5 时,专门为开集噪声设计的先前方法迅速退化. 此外,当开集噪声比增加时,不考虑开集噪声的方法的像 DivideMix 的性能降低. EDM 修改了 DivideMix 的方法来处理组合噪声,然而,报告的结果仍大大低于本文的结果. 相对于原方法,本文的改进保持了稳定的提升.

表 3 组合噪声数据集的实验结果 (%)

Methods	Best/Last	NR=0.3		NR=0.6	
		OR=0.5	OR=1	OR=0.5	OR=1
ILON ^[13]	Best	87.4	90.4	80.5	93.4
	Last	80.0	87.4	55.2	78.0
RoG ^[27]	Best	89.8	91.4	84.1	88.2
	Last	85.9	89.8	66.3	82.1
DivideMix ^[17]	Best	91.5	89.3	91.8	89.0
	Last	90.9	88.7	91.5	88.7
EDM ^[28]	Best	94.5	92.9	93.4	90.6
	Last	94.0	91.9	92.8	89.4
SSR ^[19]	Best	96.2	96.0	95.1	93.8
	Last	96.0	95.9	95.0	93.7
WSSR-2s	Best	96.4	96.2	95.3	94.0
	Last	96.3	96.1	95.3	94.0

4.3 真实噪声数据集

对于 WebVision, 本文使用 InceptionResNetv2. 具体地, 使用 SGD 优化器训练 150 个 epoch, 动量为 0.9, 权重衰减为 $1E-4$. 初始学习率为 0.01, 并且在 50 和 100 个时期之后降低 10 倍. 批大小固定为 32. 对于 ANIMAL-10N, 本文使用 VGG-19, 并进行批归一化. 具体地, 使用 SGD 优化器训练 150 个 epoch, 动量为 0.9, 权重衰减为 $5E-4$. 初始学习率为 0.02, 并且在 50 和 100 个时期之后降低 10 倍. 批大小固定为 128. 对于所有真实世界的噪声数据集, 本文设置 θ_s 为 1 从头开始训练模型, 而 θ_r 固定为 0.95. 对于新参数设置 $\theta_h = 0.9$, $\theta_l = 0.2$, $w = 2$, T' 设置为 epoch 的 1/6 为 25.

最后, 在表 4、表 5 中, 本文分别显示了 WebVision 和 ANIMAL-10N 数据集的结果. 总而言之, 本文的方法在大规模网络爬取数据集和小规模人类标注的噪声数据集方面都取得了与当前最先进技术相比更好或具有竞争力的性能. 同时也都对原方法实现了一定的改进.

4.4 分阶段票权改进的有效性验证

本文也对所提出的分阶段票权改进方法的有效性做了消融实验来验证其有效性. 具体地, 分为原方法,

原方法加一阶段方法以及分阶段二阶段改进方法在 CIFAR-10、CIFAR-100 上 90% 以及 80% 高噪声下 (用 90%N、80%N 表示) 作对比. 需要说明的是, 这里的一阶段指的是高置信度票权加权, 二阶段指的是低置信度票权转移.

表 4 WebVision 数据集的实验结果 (%)

Methods	WebVision		ILSVRC2012	
	Top1	Top5	Top1	Top5
Co-teaching ^[8]	63.58	85.20	61.48	84.70
DivideMix ^[17]	77.32	91.64	75.20	90.84
ELR+ ^[22]	77.78	91.68	70.29	89.76
NGC ^[24]	79.16	91.84	74.44	91.04
LongReMix ^[29]	78.92	92.32	—	—
RRL ^[23]	76.3	91.5	73.3	91.2
SSR ^[19]	80.92	92.80	75.76	91.76
WSSR-2s	81.77	93.51	76.21	92.32

表 5 ANIMAL-10N 数据集的实验结果 (%)

方法	Cross-entropy	SELFIE ^[30]	PLC ^[31]	NCT ^[32]	SSR	WSSR-2s
准确率	79.4	81.8	83.4	84.1	88.2	88.7

由表 6 可以看出, 本文提出的两阶段方法都分步实现了性能提升. 值得注意的是, 在高类别高噪声的环境下, 阶段 1 高置信度样本票权加权的提升效果下降, 阶段 2 低置信度票权转移的提升效果上升. 这一现象验证了本文之前的观点, 即训练后期高置信样本数量增多, 加权效果减少, 反而是对低置信度样本的票权处理反而提升效果明显, 进一步说明本文正确地处理了标签噪声样本.

表 6 分阶段票权改进有效性的实验结果 (%)

Methods	CIFAR-10		CIFAR-100	
	80%N	90%N	80%N	90%N
SSR	95.31	93.94	71.66	63.08
WSSR-1s	95.46	94.11	71.73	63.24
WSSR-2s	95.63	94.27	71.94	63.68

4.5 计算开销分析

在表 7 中, 我们记录了模型在我们实验的数据集上每个步骤的运行时间. 可以看出, 与梯度传播的时间相比, 样本选择和重新标记的时间几乎可以忽略不计. 同时, 我们的二阶段改进方法的时间开销同样可以忽略不计. 再次证明, 我们的方法不依托于任何固定的网络模型, 不需要复杂的机制组合, 不需要额外的附加信息, 不耗费大量额外的计算开销, 仅是作为一种训练策略, 就能对噪声标签的处理产生较好效果.

表7 计算开销分析 (s)

Dataset (size)	Model training	SSR		WSSR-2s	
		Feature extraction	Sample selection	Voting weighted	Voting transfer
CIFAR (50k)	112	9	1.23	1.54	1.76
WebVision (65k)	587	109	1.48	1.69	1.91

5 结论

本文基于样本选择重标签方法 SSR, 提出了高低样本置信度假设, 并基于假设提出了分阶段的票权改进方法. 与以前的整合复杂机制以及额外信息的方法相比, 本文秉承了原方法简单和强大的特点. 同时进一步利用了样本置信度的信息, 将样本选择和重标签两个相对独立的部分联系起来. 经过广泛的实验和消融分析表明, 本文利用置信度信息进行的改进以及分阶段的策略是有效的, 效果超过了大部分先前的方法, 同时对原方法实现了可靠的改进. 后续将进一步分析样本空间中噪声样本特点, 提出合理假设以实现更好的效果.

参考文献

- 1 王晓莉, 薛丽. 标签噪声学习算法综述. 计算机系统应用, 2021, 30(1): 10–18. [doi: 10.15888/j.cnki.csa.007776]
- 2 史伟, 李超群. 一种基于预过滤和聚类处理的众包标签噪声纠正方法. 计算机应用与软件, 2022, 39(7): 6–12, 37. [doi: 10.3969/j.issn.1000-386x.2022.07.002]
- 3 张增辉, 姜高霞, 王文剑. 基于动态概率抽样的标签噪声过滤方法. 计算机应用, 2021, 41(12): 3485–3491.
- 4 McNicol D. A Primer of Signal Detection Theory. Mahwah: L. Erlbaum Associates, 2005.
- 5 Zhang CY, Bengio S, Hardt M, *et al.* Understanding deep learning requires rethinking generalization. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017.
- 6 Jiang L, Zhou ZY, Leung T, *et al.* MentorNet: Learning data-driven curriculum for very deep neural networks on corrupted labels. Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018. 2309–2318.
- 7 Malach E, Shalev-Shwartz S. Decoupling “when to update” from “how to update”. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 961–971.
- 8 Han B, Yao QM, Yu XR, *et al.* Co-teaching: Robust training

of deep neural networks with extremely noisy labels. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal: Curran Associates Inc., 2018. 8536–8546.

- 9 Yu XR, Han B, Yao JC, *et al.* How does disagreement help generalization against label corruption? Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 7164–7173.
- 10 Wei HX, Feng L, Chen XY, *et al.* Combating noisy labels by agreement: A joint training method with co-regularization. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 13723–13732.
- 11 Wang XS, Hua Y, Kodirov E, *et al.* IMAE for noise-robust learning: Mean absolute error does not treat examples equally and gradient magnitude’s variance matters. arXiv:1903.12141, 2019.
- 12 Wang YS, Ma XJ, Chen ZY, *et al.* Symmetric cross entropy for robust learning with noisy labels. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 322–330.
- 13 Wang YS, Liu WY, Ma XJ, *et al.* Iterative learning with open-set noisy labels. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8688–8696.
- 14 Guo S, Huang WL, Zhang HZ, *et al.* CurriculumNet: Weakly supervised learning from large-scale web images. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 139–154.
- 15 Ren MY, Zeng WY, Yang B, *et al.* Learning to reweight examples for robust deep learning. Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018. 4331–4340.
- 16 Shu J, Xie Q, Yi LX, *et al.* Meta-weight-net: Learning an explicit mapping for sample weighting. Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 172.
- 17 Li JN, Socher R, Hoi SCH. DivideMix: Learning with noisy labels as semi-supervised learning. Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020.
- 18 Berthelot D, Carlini N, Goodfellow I, *et al.* MixMatch: A holistic approach to semi-supervised learning. Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc.,

2019. 454.
- 19 Feng C, Tzimiropoulos G, Patras I. SSR: An efficient and robust framework for learning with unknown label noise. Proceedings of the 33rd British Machine Vision Conference. London: BMVA Press, 2022. 372.
- 20 Patrini G, Rozza A, Menon AK, *et al.* Making deep neural networks robust to label noise: A loss correction approach. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2233–2241.
- 21 Yi K, Wu JX. Probabilistic end-to-end noise correction for learning with noisy labels. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7010–7018.
- 22 Liu S, Niles-Weed J, Razavian N, *et al.* Early-learning regularization prevents memorization of noisy labels. Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 1707.
- 23 Li JN, Xiong CM, Hoi SCH. Learning from noisy data with robust representation learning. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 9465–9474.
- 24 Wu ZF, Wei T, Jiang JW, *et al.* NGC: A unified framework for learning with open-world noisy data. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 62–71.
- 25 Nishi K, Ding Y, Rich A, *et al.* Augmentation strategies for learning with noisy labels. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 8018–8027.
- 26 Zheltonozhskii E, Baskin C, Mendelson A, *et al.* Contrast to divide: Self-supervised pre-training for learning with noisy labels. Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022. 387–397.
- 27 Lee K, Yun S, Lee K, *et al.* Robust inference via generative classifiers for handling noisy labels. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 3763–3772.
- 28 Sachdeva R, Cordeiro FR, Belagiannis V, *et al.* EvidentialMix: Learning with combined open-set and closed-set noisy labels. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 3606–3614.
- 29 Cordeiro FR, Sachdeva R, Belagiannis V, *et al.* LongReMix: Robust learning with high confidence samples in a noisy label environment. Pattern Recognition, 2023, 133: 109013. [doi: 10.1016/j.patcog.2022.109013]
- 30 Song H, Kim M, Lee JG. SELFIE: Refurbishing unclean samples for robust deep learning. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 5907–5915.
- 31 Zhang YK, Zheng SZ, Wu PX, *et al.* Learning with feature-dependent label noise: A progressive approach. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.
- 32 Chen YY, Shen X, Hu SX, *et al.* Boosting co-teaching with compression regularization for label noise. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville: IEEE, 2021. 2682–2686.

(校对责编: 孙君艳)