

一种基于可信度估计单元的图像分类 噪声抑制深度学习策略

邵航¹, 黄海亮¹, 郭雨晨², 戴琼海²

(1. 浙江未来技术研究院(嘉兴), 浙江嘉兴 314000; 2. 清华大学自动化系, 北京 100084)

摘要: 近年来,深度学习越来越关注噪声抑制的研究. 本文提出了一种噪声抑制深度学习策略,该策略通过构建噪声无感网络(Noise Unaware Network, NUN)和可信度估计单元(Reliability Estimation Gate, REG)来处理训练数据含有噪声的情况. 通过对每个样本的可信度进行评估,调节其在训练时的权重,从而降低标签噪声对网络训练的影响. 随着模型的迭代更新,标签可信数据的权重将会逐渐变大,而噪声数据的权重将会被抑制. 本文通过在多个标注数据集上的实验验证了所提出的噪声抑制深度学习策略的有效性.

关键词: 深度学习; 图像分类; 噪声抑制

中图分类号: TP301.6

文献标识码: A

文章编号: 0372-2112 (2020)10-1969-07

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2020.10.014

Noise-Suppression Deep Learning Strategy via Reliability Estimation Gate for Image Classification

SHAO Hang¹, HUANG Hai-liang¹, GUO Yu-chen², DAI Qiong-hai²

(1. Zhejiang Future Technology Institute (Jiaxing), Jiaxing, Zhejiang 314000, China;

2. Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: In recent years, more and more attention has been paid to the study and research of noise suppression. This paper proposes a noise suppression deep learning strategy that deals with the presence of noise in training data by constructing a Noise Unaware Network (NUN) and Reliability Estimation Gate (REG). By evaluating the reliability of each sample and adjusting its weight during training, the influence of label noise on network training can be reduced. As the model is updated iteratively, the weight of clean data will gradually increase, while the weight of noise data will be suppressed. Experiments on multiple benchmark data sets demonstrate the effectiveness of the proposed deep learning strategy for noise suppression.

Key words: deep learning; image classification; noise-suppression

1 引言

近年来深度学习方法在图像分类和识别任务中性能优异,取得了广泛应用^[1-4]. 随着人们对网络结构和算法的深入研究,训练数据的质量开始成为制约深度学习性能的重要瓶颈. 目前深度学习需要大量带有准确标签的训练数据,而数据的标签通常是人工标注,或者从互联网上直接获取的,标注普遍存在错误,这类噪声数据对深度学习模型的性能具有潜在的严重影响,文献[5]中通过理论和实验证实了标签噪声对分类

系统的危害性.

在实际场景中,大量且精确的图像标签往往难以获得. 例如在使用 Google 等搜索引擎通过关键词进行检索图片时,大部分图片都是基于用户在互联网上传的照片,用户会根据当时的心情、时间、地点以及上传者个人的理解等信息对图片进行标注,但是由于个人认知的差异可能对同一类图像有不同理解,从而给出不一致的标签,这将导致根据标签搜索时会检索到一些无关的图像. 另外,在近些年备受关注的医疗影像诊断的疾病分类任务中,因为医学数据本身的复杂性,以及

不同医生对同一医学影像会有不同的理解,要获取完全没有标签噪声的数据集是比较困难的,然而这些噪声数据又会严重影响到训练出的模型性能.如何在模型训练时降低标签噪声的影响是当前的研究热点,在实际应用中也有重要的意义.

在本文中,针对训练数据集中存在噪声的深度模型训练问题,提出了一种基于可信度估计单元(Reliability Estimation Gate, REG)的噪声抑制深度学习策略.其主要思想是在减弱可疑噪声数据对模型性能影响的同时,增强无噪数据的影响力.为此,本文构建了一种噪声无感网络(Noise Unaware Network, NUN)并应用于有噪声的训练集,同时根据数据的特征构建 REG.即训练集里的每一个训练样本都先由 REG 进行评估,通过可信度为生成网络调节训练权重. NUN 和 REG 是迭代更新的.最后,训练数据中噪声的权重将会逐渐减小,其对模型性能的影响也会逐渐变小.综上所述,本文的主要贡献如下:

(1) 针对深度学习模型训练提出了一种有效的噪声抑制策略.

(2) 提出了可信度估计单元(REG),通过迭代训练逐步降低标签噪声对深度学习模型性能的影响.

(3) 本文通过实验将该策略应用于不同的深度学习模型训练中,论证了所设计算法的优越性.

2 相关工作

传统的分类方法中,噪声数据容易导致模型对错误的标签过度拟合,对深度学习模型来说,后果将会更加严重.目前研究训练数据存在标签噪声时的模型学习工作主要有以下三类方法^[6,7].

鲁棒学习算法.该方法的整体思路是基于避免过拟合的技术(如正则化)局部地处理标签噪声^[8-10]. Manwani 和 Sastry^[8]提出了一种基于经验风险最小化框架的二元分类学习算法.然而该方法仅对特定模型和损失函数有效,泛化性能差.随机森林^[11]算法也被证明在分类任务中是鲁棒的.然而基于噪声鲁棒的方法不能完全避免对噪声的过拟合的问题. Bartlett 等人^[12]也证明了使用的大多数损失函数对标签噪声并不是完全鲁棒的,这些问题严重阻碍了分类器性能的进一步提升和改进.

半监督算法.该方法的整体思路是使用绝对没有标签噪声的精确小数据集进行模型预训练,然后使用无标签噪声的数据与大规模有标签噪声的数据采用联合训练和多视图的方法对网络进行调优^[6,13,14]. Veit 等人^[15]提出的方法不同于监督学习中使用的标签传播算法,如约束引导^[16]和基于图形的方法^[17],而是将无标签噪声的小数据集与存在噪声的数据集混合.虽然该

方法在数据集可控的情况下性能较好,但在真实的应用场景中难以实现完全没有标签噪声数据集的构建.

学习映射规则法.该方法的整体思路是通过数据学习得到图像、标签以及噪声之间的依赖关系,删除或纠正错误标签的样本^[18,19]. Veit 等人^[15]、Xiao 等人^[6]、Natarajan 等人^[20]和 Sukhbaatar 等人^[21]分别构建了独立于输入图像的条件噪声模型.但是这类方法往往可能会过多的删除样本,这将会严重降低分类器的性能^[22].

这些方法虽然考虑了图像之间的关联,但是没有充分考虑图像与类别之间的概率分布关系,及其与深度学习框架的紧密耦合.随着深度学习方法成为处理各种识别任务最流行的解决方案^[23,24],许多研究者开始将降噪算法和深度学习模型相结合. Scott 等人^[25]提出了一种基于数据一致性的标签噪声重构方法,用于处理弱监督深度学习中的标签噪声和不完全标记数据.这些方法基于噪声分布构建映射模型以标记标签噪声,然而要求标签噪声包含明显的映射规则.一旦标签噪声是随机分布的,此类方法将无法构造映射矩阵.相比之下,本文提出的方法拥有更好的场景适应性,适用于标签噪声随机分布的情况.

3 基于可信度估计单元的噪声抑制深度学习

本文提出一种具有噪声抑制机制的深度学习策略,通过对图像分类深度学习模型进行优化,构建可信度估计单元,并对重构后的模型在有标签噪声的数据进行训练,能有效降低含标签噪声数据带来的负面影响.

3.1 整体框架

标签噪声对训练模型性能的不利影响在诸多研究中都已揭示,通过识别含有标签噪声的样本并设法减弱这些样本对网络模型训练的影响已迫在眉睫.对此,一个直观的想法是使用样本的其他特征对样本标签的真实性进行分析,比如根据 CenterLoss^[26]的算法思路,每个样本应该尽可能的靠近它们的聚类中心,根据 TripletLoss^[27]算法,同一类特征之间距离要尽可能小,而不同类之间特征距离要尽可能大.本文受此启发,试图利用训练数据的特征信息构建辅助模型,并且利用辅助模型来评价新训练样本的可信度.

与一般的深度学习模型相比,本文增加了可信度估计单元(Reliability Estimation Gate, REG)来评估每个训练样本的可信度,并计算它们的训练权重.整个学习过程如下:①基于所有训练数据对噪声无感网络(Noise Unaware Network, NUN)进行初步训练;②REG 利用来自 NUN 的特征来评估每个训练样本的可信度,并给出合适的训练权重;③使用拥有训练权重的数据来重新训练 NUN,并迭代更新 NUN 和 REG 直至它们收敛.通

过该策略,将训练权重分配给各个训练样本,调节各个样本对网络训练的影响.随着模型的迭代更新,干净的数据将逐渐得到更高的训练权重,而有噪声的数据的权重将会被抑制,从而减少对网络训练的影响.本文的系统架构示意图如图 1.

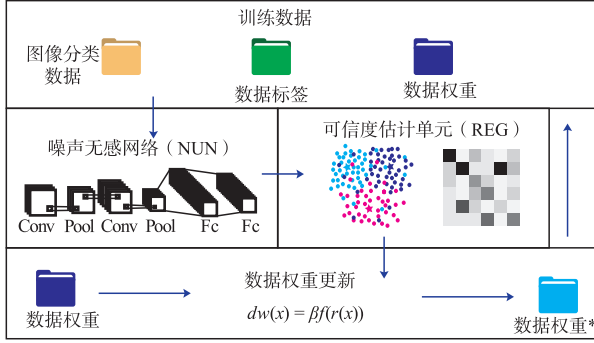


图1 基于可信度估计单元(REG)的噪声抑制网络体系结构

3.2 算法细节

算法 1 描述了本文的噪声抑制深度学习策略的整体过程.

算法 1 噪声抑制深度学习策略

输入:训练数据 X , 表示为 $\{x_i \in X | r_i, f_i\}$;

输出:REG 和准确的 NUN;

- 1: 使用整个训练数据集 X 训练噪声无感网络(NUN);
- 2: 使用 NUN 的特征图构造可信度估计单元(REG);
- 3: 根据式(1)评价可信度 $r(x)$;
- 4: 根据式(5)分配训练的权重 $dw(x)$;
- 5: 根据式(7)利用训练数据及其权重和损失函数 \tilde{l}_x 重新训练 NUN;
- 6: 迭代更新 NUN 和 REG, 直至收敛.

基于训练数据集训练 NUN. 通常输入 NUN 中的数据可以被抽象表示为 $\{x_i \in X | r_i, f_i, l_i, p_i\}$. 其中 r_i 和 f_i 表示样本 x_i 的原始信息和特征图信息, 并且 l_i 和 p_i 表示样本的标签和预测类别. 利用 NUN 中的特征图信息 f_i , 其中包含多个隐藏的线索. REG 的定义如下:

$$r(x) = \sum_{n=1}^N \alpha_n \cdot q(x | n) \quad (1)$$

类似高斯混合模型, 它的输出可以是一个正则化的数值. 其中 $r(x)$ 表示样本 x 的可信度; REG 模型中有 N 个子模块, 其中 N 个子模块分别对应 N 个聚类中心, 即样本标签, 每个模块都是对样本进行可信度评价; α_n 是模块 n 的权重, 并且 $q(x | n)$ 为模块 n 对样本 x 的计算输出.

本文使用 K-Means 模型作为 REG 的简单解释. 根据假设, 同一类别中某一特征空间中的特征应该高度聚合在一起. K-Means 算法被应用到 NUN 模型的特征图中. 直观地说, 一个样本应该尽可能接近一个它所属

的类别集群中心. 否则可以质疑其真实标签 l_i 的准确性. 定义 K-Means 模型评估以及距离定义如下:

$$q(x | \text{K-Means}) = \sum_{d=1}^D \sum_{k=1}^{k_d} g(\max(0, d_{c1} - d_{c2})) \quad (2)$$

$$d_{c1} = \sqrt{(x_f - x_d)^2 + (y_f - y_d)^2} \quad (3)$$

$$d_{c2} = \sqrt{(x_d - x_k)^2 + (y_d - y_k)^2} \quad (4)$$

其中 D 表示分类类别数量. k_d 表示类别 d 中的 K-Means 聚类中心. d_{c1} 表示特征映射 f 到标签类别中心 d 的距离, d_{c2} 表示目标类 d 到 K-Means 中心 k 的距离. 利用规则函数 $g(x)$ 对更接近其他类别的异常样本进行计数, 可以使用硬计数, 若 x 为真, 则 $g(x) = 1$, 否则为 0; 也可以选择软计数方法 $g(x) = x$. 本文选用软计数方案.

接下来, 通过每个样本的可信度来给出每个样本的训练权重,

$$dw(x) = \beta f(r(x)) \quad (5)$$

式中 $dw(x)$ 表示样本 x 的权重. β 是一个控制 REG 模型整体影响的因子, f 是确定 $r(x)$ 即可信度的某种函数. 可以简化为

$$f(r(x)) = \begin{cases} 1, & r(x) \geq 1 \\ r(x), & 0 < r(x) < 1 \\ 0, & r(x) \leq 0 \end{cases} \quad (6)$$

在 REG 的辅助下, 可以为每个训练样本分配其训练权重, 使用拥有权重的训练数据重新对 NUN 进行迭代训练. 损失函数修改为

$$\tilde{l}_x(y, z) = -dw(x) \cdot \log\left(\frac{e^z}{\sum_{j=1}^m e^{z_j}}\right) \quad (7)$$

其中 y 表示真实标签, z 表示 softmax 层的向量. NUN 和 REG 被迭代更新, 并且它们是彼此交互更新的过程.

4 实验

本节将详细介绍在不同数据集上进行的实验. 总体来说, 实验分为三个步骤, 首先在标准数据集中人为地制造噪声数据, 使用一些代表性的深度学习方法构造 NUN, 比较 REG 对 NUN 性能的影响; 其次, 使用存在标签噪声的情感分类数据集来模拟噪声抑制深度学习策略在实际应用场景中作用; 最后将本文的方法与现有的一些噪声抑制策略进行比较.

为了验证本方法的有效性, 选取了多个标准数据集进行图像分类实验, 并与多个国际前沿的算法进行了比较. 同时, 在实验中为每个数据集进行了标签噪声的人为添加, 控制标签噪声的比例, 以进行更准确的定量分析.

4.1 实验数据集

在本文的实验中, 首先在三个常用图像数据集上对本文提出的策略进行评估: Cifar10^[28]、MNIST^[29] 和

AwA (Animal with Attributes)^[30]. 它们的容量、复杂性和内容各不相同.

Cifar10 数据集是图像分类数据集,由 60000 张 32×32 的彩色图像组成,分 10 个类,每个类 6000 张图像,训练图像 50000 张,测试图像 10000 张. 为了在实验中产生噪声,本文将其重构为 60000 张原始 jpg 图像,使其可以自由地改变噪声率.

MNIST 数据库是一个手写数字的图像数据库. 每个图像大小为 28×28 ,代表 0 到 9 的一位数字. 它有一个包含 60000 个例子的训练集,以及一个包含 10000 个例子的测试集.

AwA 数据集包含 50 个动物类别的 30475 张图像,该数据集中每个类别的图像都含有 85 种语义属性值. 随机选取 80% 的数据作为训练集,剩余 20% 作为测试集.

4.2 实现细节

噪声数据产生. 记训练数据集中样本总数为 M ,需要制作的标签噪声样本的数量为 m ,噪声率为 nr ,则 $nr = m/M$,即噪声率表示标签噪声样本在总数据集中的比例. 根据所设定的噪声率以及训练集样本总数随机挑选出 m 个样本,对其标签进行人为的随机重置操作,即可得到所设定的噪声率的标签噪声数据.

训练噪声无感网络. 本文挑选了 AlexNet^[31]、LeNet^[32]、DenseNet^[33],将含有噪声的训练数据输入到上述深度学习网络中对模型参数进行重新训练,得到初期 NUN. 然后利用网络中的全连接层生成多个特征映射,建立了 REG.

可信度估计单元. REG 是噪声抑制模块的核心组成部分. 本文尝试了各种方案来实现这个模型. 将 REG 组件集成到现有的深度学习模型中,REG 与现有的深度学习模型同时训练,如式(1)所示. 对于 REG 中的参数,根据经验指导测试了几个案例,并采用了最好的一个.

训练权重. 利用 REG 可以对每个样本进行可信度评价,进而确定训练阶段样本的训练权重. 本文简单地将每个样本的可信度赋给权重因子 dw .

采用下列指标对可信度判断规则和训练权重因子进行评价. 使用 ncr 和 nhp 来调节阈值.

$$ncr = \frac{\text{发现的真实噪声数量}}{\text{真实噪声的总量}} \quad (8)$$

$$nhp = \frac{\text{发现的真实噪声数量}}{\text{识别全部噪声数量}} \quad (9)$$

其中 ncr 表示噪声覆盖率, nhp 表示噪声命中精度. 较高的 ncr 值意味着会发现更多的噪声数据,但可能会导致较低的 nhp 值,这意味着会意外地删除许多正实例. 式(8)和式(9)没有用于噪声抑制策略,本文只是用它们

来检验 REG 的效果.

网络参数. Cifar10 数据集使用 AlexNet 方法时,学习率设置为 10^{-4} ,权重衰减设置为 5×10^{-5} ,设置了最大迭代设置为 $30k$ 和步长设置为 $10k$. 对于 DenseNet 上的 Cifar10,训练 $230k$ 的迭代来初始化 DenseNet 模型,并添加 REG 来迭代地训练额外的 $60k$ 迭代. MNIST 和 AwA 数据集使用相同参数设置. 本文不使用数据扩充方法和验证集.

4.3 实验结果

本文为数据集设置不同的标签噪声率,记录模型分类准确率的变化,并在完全相同的模型上引入本文提出的学习策略,将现有的深度学习模型与 REG 模型相结合,观察分类准确率提升的效果. 实验结果见图 2.

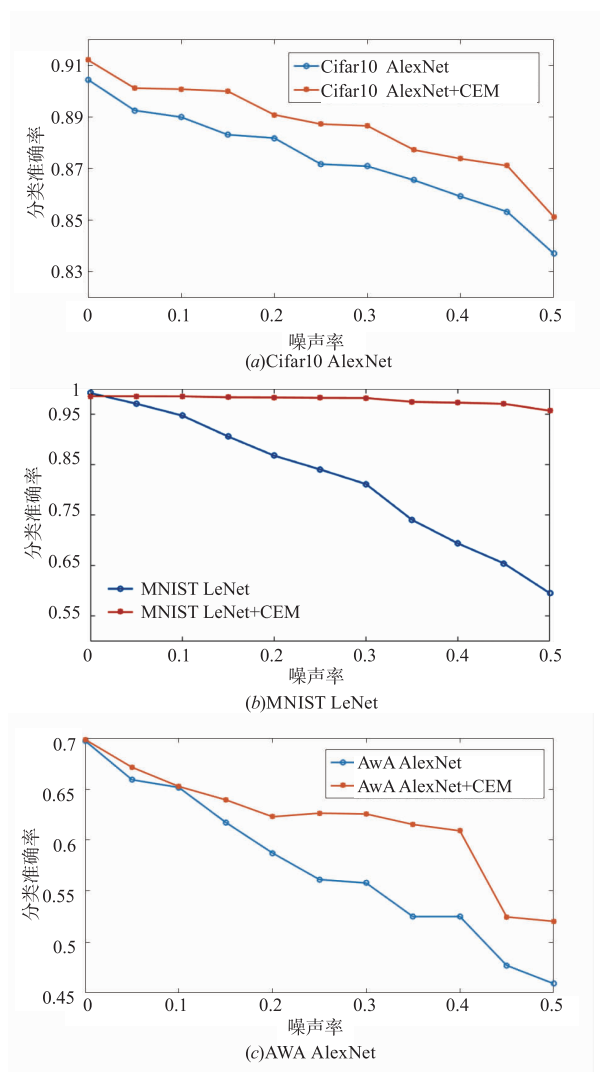


图2 用作NUN的基本模型与结合REG模型分类准确率

在 Cifar10 数据集中, AlexNet 模型的性能达到了 90.44%. 在 MNIST 数据集中, LeNet 模型达到 99.0% 的分类准确率. 对于 AwA 数据集, 本文使用 AlexNet 模型

对其进行简单评估。

从结果可以看出,随着噪声率的增加,现有的深度学习模型的性能急剧下降,这是因为噪声数据很容易造成深度学习模型过拟合。同时在大多数不同噪声率的数据下,本文的学习策略可以提高模型的性能,这意味着本文的策略是有效的。

此外,在不同的数据集中,噪声数据所带来的负面影响的程度是不同的。如图 2(a) 所示, Cifar10 数据集对噪声的鲁棒性更强,可能是由于其内容复杂多样。如图 2(b) 所示, MNIST 数据集受到噪声的影响较为严重,然而本文的学习策略对该数据集有较大的改进。如图 2(c) 所示, 本文所提出的方法在 AwA 数据集上也表现良好。

本文方法在复杂的深度神经网络中也能起到显著的效果。通过对 DenseNet 进行了优化,增加了一个全连接层来提供全局的特征信息,然后产生不同的噪声率来观察噪声的影响。实验结果见图 3。DenseNet* 表示 DenseNet 的优化版本, DenseNet* + REG 表示本文所提出的 REG 模型与 DenseNet* 模型相结合的方法。结果显示,虽然 DenseNet* 在 Cifar10 中达到了较高的准确率,但当数据有噪声时,其性能显著下降,当噪声率为 40% 时准确率下降了 18.26%。结合本文提出的 REG 模型,分类准确率达到 93.4%,证明了本文所提出的策略对于现有的复杂网络架构在噪声抑制和性能提升上有较好的效果。

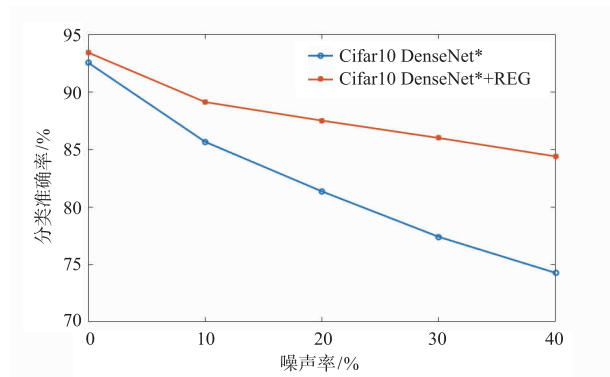


图3 DenseNet*与DenseNet*+REG不同噪声率下的性能比较

4.4 与现有方法进行比较

Scott^[25] 等人提出了一种基于一致性的噪声重构方法来处理弱监督深度学习中的噪声和不完全标记。他们在 MNIST 数据集上进行了一个实验,在本节中将进行比较。

根据 Scott 的实验设置,使用了标签噪声的固定排列,这意味着如果类别 c1 中的样本的标签是错误的,其标签类别将永远属于 c2。然而本文人工设置的噪声与 Scott 的不同。本文的标签噪声在每个类别中都是均匀

分布的,即在某个类别中,有噪声的标签可以随机分布。实验结果对比见图 4。由于噪声产生方案的不同, Scott 提出的方法在本文的噪声环境中表现的不是很好,因此 Scott 的方法在本文的环境中不能学习正确的映射矩阵。然而,本文的策略成功地对噪声数据进行了分类,提高了分类效率。

Scott 的方法和本文的策略在不同的情况下有各自的优势。本文的策略更适合随机分布的噪声,而随机分布的噪声在现实世界中更容易发生。

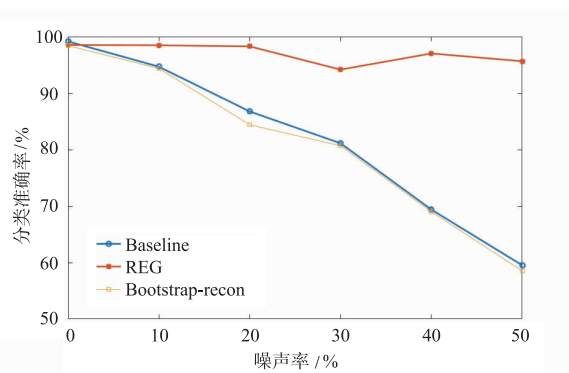


图4 比较bootstrap-recon和本文提出的噪声抑制深度学习策略在噪声随机分布的情况下,bootstrap-recon方法对分类准确率的提高很小,而本文提出的方法可以提高分类性能

4.5 实际应用场景中的结果

使用两个情感分类数据集来评估本文提出的噪声抑制深度学习策略在实际应用场景中的性能。第一个是 FI 数据集,它包含从 Flickr 和 Instagram 通过搜索相应标签的关键词^[34] 收集得到的 90000 个噪声图片。带有弱标签的图像是由通过资格测试选择出来的 225 位 Amazon Mechanical Turk (AMT) 工作人员进一步标记的。23308 张图片是从指定 5 名 AMT 工作人员那里获得至少 3 票被保留的数据。每个标签类别的图片数量至少有 1000 张。第二个是 ArtPhoto 数据集,它包含来自照片共享网站的 806 张艺术照^[35],其标签由上传照片的艺术家决定。

本文使用广泛使用的深度网络框架 AlexNet 作为框架,并使用分类准确率作为度量指标。参数设置与 4.2 节中的网络参数设置一致。

测试的比较结果如表 1 所示。可以观察到,本文的方法明显优于其他对比方法,证明了本文提出的噪声抑制模型在图像分类的有效性。靠近边界的例子是提供信息的,但是注意到,在实践中它们更有可能被错误标记,因为它们更加模糊。如果样本拥有正确的标签,它们可以获得区分两个类的重要先验知识。否则,更多的关注它们将会导致一个更糟糕的模型,因为模型在训练时被误导了。在本文中,对靠近边界的样本关注较少,因此其中的噪声对模型的影响较小。此外,这样做还可

以使模型更加关注类的一般知识,从而使它能够更好地概括测试数据.事实上,在实际应用中,当模型是由现实世界中收集的数据训练完成复杂任务时,几乎不可能确保训练集与 MNIST 一样干净.结果表明本文的方法在实践中表现良好.

表 1 情感分类数据集上的结果

方法	FI	ArtPhoto
You 等人 2016 ^[34]	0.583	0.737
Rao 等人 2016 ^[36]	0.652	0.834
Yang 等人 2018 ^[37]	0.676	0.788
AlexNet + REG	0.690	0.852

5 结论

大规模高质量的数据集对深度学习模型的训练效果起着至关重要的作用,但获取高质量的海量标签不仅费时费力,还十分昂贵.在实际应用场景中构建深度学习模型时,必须要考虑到标签噪声所带来的影响.本文提出了一种噪声抑制深度学习策略,该策略通过构建噪声无感网络(NUN)和可信度估计单元(REG)来应对训练数据集中含有标签噪声的情况.通过 REG 来评估每个样本的可信度用于调节训练权重,逐渐增强干净数据的影响,减弱可疑噪声数据的反馈.本文将该策略应用到现有的深度学习模型中,实验表明,大多数深度学习模型的性能随着噪声率的增加而下降,然而在结合了本文所提出的噪声抑制深度学习策略下,可以将深度学习模型的性能拉回到令人满意的程度.大量的实验证明了该策略的有效性.

参考文献

- [1] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE, 2016. 770 – 778.
- [2] Yang X, Molchanov P, Kautz J. Multilayer and multimodal fusion of deep neural networks for video classification [A]. Proceedings of the 24th ACM International Conference on Multimedia [C]. SA: ACM, 2016. 978 – 987.
- [3] 田艳玲, 张维桐, 张镏石, 等. 图像场景分类技术综述 [J]. 电子学报, 2019, 47(4): 915 – 926.
TIAN Yan-ling, ZHANG Wei-tong, ZHANG Qie-shi, et al. Review on image scene classification technology [J]. Acta Electronica Sinica, 2019, 47(4): 915 – 926. (in Chinese)
- [4] 余游, 冯林, 王格格, 等. 一种基于伪标签的半监督少样本学习模型 [J]. 电子学报, 2019, 47(11): 2284 – 2291.
YU You, FENG Lin, WANG Ge-ge, et al. A few-shot learning model based on semi-supervised with pseudo label [J]. Acta Electronica Sinica, 2019, 47(11): 2284 – 2291. (in Chinese)
- [5] Zhu X, Wu X. Class noise vs attribute noise: A quantitative study [J]. Artificial Intelligence Review, 2004, 22(3): 177 – 210.
- [6] Xiao T, Xia T, Yang Y, et al. Learning from massive noisy labeled data for image classification [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE, 2015. 2691 – 2699.
- [7] Frénay B, Verleysen M. Classification in the presence of label noise: a survey [J]. IEEE Transactions on Neural Networks and Learning Systems, 2013, 25(5): 845 – 869.
- [8] Manwani N, Sastry P S. Noise tolerance under risk minimization [J]. IEEE Transactions on Cybernetics, 2013, 43(3): 1146 – 1151.
- [9] Beigman E, Klebanov B B. Learning with annotation noise [A]. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP [C]. Association for Computational Linguistics, 2009. 280 – 287.
- [10] Teng C M. Evaluating noise correction [A]. Pacific Rim International Conference on Artificial Intelligence [C]. Berlin, Heidelberg: Springer, 2000. 188 – 198.
- [11] Folleco A, Khoshgoftaar T M, Van Hulse J, et al. Software quality modeling: The impact of class noise on the random forest classifier [A]. IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence) [C]. USA: IEEE, 2008. 3853 – 3859.
- [12] Bartlett P L, Jordan M I, McAuliffe J D. Convexity, classification, and risk bounds [J]. Journal of the American Statistical Association, 2006, 101(473): 138 – 156.
- [13] Misra I, Lawrence Zitnick C, Mitchell M, et al. Seeing through the human reporting bias: Visual classifiers from noisy human-centric labels [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE, 2016. 2930 – 2939.
- [14] Zhu X J. Semi-supervised learning literature survey [R]. University of Wisconsin-Madison Department of Computer Sciences, 2005.
- [15] Veit A, Alldrin N, Chechik G, et al. Learning from noisy large-scale datasets with minimal supervision [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE, 2017. 839 – 847.
- [16] Chen X, Shrivastava A, Gupta A. Neil: Extracting visual knowledge from web data [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. USA: IEEE, 2013. 1409 – 1416.
- [17] Fergus R, Weiss Y, Torralba A. Semi-supervised learning

- in gigantic image collections [A]. Advances in Neural Information Processing Systems [C]. Vancouver, BC, Canada, 2009. 522 – 530.
- [18] Miranda A L B, Garcia L P F, Carvalho A C, et al. Use of classification algorithms in noise detection and elimination [A]. International Conference on Hybrid Artificial Intelligence Systems [C]. Berlin, Heidelberg: Springer, 2009. 417 – 424.
- [19] Guan D, Yuan W, Lee Y K, et al. Identifying mislabeled training data with the aid of unlabeled data [J]. Applied Intelligence, 2011, 35(3): 345 – 358.
- [20] Natarajan N, Dhillon I S, Ravikumar P K, et al. Learning with noisy labels [A]. Advances in Neural Information Processing Systems [C]. Nevada, 2013. 1196 – 1204.
- [21] Sukhbaatar S, Bruna J, Paluri M, et al. Training convolutional networks with noisy labels [J]. arXiv Preprint, 2014, arXiv:1406. 2080.
- [22] Matic N, Guyon I, Bottou L, et al. Computer aided cleaning of large databases for character recognition [A]. Proceedings of the 11th IAPR International Conference on Pattern Recognition (Vol. II). Conference B: Pattern Recognition Methodology and Systems [C]. USA: IEEE, 1992. 330 – 333.
- [23] Wu P, Hoi S C H, Xia H, et al. Online multimodal deep similarity learning with application to image retrieval [A]. Proceedings of the 21st ACM International Conference on Multimedia [C]. USA: ACM, 2013. 153 – 162.
- [24] Wu Z, Jiang Y G, Wang J, et al. Exploring inter-feature and inter-class relationships with deep neural networks for video classification [A]. Proceedings of the 22nd ACM International Conference on Multimedia [C]. USA: ACM, 2014. 167 – 176.
- [25] Reed S, Lee H, Anguelov D, et al. Training deep neural networks on noisy labels with bootstrapping [J]. arXiv Preprint, 2014, arXiv: 1412. 6596.
- [26] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition [A]. European Conference on Computer Vision [C]. Cham: Springer, 2016. 499 – 515.
- [27] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE, 2015. 815 – 823.
- [28] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images [A]. Computer Science [C]. MIT, 2009. 1 – 14.
- [29] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278 – 2324.
- [30] Lampert C H, Nickisch H, Harmeling S. Attribute-based classification for zero-shot visual object categorization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(3): 453 – 465.
- [31] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [A]. Advances in Neural Information Processing Systems [C]. USA, 2012. 1097 – 1105.
- [32] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278 – 2324.
- [33] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE, 2017. 4700 – 4708.
- [34] You Q, Luo J, Jin H, et al. Building a large scale dataset for image emotion recognition: The fine print and the benchmark [A]. Thirtieth AAAI Conference on Artificial Intelligence [C]. USA: AAAI, 2016. 1 – 7.
- [35] Machajdik J, Hanbury A. Affective image classification using features inspired by psychology and art theory [A]. Proceedings of the 18th ACM International Conference on Multimedia [C]. USA: ACM, 2010. 83 – 92.
- [36] Rao T, Xu M, Liu H, et al. Multi-scale blocks based image emotion classification using multiple instance learning [A]. IEEE International Conference on Image Processing (ICIP) [C]. USA: IEEE, 2016. 634 – 638.
- [37] Yang J, She D, Lai Y K, et al. Retrieving and classifying affective images via deep metric learning [A]. Thirty-Second AAAI Conference on Artificial Intelligence [C]. USA: AAAI, 2018. 1 – 14.

作者简介



邵 航 男, 1986 年 9 月出生, 浙江江山人. 2011 年毕业于清华大学自动化系, 获工学硕士学位. 主要研究方向为人工智能与计算成像.
E-mail: shaohang@tsinghua.edu.cn



黄海亮 男, 1995 年 6 月出生, 江苏如皋人. 2019 年毕业于杭州电子科技大学自动化学院, 获硕士学位. 主要研究方向为人工智能与计算成像.
E-mail: huanghailiang@zfti.org