计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(10):275−283 [doi: 10.15888/j.cnki.csa.009270]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于预测对抗网络的图像二分类模型①

余筝韵[1], 李　春[2]

[1](哈尔滨工业大学 (深圳) 理学院, 深圳 518055)
[2](深圳北理莫斯科大学 计算数学与控制联合研究中心, 深圳 518172)
通信作者: 余筝韵, E-mail: 2398939452@qq.com

摘　要: 正未标记学习仅使用无标签样本和正样本训练一个二分类器, 而生成式对抗网络 (generative adversarial networks, GAN) 中通过对抗性训练得到一个图像生成器. 为将 GAN 的对抗训练方法迁移到正未标记学习中以提升正未标记学习的效果, 可将 GAN 中的生成器替换为分类器 C, 在无标签数据集中挑选样本以欺骗判别器 D, 对 C 与 D 进行迭代优化. 本文提出基于以 Jensen-Shannon 散度 (JS 散度) 为目标函数的 JS-PAN 模型. 最后, 结合数据分布特点及现状需求, 说明了 PAN 模型在医疗诊断图像二分类应用的合理性及高性能. 在 MNIST, CIFAR-10 数据集上的实验结果显示: KL-PAN 模型与同类正未标记学习模型对比有更高的精确度 (ACC) 及 F1-score; 对称化改进后, JS-PAN 模型在两个指标上均有所提升, 因此 JS-PAN 模型的提出更具有合理性. 在 Med-MNIST 的 3 个子图像数据集上的实验显示: KL-PAN 模型与 4 个 benchmark 有监督模型有几乎相同的 ACC, JS-PAN 也有更高表现. 因此, 综合 PAN 模型的出色分类效果及医疗诊断数据的分布特征, PAN 作为半监督学习方法可获得更快、更好的效果, 在医学图像的二分类的任务上具有更高的性能.

关键词: 预测对抗网络; 正未标记学习; 医学图像分类; 对抗性训练

引用格式: 余筝韵,李春.基于预测对抗网络的图像二分类模型.计算机系统应用,2023,32(10):275–283. http://www.c-s-a.org.cn/1003-3254/9270.html

Medical Image Classification Based on Predictive Adversarial Networks

YU Zheng-Yun[1], LI Chun[2]

[1](School of Science, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China)
[2](Joint Research Center for Computational Mathematics and Control, Shenzhen MSU-BIT University, Shenzhen 518172, China)

Abstract: Positive-unlabeled learning (PU learning) only uses unlabeled samples and positive samples to train a binary classifier, while generative adversarial networks (GANs) obtain an image generator through adversarial training. In order to transfer the adversarial training method of GANs to PU learning for higher PU learning performance, the generator in GANs can be replaced with a classifier C, which selects samples in the unlabeled dataset to deceive the discriminator D and optimize C and D iteratively. This study proposes the JS-PAN model, which uses the Jensen-Shannon divergence (JS-divergence) as the objective function. Finally, according to the characteristics of data distribution and current needs, the rationality and high performance of the PAN model applied in the binary classification of medical diagnostic images are explained. Experiments on MNIST and CIFAR-10 datasets show that the KL-PAN model has higher accuracy (ACC) and F1-score than the similar PU learning models, and the JS-PAN model has higher performance in terms of two indicators after symmetric improvement, so the JS-PAN model is more reasonable. Experiments on three image subdatasets of Med-MNIST show that the KL-PAN model has almost the same ACC as the four benchmark supervised models, and JS-PAN has higher performance. Therefore, in view of both the excellent classification performance of the PAN model and the distribution characteristics of medical diagnostic data, PAN, as a semi-supervised learning method, can achieve faster and

① 基金项目: 广东省基础与应用基础研究基金 (2021A1515220073)
收稿时间: 2023-02-28; 修改时间: 2023-03-30, 2023-04-07, 2023-04-12, 2023-05-11; 采用时间: 2023-05-17; csa 在线出版时间: 2023-08-21
CNKI 网络首发时间: 2023-08-22

better results and thus show higher performance in the task of binary classification of medical images.

Key words: predictive adversarial networks (PAN); positive-unlabeled learning (PU learning); medical image classification; adversarial training

目前, 深度学习在医学图像分析任务上取得很大成功[1,2], 此类模型的学习效果在很大程度上依赖于专家给出的人工诊断以及大量训练图像的可用性. 半监督学习是利用无标签数据和一部分有标签数据进行训练, PU learning 是半监督学习的一个重要分支. 正未标记学习中, 唯一可用的有标签数据是正样本, 它利用无标签数据和正样本进行训练, 学习出一个二分类器. 它不仅符合此类诊断数据的分布特征, 更好地利用无标签数据进行神经网络的训练, 而且能有效解决高质量人工标注难以获得、数量较少的问题. 同时, 近年来对抗性训练被广泛应用在深度学习尤其是半监督学习中. 将对抗性训练思想应用在医学诊断图像的深度学习模型中可以提升模型鲁棒性及稳定性, 所以基于对抗网络的医学图像分类模型也受到广泛关注[3–5].

1 模型相关工作

1.1 正未标记学习模型

正未标记学习 (PU 学习) 的研究始于 21 世纪初. 由于其在生活中的应用广泛, 近年来人们对 PU 学习的兴趣激增. Denis[6] 对 PU 学习进行理论分析; Liu 等人[7] 对 PU 数据集样本的复杂性进行分析. 这些 PU 学习的早期方法由两步骤组成: 步骤 1, 从无标签数据集中查找一些极有可能为负类的样本. 步骤 2, 使用正样本和步骤 1 的可能的负样本作为有标签的数据, 再利用剩余的未标记集来构建单独的分类器. 但是, 此类两步过程容易受到错误的负标签的影响, 因为当用错误的负类样本训练新模型时, 错误会在后续迭代中传播. 以上的 PU 学习方法都没有像 PAN 模型一样, 将对抗性思想运用在 PU 学习中.

1.2 生成式对抗网络

生成式对抗网络 (generative adversarial networks, GAN)[8] 起源于 Goodfellow 等人在 2014 年研究的噪声对比估计, 它使用与 GAN 相同的损失函数. 对抗性学习思想除了生成建模还有其他用途, 可以应用于神经网络以外的模型. 在控制论中, 基于神经网络的对抗性学习被来训练博弈思想上的鲁棒性控制器, 对抗性学习方法体现在最小化策略 (控制器) 和最大化策略 (干扰) 之间的交替迭代. 2017 年, GAN 被用于图像增强领域[9], 其用于提高图像的逼真性而不是清晰度 (像素精度), 从而在高放大倍率下产生更高质量的图像. 同年第 1 批图像被生成了, 并于 2018 年 2 月在大皇宫展出. 在 2019 年, StyleGAN[10] 可以更加灵活地生成图像. 2020 年 5 月, Nvidia 研究人员研究出了一个 AI 系统 (称为“GameGAN”) [11], 它只需通过观看游戏来重现 Pac-Man 游戏.

Hou 等人[12] 运用 GAN 生成正类和负类样本, 然后将它们作为有标签数据, 构建单独分类模型, 训练得到最终分类器. Chiaroni 等人[13] 提出了一种使用 GAN 生成负样本的方法. 这两篇论文采用的数据类型均为图像. 使用 GAN 生成文本类型数据和其他形式的数据更困难. 这些基于 GAN 的分类模型大多单独使用 GAN 作为数据生成器, 再利用其单独训练一个分类器. 这些模型都没有将对抗性学习思想直接运用在对最终的分类器的训练中, 分类器的训练数据输入也并不是只有正样本和无标签数据集.

1.3 近年来正未标记学习的发展

Xu 等人[14] 都使用传统的基于决策边际 (Margin) 的方法训练分类器, 但模型结果较差, 尤其在图像分类中; Bekker 等人[15] 研究 PU 学习的随机假设; 这些模型的缺点在于, 它们都需要一个先验概率. 研究表明, 如果先验概率的估计值不对, 结果可能会很差. 基于以上分析, Hu 等人[16] 提出的基于对抗性训练网络的分类模型 (PAN), 在该模型中, 作者们将对抗性思想引入 PU 学习中, 其分类效果较好, 敏感性及鲁棒性较高, 并且不需要先验概率, 可运用在多种类型的数据上.

2 PAN 模型背景

2.1 正未标记学习与生成式对抗网络

PU learning 是半监督学习的一个重要分支. 它利用无标签数据和正样本进行训练, 学习出一个二分类器. GAN 是一种深度学习中的无监督学习模型, 因其善于处理复杂分布的数据而广受关注. 此模型框架中

含有 (至少) 两个模块: 生成器 (generator) 和判别器 (discriminator), 以下简称为 G 和 D. 生成器 G 生成新的样本数据, 鉴别器 D 评估它们的真实性, 即决定它每个生成的数据样本是否来自有标记的实际训练集. 生成式对抗网络就是通过 G 和 D 的互相博弈, 迭代优化模型, 最终学习出一个用于生成图像等数据的一个生成器. 其中, 对抗性神经网络的流程图如图 1 所示.

以图像数据为例, 下面具体阐述 GAN 模型的原理. 首先, 本文有一组真实的图片作为训练集中有标签的正样本. 生成器 G 先接收一组随机噪声 Z, 把真实的图片加上这个噪声, 生成假的图片, 记作 $G(z)$. 因为本文只向网络输入噪声 Z, 而不给它们添加任何标签 (或期望的输出), 所以这种训练过程是无监督学习. D 是一个判别网络, 它的输入参数是真实图片 x 和假数据 $G(z)$, 分别输出 $D(x)$ 和 $D(G(z))$, 代表 x 和 $G(z)$ 被 D 鉴别为正样本的概率. 若概率为 1, 就代表 100% 是真实的图片, 若输出为 0, 就代表不可能是真实的图片.

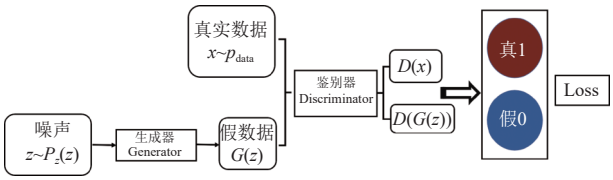


图 1 对抗性神经网络的流程图

在 GAN 中, G 的功能是产生大量“假图片”来欺骗 D, 而 D 的功能是尽力将生成的假数据 (负样本) 与真实的有标签的训练数据 (正样本) 区分开来. 所谓对抗性学习就是一个 G 和 D 动态博弈的过程. 最后博弈的结果是: 在最理想的状态下, G 可以生成“以假乱真”的图片 $G(z)$, 使得 D 难以判定 G 生成的图片究竟是不是真实的, 因此 $D(G(z))$=0.5. 通过轮番优化 G 和 D, 从而可得一个生成数据的模型 G, 训练的目的就达成了. GAN 的目标函数如下:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim P_z(z)}[\log(1-D(G(z)))] \quad (1)$$

首先, 设 x 代表有标签的训练集样本, 其中 z 代表传入生成器 G 的随机噪声. $p_{data}(x)$代表真实数据的数据生成分布, $P_z(z)$代表生成器生成的数据分布. 生成器 G 是一个网络, 生成的数据为$G(z)$, 在鉴别器 D 中得到的正概率为$D(G(z))$, 则被鉴别器识别出假数据的概率即为$1-D(G(z))$. 鉴别器 D 是一个二分类器, 传入了真实数据 x 及假数据$G(z)$, 对这两种数据 D 输出它们是正样本的概率分别为$D(x)$和$D(G(z))$, 最终将概率$D(x)$和$D(G(z))$与 0.5 相比较, 分别为其分配一个标签 (y=0 表示 D 认为这是 G 生成的假数据, y=1 代表是训练集的正样本).

(1) D 的损失函数

D 的损失函数可以用常见的二分类器损失函数——二元交叉熵损失函数来定义. 其中 k 为输入数据. 由于当 k 为真实数据时 y=1, 当 k 为生成器 G 生成的数据时 y=0, 于是公式简化成:

$$D_{loss} = -\frac{1}{n}\cdot\sum_{i}^{n}\log(D(x_i)) + \log(1-D(G(z))) \quad (2)$$

(2) GAN 的目标函数

对 D 而言, 目的是将损失函数最小化D_{loss}, 也就是最大化$-D_{loss}$. 于是, 其目标函数为:

$$\max_D V(G,D) = E_{x\sim p_{data}(x)}[\log(D(x))] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (3)$$

对 G 而言, 它的目的是缩小真实数据和产生的假数据的差异. 对所有样本进行总体考虑, 也就是最小化 $p_{data}(x)$和$P_G(z)$之间的真假数据分布差异, 这恰恰与 D 的目标相反, 于是优化 G 就是求解上述$V(G,Z)$的最小值. 于是整体优化函数可以定义为:

$$\min_G \max_D \mathrm{V}(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + \mathrm{E}_{z\sim P_z(z)}[\log(1-D(G(z)))] \quad (4)$$

2.2 适应性对抗网络 (Adapted-GAN)

本文将对抗性思想直接运用在 PU 学习中得到的初级的 Adapted-GAN (A-GAN). 由第 2.1 节, PU learning 与 GAN 两者都是利用一部分有标签数据和一部分无标签数据的半监督模型, 且最终都训练出一个二分类器. 将 GAN 的对抗性思想直接引进正未标记学习里, 这个新型的正未标记学习模型可以理解为: 模型在未标注的数据集里寻找被分为正类的概率大的样本, 去欺骗鉴别器 D, 让鉴别器 D 判别数据是有标签的正样本还是未标注的假样本. 具体对比两者的差别, 不同点是 PU learning 里面的假数据是从无标签的数据集里面找的; 而 GAN 的假数据是自身通过生成器生成的.

因此, 将对抗性思想直接运用在正未标记学习中, 可将 GAN 中的生成器 G 换成一个分类器 C, 当于将 GAN 改造成一个对抗性 PU 学习模型, 这就得到了最初级的 Adapted-GAN (A-GAN). 其目标函数如下:

$$\min_C \max_D V(D,C) = \mathbb{E}_{x^p \sim P^p(x^p)}[\log D(x^p)] + \mathbb{E}_{x^s = \arg_{x^u \sim P^u(x^u)} C(x^u)=1}[\log(1-D(x^s))] \tag{5}$$

其中, X^p代表有标签的训练集样本, $D(X^p)$代表鉴别器对正样本的判别概率; X^s代表无标签数据集中被分类器 C 选出的数据样本, $D(X^s)$代表鉴别器 D 将它们归到正类的概率. $P^p(X^p)$代表真实数据的数据生成分布, $P^u(X^u)$代表未标记集的数据分布.

和 GAN 模型的公式推导原理相似: 对于 D 而言, 要尽量把标记的正样本判别为正类, 最大化$D(X^p)=1$的概率, 同时要尽量把 C 从未标记集中挑出的假数据X^s判别为负类, 最小化$D(X^s)=0$的概率, 也就是最大化$1-D(X^s)=1$的概率. 对于 C 而言, 它无法影响第 1 项(即 D 对标记的正样本判别为正类的概率), 而对第 2 项的影响与 D 的目标恰恰相反, 由相似的推导, 轮番优化 D 和 C, 得出优化函数如上.

值得注意的是, 当本文在训练 C 时, 相当于保持鉴别器 D 的各参数不变 (固定 D 的分类标准). 首先, 此时函数中第 2 项 (即鉴别器 D 将假数据辨认出来的概率) 对于固定的无标签数据集 U 而言, 它的分布与 U 中数据的分布是有一一映射关系. 而 U 中的数据分布为离散型, 则函数中第 2 项的数据分布在训练 C 时是离散的; 其次, 对于 C 而言, 第 1 项是一个固定的无关量, 优化 C 只与第 2 项有关. 于是, 对于此离散型的优化函数, 本文采用强化学习中的策略梯度算法来训练 C. 在强化学习的训练中, 这一项被视为优化 C 的奖励. 此种策略梯度算法在后面提出的 PAN 模型中也有运用.

3 PAN 模型

本文对 Adapted-GAN (A-GAN) 模型在对抗性思想上的运用进行改进, 研究了一个更高级的模型——预测性对抗网络 (PAN). PAN 同样构建 PU 分类器 C 替代 GAN 中生成器 G, 但它采用了一种新的方式来运用对抗性学习思想.

3.1 PAN 模型提出

以上 GAN 及 A-GAN 两个模型均用二分类模型的交叉熵函数来定义 D 的损失函数, 再基于此考虑 C 的功能, 提出整体优化函数. 这个优化函数本质就是通过迭代更新 C 和 D, 优化鉴别器 D 分类正确的概率. 而这两个模型的对抗性思想运用在 C 和 D 的博弈中, 直接优化鉴别器 D 分类的总体准确度 (ACC). 本文研究的 PAN 模型则做出以下改进: 它不在 D 分类的总体准确度上使用对抗性学习思想, 而是在 D 和 C 的整体分布上使用对抗性学习思想.

对于单个无标签数据集 U 中的样本 x, 若 C 选中则 $C(X)=1$, 否则$C(X)=0$, 若 D 判断为正类则$D(X)=1$, 否则$D(X)=0$. 对于 U 中的所有样本点, 也就是总体样本 X, 将 C 和 D 分别作用后得到两个矩阵, 记为$C(X)$和$D(X)$.

对 D 而言, D 试图区分 C 选中的负样本与 P 中的正样本, 也就是增大$C(X)$和$D(X)$的数据分布差异. 对 C 而言, C 又试图在无标签数据集 U 中找到最可能分为正类的无标签样本 (负样本) 去迷惑 D, 也就是缩小$C(X)$和$D(X)$的数据分布差异. 于是对抗性体现在: C 和 D 分别希望减小、增大$D(X)$和$C(X)$的数据分布差异. 于是本文可以将这个“数据分布差异”作为优化函数, 用对抗学习的思想轮番对其求最大最小值. 因此, PAN 首先使用$C(X)$来拟合 D 在总体样本上的概率分布, 然后根据衡量分布差异的目标函数, 轮流对 D 和 C 进行参数的更新, 最终训练出一个而分类器 D.

3.2 PAN 模型中 C 的分类原理

PAN 模型计算 D 和 C 总体概率分布差异, 通过迭代优化 D 和 C 进行对抗性训练, 学习出一个二分类器 D. 由于 PAN 以 D 作为二分类器的模型输出, 它不仅希望 C 选出的无标签样本很难用鉴别器 D 来区分, 以达到迷惑鉴别器 D 的效果, 同时也希望剩余的 U 中样本点很容易被鉴别器 D 区分. 于是, C 在训练过程中, 不仅要给 U 中 D 难以区分的样本赋予高概率, 而且还要给 D 容易区分的数据赋予低概率.

针对以上目标, C 从无标签数据集 U 中选取假的“正类样本”的原理可以表述如下: C 是一个二分类器, 功能是找到一个超平面将数据分成正负两类, 两类样本点的距离就叫做边际 (Margin). C 的训练过程就是在不断调整这个超平面的参数, 最终训练目标是找到使得 Margin 最大的超平面. 这与支持向量机 (support vector machine, SVM) 的二分类原理相同. 总的来说, PAN 追求的是 C 分类整体的精准性, 而不仅只追求把正样本识别出来, 而忽略对负样本输出概率的考虑 (后者容易导致高精确度, 低召回率的现象).

3.3 PAN 模型中衡量分布差异的函数

在 PAN 中, 本文先利用 KL 散度度量分类器 C 和鉴别器 D 的概率分布差异. 随后研究 KL 散度的对称化, 得到 JS 散度作为差异衡量函数.

(1) Kullback-Leibler 散度 (KL 散度)

对连续型概率分布, KL 散度的定义如下:

$$KL(p\|q) = -\int_{\mathbb{R}} p(x)\ln\frac{q(x)}{p(x)}dx \tag{6}$$

对离散型概率分布, KL 散度的定义如下:

$$KL(P\|Q) = \Sigma_i P_P(x_i)\log\left(\frac{P_P(x_i)}{P_Q(x_i)}\right) \tag{7}$$

(2) Jensen-Shannon 散度 (JS 散度)

采用 KL 散度衡量 D 和 C 概率分布差异可以最大限度地减少两个概率分布之间的信息损失, 但是 KL 散度不符合三角不等式及对称性, 所以不符合严格的“距离”的定义. 因此, 我们对 KL 散度进行对称化, 设 $M=(P+Q)/2$, 从而可得到如下 JS 散度:

$$JS(P\|Q) = \frac{1}{2}KL(P\|M) + \frac{1}{2}KL(Q\|M) \tag{8}$$

3.4 PAN 模型的目标函数

基于 KL 散度及 JS 散度的 PAN 模型的目标函数分别如式 (6) 和式 (7):

$$\min_C\max_D V(D,C) = \underbrace{-\sum_{i=1}^{n} KL\left(P_i^{pu}\|D_i^{pu}\right)}_{\text{I}} + \lambda\left(\underbrace{\sum_{i=1}^{n_0} KL\left(D_i^{u}\|C_i^{u}\right)}_{\text{II}} - \underbrace{\sum_{i=1}^{n_0} KL\left(D_i^{u}\|\hat{C}_i^{u}\right)}_{\text{III}}\right) \tag{9}$$

$$\min_C\max_D V(D,C) = \underbrace{-\sum_{i=1}^{n} JS\left(P_i^{pu}\|D_i^{pu}\right)}_{\text{I}} + \lambda\left(\underbrace{\sum_{i=1}^{n_0} JS\left(D_i^{u}\|C_i^{u}\right)}_{\text{II}} - \underbrace{\sum_{i=1}^{n_0} JS\left(D_i^{u}\|\hat{C}_i^{u}\right)}_{\text{III}}\right) \tag{10}$$

本文规定这个二分类模型中, 正类概率为 1, 负类为 0, 本文默认无标签数据集中样本全为负类 (即 U 中样本真实标签均为 0). 样本通过 D 和 C 后输出结果均满足 0-1 分布. 其中, P_i^{pu} 是给定 PU 数据X^{pu}(包括正X^p和未标记X^u数据) 中第 i 个样本的概率分布, D_i^{pu} 是鉴别器 D 对这个样本的分类结果的概率分布. n和n_0分别是 PU 数据集中无标签数据集 U 中的样本的总数. $\hat{C}_i^u$ 表示C_i^u的相反分布.

目标函数中, 第 I 项代表 PU 数据集中第 i 个样本的真实概率分布与鉴别器输出结果的概率分布差异, 由于 PU 数据集中的样本满足 0-1 分布, 是离散型随机变量, 本文对n个样本求和得到 PU 中总体样本的概率分布差异. 注意到这一项与 C 无关, 只有在优化 D 的时候起作用. 对于 D 而言, 要最小化这个和函数, 也就是最大化第 1 项, 这一项的功能是帮助 D 从 PU 数据集里有标签的正样本中学习判别规则. 在训练初期 D 的判别能力较弱, 这一项对 D 的学习就尤其重要; 第 II 项代表无标记数据集 U 中第 i 个样本通过分类器 C 与鉴别器 D 后输出结果的概率分布差异, 由于 U 数据集中的样本满足 0-1 分布, 是离散型随机变量, 本文对n_0个样本求和得到 U 中总体样本通过 C 和 D 分类后的概率分布差异. 优化 D 时, D 试图检测假样本, 要最大化与 C 的分布差异; 优化 C 时, 为了实现了用 C 欺骗 D 的目标, 要最小化它, 于是本文得到了优化函数的第 2 项. 这两个项已经可以执行 PAN 的功能; 对于第 III 项, 如上文 PAN 中 C 的分类原理所提, PAN 模型可以同时考虑和优化无标签数据集中的正样本和负样本的输出结果. 但是如果优化函数只有前两项, 则第 2 项中 C 和 D 的正负样本将在训练过程中产生不对称梯度, 可能导致正样本和负样本点间的不平衡训练, 因此导致高精度和低召回率. 为此, 本文需要找到一个对称的函数, 使得对 C 优化时, C 分类为正样本时不存在梯度, 而分类为负样本时存在与第 2 项相同的梯度. 于是, 根据对称性构造出第 III 项, 此项作为梯度平衡项. 最后, 再在第 2 和第 3 项整体前面加上超参数λ, 便于在训练时通过调整它的值提升训练效果, 利用这 3 个项, 本文得到了 PAN 模型的优化函数, 通过上述最小化和最大化操作, 即 D 和 C 之间的最小值博弈, 构建了 PU 学习的对抗性学习方法.

3.5 PAN 模型的优势

(1) 它不需要先验概率, 而很多同类的先进分类器需要先验概率, 但是在实际问题中先验概率是未知的, 所以这是 PAN 模型对比起同类模型的一大优势.

(2) 应用范围广. 由于 PAN 没有生成器, 它可以应用在不同类型的数据上, 而不仅限于图像类. 实验得出, 对比起同类的 PU 模型, 在文本类数据上 PAN 模型表现优异.

(3) 模型效果好. 将 PAN 模型 和 PU 学习的其他高级模型在不同数据集上进行对比. 实验证明, 在给予其他模型准确的先验概率条件下, PAN 的精确度、$F1$-score 和鲁棒性都强于所有同类模型. 对结果进行灵敏

性分析后, 本文得出如果在未知先验概率的条件下, 其他模型的表现会更差.

4 训练过程

本文利用小批量随机梯度下降算法, 对 PAN 模型进行训练. 设置总训练次数为 k, 流程图如图 2.

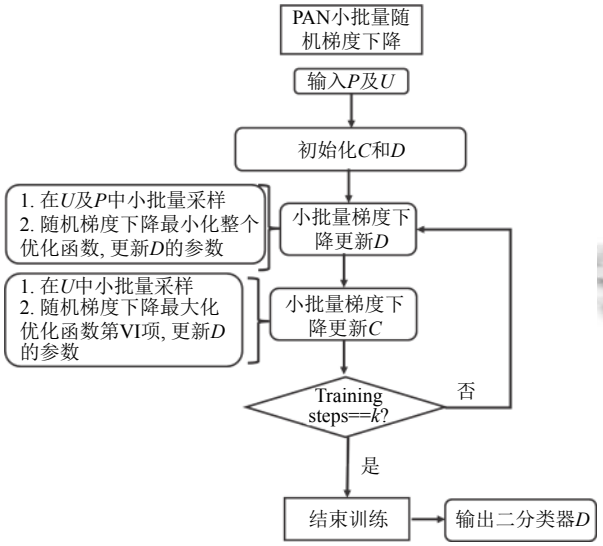


图 2 小批量随机梯度下降的流程图

5 实验分析

实验分别采用两个普通图像数据集、3 个医学图像数据集, 以同类正未标记学习模型, 官网给出的基准模型及原论文的 PAN 模型结果作为对比, 测试模型的精确度、$F1$-score 指标.

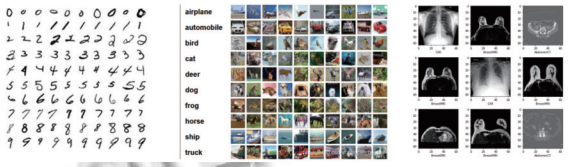
5.1 相关指标定义

精确度 (Accuracy, ACC), 也称准确度, 表示的是分类器正确预测占总预测的比例; 召回率 (Recall) 表示正样本中被预测正确的比例; 误报率表示负样本中预测为正样本的比例; $F1$-score 是精确率和召回率的调和平均数, 最大为 1, 最小为 0.

对二分类模型而言, 精确度 (ACC) 及 $F1$-score 是重要的衡量指标. 研究者在 PU 学习中希望通过网络找出所有的正类数据, 从而进行后续分析; 如果模型在找出部分正类数据的同时, 漏掉了很多其他正类样本, 则将大大降低模型的可解释性, 达不到研究目的.

5.2 数据集介绍及准备工作

PU 数据集有如下特点: 其一, 包含了少量标签数据和大量无标签数据; 其二, 数据集的分类结果只有两类 (正类与负类) 因此, 我们需要将实验数据集改造成以一定比例的标签数据组成的, 且标签数据只含有正负两类的 PU 数据集. 本文分别以 ratio=0.5, 0.6, 0.7, 0.8, 0.9 共 5 种无标签数据的比例进行实验, 以模拟不同情况下的数据集实验情况. 关于实验采用的 3 个数据集的示意图如图 3. 以下分别介绍本文对每个数据集的二分类标签准备工作.



(a) MNIST (b) CIFAR-10 (c) Medical-MNIST

图 3 数据集示意图

(1) MNIST[17]

MNIST 是一个手写体数字的图片数据集, 标签为 0–9 的数字, 共 10 种. 本文将标签为奇数的 5 类样本作为负类, 偶数的 5 类样本作为正类, 创建二分类的 MNIST 数据集.

(2) CIFAR-10[18]

CIFAR-10 是一个 60 万张用于识别普适物体的彩色图片的数据集. 这个数据集一共包含 10 个不同类别的彩色图像, 图片的尺寸为32×32. 本文挑选飞机、手机、卡车和轮船 4 类图像作为正类数据, 其余 6 类作为负类数据, 构建二分类数据集.

(3) Medical-MNIST[19]

Medical- MNIST (简称 Med-MNIST) 是一个新型开放医疗图像数据集, 提供共十类的医学图像分类数据集. 数据集中的图像分辨率为28×28, 标签均为个位数字且奇偶标签的样本个数相当. Med-MNIST 数据集在轻量级28×28图像上执行分类任务, 主要用于医疗图像模态分类和多样化的数据规模. 本文挑选 Medical-MNIST 中不同尺寸输入的 3 个数据集: 3D 数据集 Synapse-MNIST3D, 2D-1 通道数据集 Tissue-MNIST 及 2D-3 通道 Retina-MNIST. 本文将标签为奇数的样本作为负类, 偶数的样本作为正类, 创建 Med-MNIST 数据集种有标签的部分.

5.3 实验结果

首先, 本文在 MNIST 和 CIFAR-10 数据集上, 以 NNPU[20], UPU[21] 和 NNPUSB[22] 模型为对比的 baseline, 设置 CPU 个数为 8, epochs 为 200, 间隔 400 个点取样, 实验结果分别为表 1 及表 2. 在 MNIST 数据集上, PAN

模型的 ACC 及 *F*1-score 分别为 96.35% 及 96.49%, 均远胜于其他 baseline 的模型, 这说明 PAN 模型分类精确、真实可信, 与预期分析一致. 运用对称化 KL 散度后, 本文得到 ACC 及 *F*1-score 分别为 96.68% 及 96.61%, 分别有了约 0.33% 及 0.12% 的效果提升. 在 CIFAR-10 数据集上, 由表 2 可见 PAN 模型的 ACC 及 *F*1-score 都略胜于其他 baseline 的模型, 说明 PAN 模型分类效果较好; 运用对称化 KL 散度后, 本文得到 ACC 及 *F*1-score 分别为 89.81% 及 87.29%, 分别有了约 0.40% 及 0.65% 的效果提升. 由于篇幅有限, 本文以基于 KL 散度的 PAN 模型在 MNIST 数据集上的训练数据为代表, 给出损失收敛曲线及精确度曲线随训练轮数 epoch 的变化图, 如图 4 所示.

表 1 在 MNIST[17] 数据集上的测试结果

模型	ACC	*F*1-score
NNPU[20]	0.9535	0.9540
UPU[21]	0.9555	0.9560
NNPUSB[22]	0.9651	0.9555
KL-PAN[16]	0.9635	**0.9649**
JS-PAN	0.9668	**0.9661**

表 2 在 CIFAR-10[18] 数据集上的测试结果

模型	ACC	*F*1-score
NNPU[20]	0.8884	0.8609
UPU[21]	0.8896	0.8620
NNPUSB[22]	0.8859	0.8656
KL-PAN[16]	0.8941	**0.8664**
JS-PAN	0.8981	**0.8729**

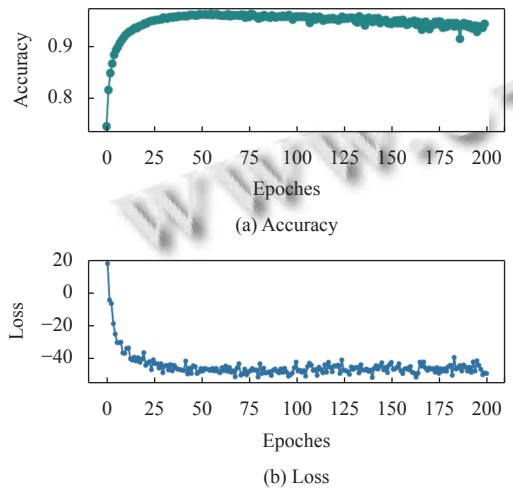


图 4 MNIST 数据集的精确度及损失收敛曲线

最后, 本文在 Med-MNIST 的 3 个子数据集 Synapse-MNIST3D, Tissue-MNIST, Retina-MNIST 上, 实验设置 CPU 个数为 8, epochs 为 200, 间隔 400 个点取样, 以 Auto-Sklearn[23], ResNet-18 (28)[24], ResNet-50 (28)[25] 和 AutoKeras[26] 模型为对比 benchmark, 实验结果如表 3. 可以看到, 基于 KL 散度的 PAN 模型的 ACC 都远胜于其他 baseline 模型, 说明 PAN 模型在 Medical-MNIST 医疗图像数据集上分类精确度高, 与预期分析一致; 同时, 对称化改进后的 PAN 模型也在 3 个数据集的 ACC 上有了不同程度的提高, 说明对称化的合理性. 证实了 PAN 模型在正未标记学习中有出色的二分类表现.

表 3 在 Med-MNIST[19] 的 3 个子数据集上的精确度

模型	Synapse-MNIST3D	Tissue-MNIST	Retina-MNIST
Auto-Sklearn[23]	0.730	0.532	0.515
ResNet-18 (28)[24]	0.696	0.676	0.524
ResNet-50 (28)[25]	0.709	0.681	0.528
AutoKeras[26]	0.724	0.703	0.503
KL-PAN[16]	**0.8412**	**0.6966**	**0.7775**
JS-PAN	**0.8531**	**0.6987**	**0.7925**

6 PAN 模型在医疗诊断图像的分类运用

本文着重研究 PAN 模型在医疗诊断图像数据集上的二分类表现. 以下将从领域特点、实验分析两个角度说明正未标记学习模型 PAN 在医疗诊断领域图像二分类中有较高的性能.

6.1 领域特点

(1) 数据集特点

医学诊断领域的数据特点突出. 除了少量确诊病例的各类数据 (例如 CT 图像、MRI 和 PET 等), 医院接收新的病人将有海量未诊断的新数据产生, 未知疾病状况, 没有对应的标签; 同时数据库中较少留存被完全排除疾病可能性的病例样本. 结合以上分析, 医学诊断领域数据集大致满足“包含少量正样本 (positive samples) 和大量无标签样本点 (unlabeled samples)”的特征, 与深度学习中正未标记学习 (positive-unlabeled learning, PU learning) 的训练数据条件较为相符. PAN 作为正未标记学习的模型, 十分符合医学诊断图像的分类应用背景.

(2) 现状与需求

目前, 深度学习在医学图像分类模型有以下几个局限性: 一方面, 由于诊断费时且只有相关疾病专家才能对数据进行标注, 获取高质量的人工标注进行训练成本高、强度大, 精确标注的训练数据少成为在医学图像上运用深度学习模型的主要障碍; 另一方面, 如何最大化利用无标签的数据进行神经网络的训练是此领

域的另一个难题. 最后, 由于有标签的数据少, 训练时需要采取加噪声等措施进一步提升鲁棒性及稳定性, 以避免模型过拟合.

综上, 此领域中获取高质量的数据标记的成本大且耗时长, 模型的学习效果在很大程度上依赖于专家给出的人工诊断以及大量训练图像的可用性. PAN 作为正未标记学习模型这种无监督模型之一, 需要的标注数据较少, 可以利用大量无标签数据; 同时 PAN 的半监督特点又可获得更快、更好的训练效果; 另外, PAN 模型中采取的鉴别器与分类器互相对抗的思想, 进一步提升了模型的鲁棒性及稳定性. 因此, 本文研究的 PAN 模型在医学图像的二分类的任务上具有更高的性能.

6.2 实验分析

(1) 模型评价指标

在医疗诊断领域中, 若模型主要用于辅助诊断, 则精确度 (ACC) 无疑是一个最重要的评价指标——这将关系到患者是否得到准确的疾病信息, 从而接受及时的治疗; 若模型主要用于挑选样本以供医疗工作者研究, 对比起无病样本, 医疗研究者更关注患病样本, 以从中找到疾病相关信息. 这就需要模型尽可能将正类样本全部识别出来, 因此除了 ACC 外, 召回率 (Recall) 这一指标也非常重要. 而 *F*1-score 作为 ACC 和 Recall 两者的调和平均数, 是针对正样本的一个综合衡量性指标, 可以很好地综合代表模型用于医疗领域研究的图像二分类效果. 综上, 在医疗诊断领域中的二分类模型效果主要由 ACC 和 *F*1-score 作为主要衡量指标. 因此我们主要采取这两个指标对 PAN 模型在 CIFAR-10 和 MNIST 两个数据集上的分类效果进行评估, 同时由于 Med-MNIST 官网没有 *F*1-score 数据, 只采取了 ACC 作为对比指标.

(2) 实验结果对比

在 Med-MNIST 数据集上, 官网上提供的 benchmark 模型均为有监督模型, 其分类效果在同等条件下应强于同类半监督模型. 而 PAN 模型作为半监督模型, 在 3 个数据集上却获得了更好的精确度 (ACC), 这更加证明出 PAN 模型在医学诊断领域图像的二分类效果出色. 另外, 由于半监督模型数据获取教易, 训练较快等特点, 在医学图像分类中也有更广泛的应用空间.

7 结论与展望

本文主要针对半监督学习中的正未标记学习提出了一种基于 KL 散度及 JS 散度的预测性对抗网络, 并将其在图像数据集 CIFAR-10, MNIST 及医学图像数据集 Med-MNIST 的 3 个子数据集中进行测试, 得出模型的良好的分类效果及强灵敏度.

PAN 模型借鉴了 GAN 的对抗性学习思想, 通过迭代优化鉴别器 *D* 及分类器 *C*, 在两者相互博弈的过程中对网络进行对抗性训练. PAN 模型将 GAN 中的生成器 *G* 换成了自身构建的 PU 分类器 *C*, 它在无标签数据集 *U* 里选出可能是正类的样本, 去欺骗鉴别器 *D*, 让 *D* 判别数据是否真正是训练集中有标签的正样本; 训练过程中, *C* 尽可能拟合 *D* 的数据分布, 以在无标签数据集中选出假数据欺骗 *D*; 而 *D* 通过尽可能拉大与 *C* 的分布差异, 辨别出 *C* 选择的假数据, 以此构成对抗性学习网络; PAN 模型将 *C* 和 *D* 间的总体分布差异作为目标函数, 从而实现对 *D* 和 *C* 的迭代优化. 本文先研究基于 KL 散度的 PAN 模型, 再利用对称化思想创新性提出基于 JS 散度的 PAN 模型, 两个模型分别利用 KL 散度及 JS 散度作为衡量两个分布的差异函数构建目标函数. 在 5 个数据集的实验结果表明, PAN 模型在精确度、*F*1-score 上优于几乎所有同类模型, 对称化后的 JS_PAN 模型具有更好的指标, 说明对称思想的合理性及 PAN 在正未标记学习的出色二分类效果.

由于医疗诊断数据集中含有大量无标签样本及少量正类样本, 适合运用 PU 学习最大化利用数据信息. PAN 模型正是基于这样的数据特征, 是 PU 学习中的创新模型之一. 对比起同类的 PU 模型, PAN 模型不需要先验概率且不仅仅能运用在图像数据上, 在其他数据类型如文本数据也有较好的分类表现. 经过在 Med-MNIST 数据集上对比实验, PAN 模型作为半监督模型, 其精确度 (ACC) 明显高于其他 benchmark 的模型, 说明 PAN 模型效果好、灵敏度高、应用范围广.

参考文献

1 Medley DO, Santiago C, Nascimento JC. CyCoSeg: A cyclic collaborative framework for automated medical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(10): 8167–8182.

2 Lin AL, Chen BZ, Xu JY, *et al*. DS-TransUNet: Dual swin transformer u-net for medical image segmentation. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 4005615.

3 Liu JF, Shen C, Aguilera N, *et al*. Active cell appearance model induced generative adversarial networks for annotation-efficient cell segmentation and identification on adaptive optics retinal images. IEEE Transactions on Medical

Imaging, 2021, 40(10): 2820–2831. [doi: 10.1109/TMI.2021.3055483]

4 Lei BY, Xia ZM, Jiang F, *et al*. Skin lesion segmentation via generative adversarial networks with dual discriminators. Medical Image Analysis, 2020, 64: 101716. [doi: 10.1016/j.media.2020.101716]

5 Yuan WG, Wei J, Wang JB, *et al*. Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images. Medical Image Analysis, 2020, 64: 101731. [doi: 10.1016/j.media.2020.101731]

6 Denis F. PAC learning from positive statistical queries. Proceedings of the 9th International Conference on Algorithmic Learning Theory. Otzenhausen: Springer, 1998. 112–126.

7 Liu B, Lee WS, Yu PS, *et al*. Partially supervised classification of text documents. Proceedings of the 19th International Conference on Machine Learning. Sydney: Morgan Kaufmann Publishers Inc., 2002. 387–394.

8 Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 2672–2680.

9 Sajjadi MSM, Schölkopf B, Hirsch M. EnhanceNet: Single image super-resolution through automated texture synthesis. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 4501–4510.

10 Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. 4401–4410.

11 Kim SW, Zhou YH, Philion J, *et al*. Learning to simulate dynamic environments with GameGAN. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 1228–1237.

12 Hou M, Chaib-Draa B, Li C, *et al*. Generative adversarial positive-unlabelled learning. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: IJCAI, 2018. 2255–2261.

13 Chiaroni F, Rahal MC, Hueber N, *et al*. Learning with a generative adversarial network from a positive unlabeled dataset for image classification. Proceedings of the 25th IEEE International Conference on Image Processing (ICIP). Athens: IEEE, 2018. 1368–1372.

14 Xu YX, Xu C, Xu C, *et al*. Multi-positive and unlabeled learning. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2017. 3182–3188.

15 Bekker J, Robberechts P, Davis J. Beyond the selected completely at random assumption for learning from positive and unlabeled data. Proceedings of the 2019 Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Würzburg: Springer, 2019. 71–85.

16 Hu WP, Le R, Liu B, *et al*. Predictive adversarial learning from positive and unlabeled data. AAAI, 2021. 7806–7814.

17 LeCun Y, Bottou L, Bengio Y, *et al*. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324. [doi: 10.1109/5.726791]

18 Krizhevsky A. Learning multiple layers of features from tiny images. Technical Report, Toronto: University of Toronto, 2009. 7.

19 Yang JC, Shi R, Wei DL, *et al*. MedMNIST v2—A large-scale lightweight benchmark for 2D and 3D biomedical image classification. Scientific Data, 2023, 10(1): 41. [doi: 10.1038/s41597-022-01721-8]

20 du Plessis MC, Niu G, Sugiyama M. Convex formulation for learning from positive and unlabeled data. Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille: JMLR.org, 2015. 1386–1394.

21 Kiryo R, Niu G, du Plessis MC, *et al*. Positive-unlabeled learning with non-negative risk estimator. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 1674–1684.

22 Kato M, Teshima T, Honda J. Learning from positive and unlabeled data with a selection bias. Proceedings of the 7th International Conference on Learning Representations. New Orleans: ICLR, 2019.

23 Feurer M, Eggensperger K, Falkner S, *et al*. Auto-sklearn 2.0: Hands-free automl via meta-learning. The Journal of Machine Learning Research, 2022, 23(1): 261.

24 Odusami M, Maskeliūnas R, Damaševičius R, *et al*. Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network. Diagnostics, 2021, 11(6): 1071. [doi: 10.3390/diagnostics11061071]

25 Theckedath D, Sedamkar RR. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. SN Computer Science, 2020, 1(2): 79. [doi: 10.1007/s42979-020-0114-9]

26 Jin HF, Song QQ, Hu X. Auto-Keras: An efficient neural architecture search system. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 1946–1956.

(校对责编: 孙君艳)