



计算机工程
Computer Engineering
ISSN 1000-3428,CN 31-1289/TP

《计算机工程》网络首发论文

题目: 基于分段损失的生成对抗网络
作者: 刘其开, 姜代红, 李文吉
DOI: 10.19678/j.issn.1000-3428.0050529
网络首发日期: 2018-05-19
引用格式: 刘其开, 姜代红, 李文吉. 基于分段损失的生成对抗网络. 计算机工程.
<https://doi.org/10.19678/j.issn.1000-3428.0050529>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于分段损失的生成对抗网络

刘其开¹ 姜代红² 李文吉^{3,4}

(1.中国矿业大学 信息与控制工程学院, 徐州 221116; 2.徐州工程学院, 徐州 221111; 3.国土资源部航空地球物理与遥感地质重点实验室, 北京 100083; 4.中国国土资源航空物探遥感中心, 北京 100083)

摘 要：近几年生成对抗网络在图像生成和半监督学习上应用广泛。但 GAN 在训练过程中难以把握生成器与鉴别器的同步更新，导致模型训练不稳定以及模式崩溃的现象，鉴别器提取的特征鲁棒性较差。为此，本文提出一种基于分段损失的生成对抗网络，一方面让生成器在不同的训练时期采用不同形式的损失函数，另一方面为生成器引入真实样本与生成样本之间特征级 (feature-wise) 损失，使网络训练更加稳定，鉴别器提取的特征更加鲁棒。实验在 MNIST 和 CIFAR-10 数据集上验证了所提算法的有效性。

关键词：生成对抗网络；模式崩溃；特征级损失；分段损失；半监督学习

Piecewise-Loss Generative Adversarial Network

Liu Qikai¹, Jiang Daihong², Li Wenji^{3,4}

(1.College of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221116; 2.college of Xuzhou engineering, Xuzhou, 221111; 3. Key laboratory of airborne geophysics and remote sensing geology, Ministry of land and resources, Beijing 100083; 4. China Aero Geophysical Survey and Remote Sensing Center for Land and Resources, Beijing, 100083)

Abstract: In recent years, Generative Adversarial Network (GAN) has been applied widely in image generation and semi-supervised learning. But GAN fail to grasp the synchronous update of generator and discriminator during the training process, which leads to the instability of model training and the phenomenon of mode collapse, and the robustness of feature extracted from discriminator is poor. To this end, this paper proposes a piecewise loss GAN, on the one hand, the generators will apply different loss function in different periods of training, on the other hand, the generator will be introduced into the feature-wise loss between real sample and generate sample. As a result, the network training is more stable, the feature extracted from discriminator is more robust. The validity of the proposed algorithm is verified by experiments in MNIST and CIFAR-10 datasets.

Keywords: Generative Adversarial Network; mode collapse; feature-wise loss; Piecewise-loss; semi-supervised learning

DOI:10.19678/j.issn.1000-3428.0050529

1 概述

机器学习算法以训练样本有无标签，可以分为带标签的监督学习和无标签的无监督学习。由于监督学习的标签数据获取的成本很高，无监督学习算法表现不足，因此半监督学习^[1] (semi-supervised learning, SSL) 成为研究学者重要的一个研究热点。

SSL 利用海量的无标签样本和少量标签样本能够学习具有鲁棒性的特征，在图像分类方面有着较好的表现。Lee^[2]等提出了一种对无标签数据的伪标签来帮助模型训练的高效方法。Rasmus^[3]等人提出了基于自动编码器的阶梯网络，编码器用于监督学习，解码器的每一层与编码器一一对应，形成阶梯，用于无监督学习训练。

基金项目：国家自然科学基金资助 (51574232)；国土资源部航空地球物理与遥感地质重点实验室航遥青年创新基金资助 (编号：2016YFL02)；徐州市科技计划项目基金资助 (KC16SQ78)

作者简介：刘其开 (1992-)，男，硕士，主要研究方向：深度学习、计算机视觉；姜代红 (1969-)，女，教授，博士，主要研究方向：智能计算；李文吉 (1988-)，男，工程师，硕士，主要研究方向：GIS 软件工程、云计算、大数据

E-mail: liuqikai_0127728@163.com

近年来,深度生成模型与生成对抗网络(Generative Adversarial Network, GAN)^[4]在SSL上有着良好的表现。Kingma^[5]等提出了灵活的深度生成模型算法,其中包括隐含特征判别模型(M1)和生成半监督模型(M2),前者学习输入样本在高维空间的流形,后者结合标签信息推断样本的后验概率 $p_\theta(y|x)$,最后将两者结合堆叠为生成半监督模型(M1+M2)。Maaløe^[6]等提出了辅助深度生成网络(ADGM),通过引入辅助的隐变量,改进了变分近似算法。Dumoulin^[7]等提出对抗学习推理模型(ALI),通过引入一个推理器,将样本数据空间映射到隐变量空间,输出样本和隐变量的联合分布;相应地生成器输出隐变量与生成样本两者的联合分布,鉴别器判断输入的联合分布的来源。Springenberg^[8]等提出的Cat-GAN在目标函数中引入了权衡输入样本与对应预测类别的互信息,通过最大化生成数据类别的条件交叉熵来训练分类器。Salimans^[9]等将鉴别器的二分类的概率输出扩展到 $N+1$ 类别概率输出。Li^[10]等提出了Triple GAN,通过引入额外的分类器,改善生成器和鉴别器在训练时无法同时达到最优的问题。文献^[11]中提到利用无标签数据对GAN的鉴别器进行预训练,用少量有标签数据进行微调,再用于分类任务。

GAN在训练过程中出现不稳定以及模式崩溃的现象,Arjovsky^[12]等从理论分析了原因,当生成样本分布与真实样本分布之间没有重叠或可忽略的重叠部分为0时,生成器的损失梯度近似常数。常规GAN采用的JS散度受限于分布之间没有重叠。Wasserstein-GAN^[13]通过设计满足一定约束条件的神经网络逼近函数来近似度量两者分布的距离。Metz^[14]等针对模式崩溃的问题,提出了基于梯度损失的Unrolled GAN,通过计算二阶梯度来指示生成器一阶梯度变化的方向。

针对上述问题,本文提出一种基于分段损失的生成对抗网络(Piecewise-Loss GAN, PL-GAN),通过引入时间参数,使生成器在不同的训练阶段采用不同形式的损失;此外,引入了生成样本和真实样本在特征空间的均方差损失。PL-GAN不仅有效改善了生成器梯度消失的问题,而且在SSL方面有着不俗的表现。

2 基于分段损失的生成对抗网络

GAN可以理解成两个玩家相互博弈的二人游戏,即生成器(Generator, G)和鉴别器(Discriminator, D)之间的“零和游戏”。G的输入是噪声变量 z ,目的是拟合真实样本的数据分布,输出生成样本 $G(z)$ 。D来判断输入的样本是来自真实样本还是生成样本,输出输入样本来自真实样本的概率 $D(x)$ 。因此,D的优化目标对输入为真实样本的概率尽可能为1,输入为生成样本的概率尽可能为0。G的优化目的是最大化D给出错误判断的概率,最终的优化目的是寻找两者之间的一个纳什均衡。其优化过程是一个极大极小的一个对抗过程,其目标函数为 $V(D, G)$:

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

其中, $p_{data}(x)$ 表示真实样本的分布, $p_{noise}(z)$ 表示随机噪声分布。

2.1 算法思想

GAN的最终目标是通过双玩家游戏策略隐式地训练一个逼近真实样本的生成器 $p_g = G_\theta(z)$,其中,D要尽可能鉴别出真实样本与生成样本的真伪,其损失函数为:

$$D_{loss} = -E_{x \sim p_{data}} [\log D(x)] - E_{z \sim p_{noise}} [\log(1 - D(G(z)))] \quad (2)$$

G的损失函数有以下两种形式:

$$G_{loss_first} = E_{z \sim p_{noise}} \log(1 - D(G(z))) = E_{x \sim p_g} \log(1 - D(x)) \quad (3)$$

$$G_{loss_second} = -E_{z \sim p_{noise}} \log D(G(z)) \quad (4)$$

当 $p_{data}(x) = p_g(x)$ 时,最优鉴别器表示为:

$$D^*(x) = \frac{P_{data}(x)}{P_{data}(x) + P_g(x)} = \frac{1}{2} \quad (5)$$

最优鉴别器下,式(5)代入式(1)得:

$$E_{x \sim p_g} \log(1 - D^*(x)) + E_{x \sim p_{data}} \log(D^*(x)) \quad (6)$$

式(6)代入 $D^*(x)$ 后,引入衡量相似度的两个指标KL散度和JS散度。

$$KL(P_1 \| P_2) = E_{x \sim P_1} \log \frac{P_1}{P_2}$$

$$JS(P_1 \| P_2) = \frac{1}{2} KL(P_1 \| \frac{P_1 + P_2}{2}) + \frac{1}{2} KL(P_2 \| \frac{P_1 + P_2}{2})$$

则最优鉴别器下G第一种形式损失函数等价于:

$$G_{loss_first|D^*} = 2JS(P_{data} \| P_g) - 2\log 2 \quad (7)$$

即在最优鉴别器下,最小化生成器的损失等价

为最小化生成样本与真实样本之间的 JS 散度。但 Arjovsky^[12]等证明了在第一种形式损失函数下, JS 散度衡量分布差异的前提是两者的分布要有所重叠或有不可忽略的重叠, 否则 JS 散度将会是一个常数。但网络初始化后的生成样本分布很难与真实样本分布有不可忽略的重叠。

同样, 由式 (6) 和式 (7) 可得 G 的第二种形式的损失函数如下:

$$\begin{aligned} G_{\text{loss_second}} &= \text{KL}(P_g \| P_{\text{data}}) - E_{x \sim P_g} \log(1 - D^*(x)) \\ &= \text{KL}(P_g \| P_{\text{data}}) - 2JS(P_{\text{data}} \| P_g) + 2\log 2 + E_{x \sim P_{\text{data}}} \log(D^*(x)) \end{aligned} \quad (8)$$

由于后两项不依赖于 G, 最终最小化式 (4) 等价于最小化

$$G_{\text{loss_second}|D^*} = \text{KL}(P_g \| P_{\text{data}}) - 2JS(P_{\text{data}} \| P_g) \quad (9)$$

该目标形式一方面要求最小化生成分布与真实分布的 KL 散度, 另一方面又要求最大化两者的 JS 散度, 优化目标相互矛盾。且 $\text{KL}(P_g \| P_{\text{data}})$ 不是一个对称的度量, 当 P_{data} 与 P_g 的取值相对改变时, KL 散度也会变化, 这就迫使生成器生成大量重复且置信度较高的样本, 导致了模式崩溃。

WGAN 针对生成器第一种形式损失函数存在的缺点, 提出了对生成样本和真实样本加噪声的方法, 使得原本的两个低维流形弥散到整个高维的空间, 迫使它们产生不可忽略的重叠; 此时, JS 散度真正发挥作用, 避免梯度消失。随着训练的进行, 逐渐对噪声退火, JS 散度继续产生有意义的梯度把两个低维流形拉近, 直到完全重合。

本文借鉴了噪声退火的思想, 通过引入时间参数 $w(t) = \exp[-10 * (1-t)^2]$, 来控制生成器在不同的训练阶段采用不同形式的损失。训练的前期以 $E_{z \sim P_{\text{noise}}} [-\log D(G(z))]$ 为主, 随着训练的进行, 真实样本和生成样本就能够有所重叠, 训练进行到某一阶段, 再切换到以 $E_{z \sim P_{\text{noise}}} [\log(1 - D(G(z)))]$ 为主, 此时 JS 散度就可以发挥良性作用。同时, 为了给生成器提供足够的梯度, 引入生成样本和真实样本之间特征级的均方差损失。最终生成器的损失函数如下:

$$\begin{aligned} \min_G V(G) &= \alpha \{w(t) E_{z \sim P_{\text{noise}}} [\log(1 - D(G(z)))] \\ &\quad + (1-w(t)) E_{z \sim P_{\text{noise}}} \log(-D(G(z)))\} \\ &\quad + \beta \|E_{x \sim P_{\text{data}}} D_f(x) - E_{z \sim P_{\text{noise}}} D_f(G(z))\|_2^2 \end{aligned} \quad (10)$$

其中, $D_f(*)$ 表示鉴别器特征层的输出, α 表示对抗损失的权重系数, β 表示特征级损失系数。有关

PL-GAN 的计算流程如图 1 所示, epoch 表示当前迭代次数, T 表示损失切换迭代次数。

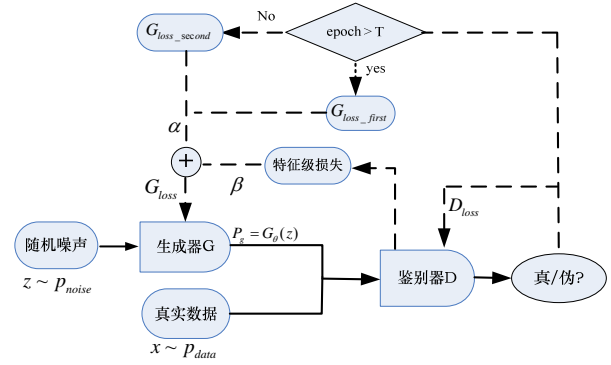


图 1 PL-GAN 的计算流程

考虑到 GAN 的监督损失, 假设标准的分类器输出是 N 维向量 $\text{logits} = \{l_1, l_2, \dots, l_N\}$, N 为样本的类别数。则用 softmax 计算输出的概率为:

$$P_{\text{model}}(y = j | x) = \exp(l_j) / \sum_{n=1}^N \exp(l_n) \quad (11)$$

将生成样本所属的类别定义为第 $N+1$ 类, 则来自生成样本的概率为 $P_{\text{model}}(y = N+1 | x)$, 对应常规 GAN 的 $1-D(x)$ 。假设鉴别器的训练样本一半来自真实样本, 一半来自生成样本, 则 D 的损失函数为:

$$\begin{aligned} C(D) &= -E_{x, y \sim P_{\text{data}}(x, y)} [\log P_{\text{model}}(y | x)] - E_{x \sim P_g} [\log P_{\text{model}}(y = N+1 | x)] \\ &= C_{\text{sup}} + C_{\text{adv}} \end{aligned} \quad (12)$$

其中, 真实样本由带标签样本和不带标签样本组成。带标签样本参与的监督损失为:

$$C_{\text{sup}} = -E_{x, y \sim P_{\text{data}}(x, y)} [\log P_{\text{model}}(y | x, y < N+1)] \quad (13)$$

不带标签的真实样本和生成样本参与的无监督对抗损失为:

$$\begin{aligned} C_{\text{adv}} &= -E_{x \sim P_{\text{data}}(x)} [\log(1 - P_{\text{model}}(y = N+1 | x))] \\ &\quad - E_{x \sim P_g} [\log P_{\text{model}}(y = N+1 | x)] \end{aligned} \quad (14)$$

令 $D(x) = 1 - P_{\text{model}}(y = N+1 | x)$, 则

$$C_{\text{adv}} = -E_{x \sim P_{\text{data}}(x)} \log(1 - D(x)) - E_{x \sim P_g} \log(1 - D(x)) \quad (15)$$

如何计算监督损失和无监督对抗损失成为问题的关键。如图 2 所示为 GAN 半监督图像分类流程。

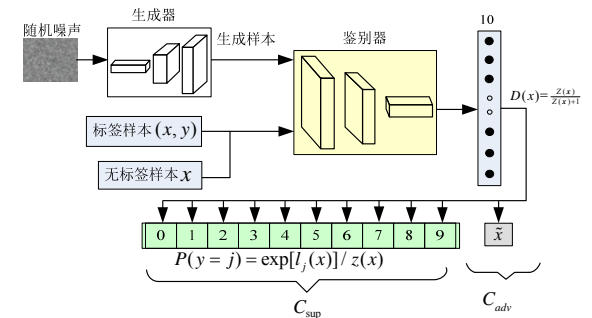


图 2 GAN 半监督分类流程图

其中, 标签样本为 D 贡献监督损失, 无标签样本为

D 贡献无监督对抗损失。

从优化目标分析, 存在一个未知的映射函数 $f(x)$, 使 $\forall j < N+1$, $P(y=j, x)=f(x) \cdot \exp[l_j(x)]$, 且 $P_G(x)=f(x) \cdot \exp[l_{N+1}(x)]$ 成立^[9]。由于鉴别器输出维度为 $N+1$ 的概率是过参数化的, 假设 $\forall x$, $l_{N+1}(x)=0$ 成立, 则不会改变鉴别器 softmax 概率值。此时

$$D(x)=z(x)/(z(x)+1), \quad z(x)=\sum_{n=1}^N \exp[l_n(x)]。$$

2.2 算法步骤

基于分段损失的生成对抗网络半监督分类算法步骤如表 1 所示。确定最大迭代次数以及损失切换次数, 设时间参数 $w(t)=\exp(-10 \cdot (1-t)^2)$, t 等于当前迭代次数 epoch 与损失切换次数 T 的比值。

表 1 PL-GAN 的半监督分类算法步骤

初始化: 设批大小 (Batchsize) $m=100$, 即每一次参数更新时所需的样本数; iterations 为批迭代次数, 即总训练样本与批大小 m 的整数比; Epoch 为迭代次数, 即遍历整个训练集的循环次数;

超参数 $k=1$, 即训练鉴别器 k 次才训练生成器 1 次; 对抗损失和特征级均方差损失权重系数分别为: $\alpha=\beta=0.5$; 用 Xavier 方法进行参数初始化。噪声分布 $P_{\text{noise}}(z)$ 满足 $Z \sim U[-1, 1]$;

for number of training iterations do

for k steps do

- 抽样 m 个服从噪声分布的样本 $\{z^{(1)}, z^{(2)} \dots z^{(m)}\}$
- 抽样 m 个无标签真实样本 $\{x^{(1)}, x^{(2)} \dots x^{(m)}\}$
- 抽样 m 个有标签真实样本 $\{(x_l^{(1)}, y^{(1)}), (x_l^{(2)}, y^{(2)}) \dots (x_l^{(m)}, y^{(m)})\}$
- 计算鉴别器的无监督对抗损失:

$$C_{adv} = -\frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))]$$

- 计算鉴别器的监督损失:

$$C_{sup} = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} * \log(\exp(x_l^{(i)}) / z(x_l^{(i)})) + (1 - y^{(i)}) * \log(1 - \exp(x_l^{(i)}) / z(x_l^{(i)}))]$$

- 通过 Adam 优化算法更新鉴别器的参数:

$$\nabla_{\theta_d} [C_{adv} + C_{sup}]$$

end for

- 抽样 m 个服从噪声分布噪声样本 $\{z^{(1)}, z^{(2)} \dots z^{(m)}\}$
- 抽样 m 个无标签样本 $\{x^{(1)}, x^{(2)} \dots x^{(m)}\}$
- 计算生成器的特征级均方差损失:

$$V_{fea}(x, z) = \|E_{x \sim p_{data}(x)} D_f(x) - E_{z \sim p_{noise}(z)} D_f(G(z))\|_2^2$$

- 计算生成器的对抗损失

$$V_{adv}(z) = w(t) E_{z \sim p_{noise}} \log(1 - D(G(z))) + (1 - w(t)) E_{z \sim p_{noise}} \log(-D(G(z)))$$

- 通过 Adam 优化算法更新生成器的参数:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \alpha V_{\log}(z^{(i)}) + \beta V_{fea}(x^{(i)}, z^{(i)})$$

end for

3 实验结果与分析

3.1 实验数据集

实验采用 MNIST^[15]、CIFAR-10^[16]数据集进行验证。MNIST 是深度学习领域常见的手写字体数据集, 共十类 (数字 0~9), 每类包含单通道的 6000 个训练样本和 1000 个测试样本。CIFAR-10 包含着 6 万张包括 10 类的三通道样本, 每类有 5000 个训练样本和 1000 个测试样本。

3.2 网络参数分析

实验在单块 GPU 型号为 GTX1080 上运行, 基

于 theano^[17]深度学习框架上实现。对于生成器的损失函数而言, 损失切换迭代次数 T 以及学习率衰减策略非常重要。

表 2 参数设置

数据集	MNIST	CIFAR-10
总迭代次数	300	600
T	60	100
初始学习率	0.003	0.0003
衰减因子	min(2.0-epoch/250, 1.0)	min(2.0-epoch/450, 1.0)

关于数据集中对应的参数 T 与学习率衰减因子

参数设置如下表 2 所示, 学习率为初始学习率与衰减因子的乘积

3.3 实验对比分析

1) MNIST 数据集

本实验的网络结构主要是由多层感知机组成, 关于 PL-GAN 模型的参数配置如表 3 所示, 为了加速模型的训练速度, 在鉴别器结构中, 使用了权重归一化^[18] (weight normalization, WN) 策略。

表 3 MNIST 数据集网络参数配置

Discriminator D	Generator G
Input: 28×28 gray image one-hot labels $y \in R^{10}$	Input Noise $\in R^{100}$
Denselayer 1000 Units lReLU, gaussian noise, WN	Denselayer 500 Units,
Denselayer 500 Units, lReLU, gaussian noise, WN	Softplus, batch norm
Denselayer 250 Units, lReLU, gaussian noise, WN	Denselayer 500 Units
Denselayer 250 Units, lReLU, gaussian noise, WN	Softplus, batch norm
Denselayer 10 Units, lReLU, gaussian noise, WN	Denselayer 784 Units, sigmoid

实验的评价标准一方面是生成样本的质量, 另一方面是半监督分类的精度。对比的算法主要是常规 GAN (regular GAN), 其生成器的损失为对抗损失 $E_{z \sim p_{noise}} \log(-D(G(z)))$; 特征级损失 GAN (feature-wise GAN), 其生成器的损失仅仅是特征级均方差损失 $\|E_{x \sim p_{data}} D_f(x) - E_{z \sim p_{noise}} D_f(G(z))\|_2^2$; PL-GAN 在常规 GAN 的基础上改变了对抗损失的形式且引入了特征级损失, 然后经过加权, 作为最终的损失。

如图 3 所示为 PL-GAN 生成样本对比图。其中图 (a) 表示真实样本的输入, 图 (b) 表示 feature-wise GAN 下的生成样本, 图 (c) 表示 regular GAN 的生成样本, 图 (d) 表示 PL-GAN 生成的样本; 与 feature-wise GAN 相比, PL-GAN 生成的样本质量较好, 与 regular GAN 相比, PL-GAN 生成的样本虽然质量上稍差, 但多样性明显好于前者。即改善了常规的 GAN 模式崩溃的现象。

如图 4 所示为 PL-GAN 与 regular GAN 及 feature-wise GAN 在训练过程中的损失变化对比图。其中对生成器的损失而言, PL-GAN 相比 feature-wise GAN 损失下降得更加稳定; 与 regular GAN 相比, PL-GAN 损失的变化呈现逐渐下降的趋

势, 而不是趋近于一个常数。对鉴别器的损失而言, PL-GAN 的变化趋势相比 regular GAN 较好, 与 feature-wise GAN 相当。

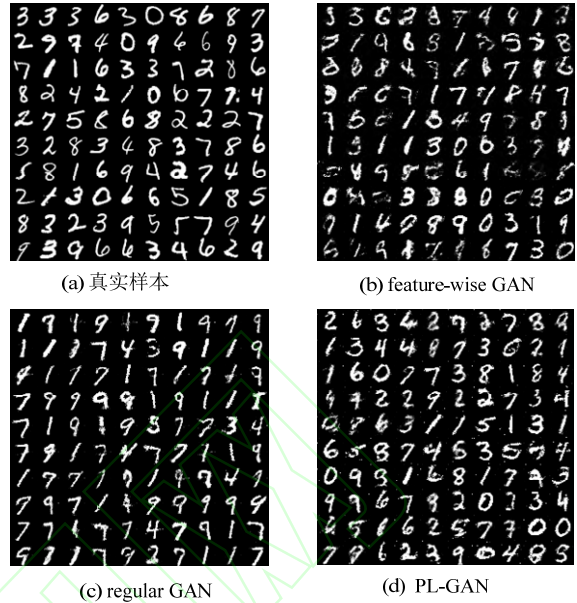


图 3 MNIST 数据集生成样本的对比图

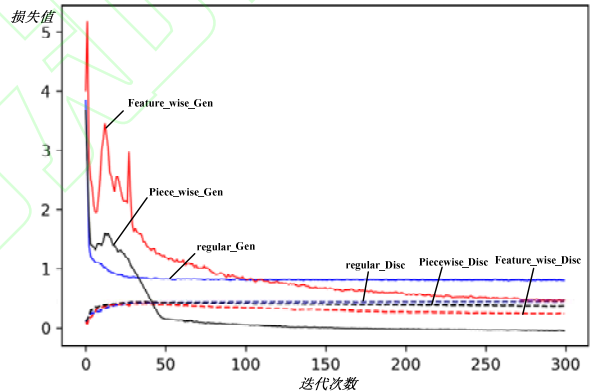


图 4 PL-GAN 损失变化对比曲线 (无单位)

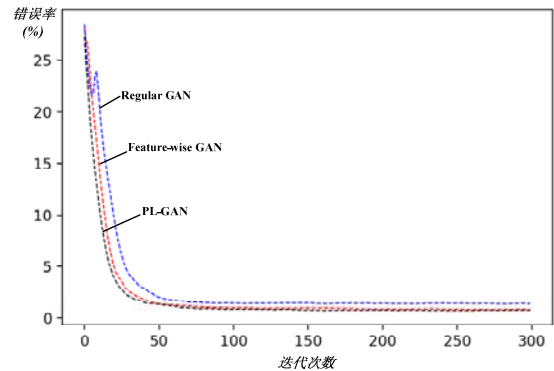


图 5 PL-GAN 在 MNIST 数据集测试对比结果 (无单位)

在保证模型的结构相同的情况下, 当标签样本为 100 时, 比较半监督分类性能。分类对比结果如图 5 所示, PL-GAN 相比 regular GAN 分类错误率

较低, 相比 feature-wise GAN, 分类性能虽基本接近, 但收敛速度更快。

同样设标签样本为 100, 与传统的半监督学习

算法, 如基于深度生成模型算法的 ADGM、M1+M2、基于虚拟对抗网络 VAT、CatGAN 等方法比较, PL-GAN 有较好的表现。比较结果如表 4 所示。

表 4 MNIST 数据集测试错误率对比结果 (labels=100)

方法	Ladder ^[3]	M1+M2 ^[5]	ADGM ^[6]	CatGAN ^[8]	Improved-GAN ^[9]	Triple-GAN ^[10]	VAT ^[19]	PL-GAN
错误率(%)	1.06(± 0.37)	3.33(± 0.14)	0.96(± 0.02)	1.39(± 0.28)	0.93(± 0.07)	0.91(± 0.58)	2.33	0.90(± 0.06)

2) CIFAR-10 数据集

本实验的生成器结构以 DCGAN^[20]的框架为基础, 初始学习率设为 0.0003, 为了加速模型的训练和防止模型的过拟合, 鉴别器使用了 WN 和 dropout 策略。关于网络模型参数配置如表 5 所示。

表 5 CIFAR10 网络参数配置

Discriminator D	Generator G
Input: 32×32 Colored image one-hot labels class $y \in R^{10}$	Input Noise $\in R^{100}$
Dropout=0.2 3×3 conv, 96, lReLU, WN 3×3 conv, 96, lReLU, WN 3×3 conv, 96, lReLU, WN	MLP 8192 units ReLU, batch norm Reshape $512 \times 4 \times 4$ 5×5 deconv, 256, stride 2 ReLU, batch norm
Dropout=0.2 3×3 conv, 192, lReLU, WN 3×3 conv, 192, lReLU, WN 3×3 conv, 192, lReLU, WN	5×5 deconv. 128. stride 2 ReLU, batch norm
Dropout=0.2 3×3 conv, 192, lReLU, WN NIN, 192, lReLU, WN NIN, 192, lReLU, WN Global pool layer Denselayer 10 Units with WN	5×5 deconv. 3. stride 2, tanh, weight norm

如图 6 所示是 PL-GAN 在 CIFAR-10 数据集生成样本的视觉对比图。为了定量衡量生成样本质量的优劣, 利用 Inception Score (IS) 的指标依次对不同方法生成的 5000 个样本进行 10 次平均评估, 得到相应的 IS 值, 比较结果如表 6 所示。

表 6 不同模型生成样本的 IS 值

方法	真实样本	regular GAN	feature-wise GAN	PL-GAN
score \pm std.	10.71 \pm 0.44	6.45 \pm 0.22	5.46 \pm 0.20	7.30 \pm 0.17

在保证模型的结构框架相同的情况下, 当标签样本为 4000 时, 半监督分类对比结果如图 7 所示, PL-GAN 相比 regular GAN 分类错误率较低, 相比 feature-wise GAN, 分类性能虽基本接近, 但收敛性更好。

与其他传统的半监督分类算法相比, 设带标签的训练样本为 4000, 对比结果如表 7 所示, PL-GAN 有较好的表现, 优于大部分传统算法。

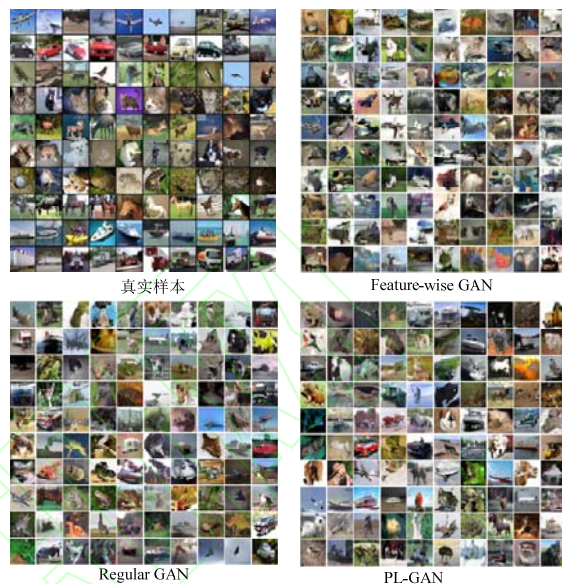


图 6 PL-GAN 生成样本的对比结果

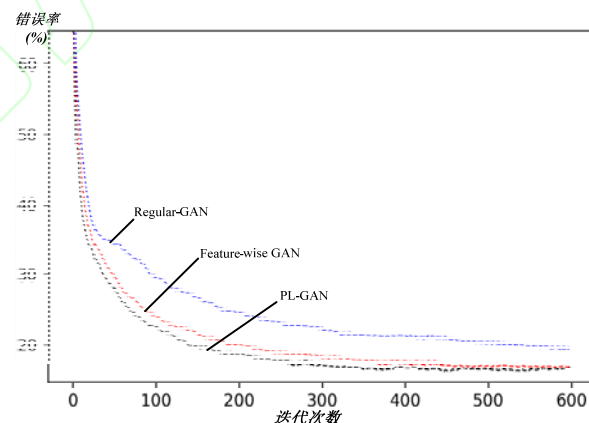


图 7 PL-GAN 在 cifar10 数据集测试对比结果 (无单位)

4 结束语

针对生成对抗网络模型训练不稳定以及模式崩溃问题, 本文提出了基于分段损失的生成对抗网络。通过引入时间参数来改变生成器与鉴别器的训练过程, 使衡量生成分布与真实分布之间差异的 JS 散度能够更好发挥良性作用; 对生成器引入真实样本与生成样本之间的特征级损失, 使训练过程更加稳定, 一定程度上能够改善模型的模式崩溃现象, 同时鉴别器提取到的特征更加鲁棒。模型在半监督图像分

类上有较好的表现，其分类精度与其他算法相比有一定的优势。

表 7 CIFAR10 测试错误数据集率对比结果 (labels=4000)

方法	Ladder ^[3]	ALI ^[7]	CatGAN ^[8]	Improved-GAN ^[9]	Triple GAN ^[10]	VAT ^[19]	PL-GAN
错误率(%)	20.40 (± 0.47)	18.3	19.58 (± 0.58)	18.63 (± 2.32)	16.99 (± 0.36)	24.65	17.30 (± 0.56)

参考文献:

- [1]. Chapelle O, Scholkopf B, Zien A. Semi-supervised learning[J]. IEEE Transactions on Neural Networks, 2009, 20(3): 542-542.
- [2]. Lee D H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks[C]//Proceedings of ICML 2013 Workshop on Challenges in Representation Learning, Atlanta, Georgia, USA, 2013.
- [3]. Rasmus A, Berglund M, Honkala M, et al. Semi-supervised learning with ladder networks [C]//Proceedings of Advances in Neural Information Processing Systems. Montreal , CANADA: NIPS, 2015: 3546-3554.
- [4]. Kingma D P, Mohamed S, Rezende D J, et al. Semi-supervised learning with deep generative models[C]//Proceedings of Advances in Neural Information Processing Systems. Montreal , CANADA: NIPS, 2014: 3581-3589.
- [5]. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// Proceedings of Advances in neural information processing systems. Montreal , CANADA: NIPS, 2014: 2672-2680.
- [6]. Maaløe L, Sønderby C K, Sønderby S K, et al. Auxiliary deep generative models[J]. arXiv preprint arXiv:1602.05473, 2016.
- [7]. Dumoulin V, Belghazi I, Poole B, et al. Adversarially learned inference[J].arXiv preprint arXiv:1606.00704, 2016.
- [8]. Springenberg J T. Unsupervised and semi-supervised learning with categorical generative adversarial networks[J]. arXiv preprint arXiv:1511.06390, 2015.
- [9]. Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans[C]// Proceedings of Advances in Neural Information Processing Systems. Barcelona, SPAIN: NIPS,2016: 2234-2242.
- [10]. Chongxuan L I, Xu T, Zhu J, et al. Triple generative adversarial nets[C]//Proceedings of Advances in Neural Information Processing Systems. Long Beach, USA: NIPS, 2017: 4091-4101.
- [11]. 王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望[J]. 自动化学报, 2017, 43(3): 321-332.
- [12]. Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks[J]. arXiv preprint arXiv:1701.04862, 2017.
- [13]. Arjovsky M, Chintala S, Bottou L. Wasserstein gan[J]. arXiv preprint arXiv:1701.07875, 2017.
- [14]. Metz L, Poole B, Pfau D, et al. Unrolled generative adversarial networks[J]. arXiv preprint arXiv:1611.02163, 2016.
- [15]. 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.
- [16]. Krizhevsky, A & Hinton, G. Learning multiple layers of features from tiny images[J]. Computer Science Department University of Toronto, 2009.1(4):32-33.
- [17]. Bastien F, Lamblin P, Pascanu R, et al. Theano: new features and speed improvements[J]. arXiv preprint arXiv:1211.5590, 2012.
- [18]. Salimans T, Kingma D P. Weight normalization: A simple reparameterization to accelerate training of deep neural networks[C]//Proceedings of Advances in Neural Information Processing Systems. Barcelona, SPAIN: NIPS,2016: 901-909.
- [19]. Miyato T, Maeda S, Koyama M, et al. Distributional smoothing with virtual adversarial training[J]. arXiv preprint arXiv:1507.00677, 2015.
- [20]. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.