

GAN 模型研究综述

刘鹤丹¹, 赵旭磊¹, 叶汉平¹, 王健²

(1. 厦门大学嘉庚学院 信息科学与技术学院, 福建 漳州 363105;

2. 中国船舶第七六〇研究所 海上试验测绘中心, 辽宁 大连 116000)

摘要: 自 2014 年生成对抗网络 GAN (Generative Adversarial Network) 模型被提出以来, GAN 已成为无监督学习中最受关注的方法之一。近年来, 众多学者基于对经典 GAN 的研究提出了一系列变体模型及适用于各类应用场景的方法。本文对 GAN 的发展历程及重要的 GAN 变体模型进行了概括、分析、总结。首先对 GAN 的基本理论进行了简要概述, 分析了 GAN 经典模型的架构和特点; 然后介绍了具有一定影响力的生成对抗网络变体模型的结构、改进方法; 最后总结了生成对抗网络与其他生成模型在优势和局限性方面的差异。

关键词: 人工智能; 生成对抗网络; 深度学习; 生成器; 判别器; 生成模型

中图分类号: TP391

文献标识码: A

文章编号: 2095-1302 (2023) 01-0091-04

0 引言

近年来, 研究人员一直在努力提升计算机的自主学习能力, 计算机学习逐步进入了人工智能阶段。人工智能作为一门交叉性强的学科, 致力于使计算机可以模拟人类的学习方法。在常见的机器学习模型中, 大部分模型是需要监督学习的, 而监督学习需要收集大量的数据集才能得到较好的训练结果, 并且数据集的标注往往是需要通过人工的手段, 这无疑会提高人工成本。在 2014 年, 生成对抗网络模型 GAN 的出现为解决该类问题提出了很好的方案, GAN 拥有无监督学习和优越的样本输出等优势, 受到大批学者的喜爱, 特别是在图像生成和目标检测等领域得到了广泛的应用与发展。

1 基本理论

1.1 理论基础

GAN 采用博弈论中零和博弈的思想^[1], 其中包含两个模型: 生成器和判别器。生成器会生成一系列的样本欺骗判别器, 而判别器的目的是识别出这些样本的真实性, 在两者的对抗过程中, 会不断地提升模型的效率, 直到生成器可生成能够欺骗判别器的样本为止, 此时达到纳什均衡^[2]。

1.2 GAN 模型的两个基本要素

GAN 模型^[3]包含生成式模型 G 和判别式模型 D (以下

生成器与判别器均用 G 和 D 替代), G 从统计学的角度直观地展示了数据分布情况, 并对其进行建模 $P(X|Y)$, 再通过贝叶斯公式进行预测, 最后根据训练数据学习联合概率分布 $P(X, Y)$ 。模型对比情况见表 1 所列。

表 1 模型对比情况

项 目	生成模型	判别模型
特 点	使用不平衡数据	不涉及不平衡数据
	对实际分布进行建模	对决策边界进行建模
	对完整的联合分布进行建模	只对条件分布进行建模
	对 $P(Y)$ 、 $P(X Y)$ 假设某种函数形式	对 $P(X Y)$ 假设某种函数形式
	基于贝叶斯定理预测条件概率	判别模型学习条件概率分布
应 用 举 例	用于生成器, 例如生成分类器、朴素贝叶斯网络、隐马尔可夫模型等	用于判别分类器, 例如标量向量机、条件随机场、Logistic 回归等

2 GAN 模型的发展历程

经典 GAN 网络架构如图 1 所示, 两个神经网络 G 和 D 构成了一个动态博弈的过程。G 和 D 使用联合损失函数, 其中 x 表示真实数据, $D(x)$ 是每个图像 x 的单个标量值, 表示此图像 x 是来自数据集的真实图像的可能性。此外, 噪声 z 的生成器输出的 $G(z)$ 、 $D(G(z))$ 是鉴别器对假实例的真实概率的估计。 E_z 是 G 的所有输入的预期值。生成器 G 试图最小化函数, 而 D 试图将函数最大化。优化目标公式如下:

$$\min(G) \max(D) E_{x \sim p_r} \lg[D(x)] + E_{z \sim p_z} \lg[1 - D(G(z))] \quad (1)$$

收稿日期: 2022-04-20 修回日期: 2022-05-18

基金项目: 厦门大学嘉庚学院大学生创新创业训练计划项目 (243); 福建省中青年教育科研项目 (JAT2009 22); 漳州市自然科学基金 (ZZ2021J10); 厦门大学嘉庚学院预研项目 (YY2019L02)

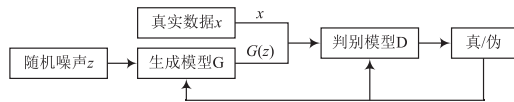


图1 经典 GAN 架构

2.1 DCGAN: 深度卷积 GAN

DCGAN (Deep Convolutional Generative Adversarial Networks) [4] 于 2015 年被提出, 其架构如图 2 所示。对 GAN 架构进行了三方面改进: (1) 代替空间池化函数为步幅卷积, 允许网络对自身空间下采样学习, 允许 D 对自身空间上采样学习; (2) 最顶层的卷积后面消除全连接层; (3) 批处理规范化。其中 G 使用了卷积神经网络, 除了最终输出层使用了 \tanh , 其余使用激活函数 ReLu, 模型目标是最小化判别网络的准确率。DCGAN 的 Generate 网络结构如图 3 所示。D 同样使用卷积神经网络, 使用的激活函数是 Leaky ReLu, 以实现最大化判别网络的准确率。

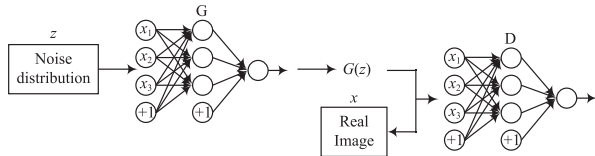


图2 DCGAN 的网络结构

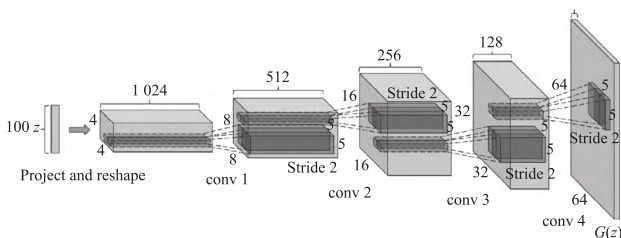


图3 DCGAN 的 Generate 网络结构

2.2 CGAN: 条件 GAN

Mirza 等人 [5] 提出了两个问题: 如何扩大已有模型以适应预测的大量输出类别, 如何实现 GAN 从一对多的映射; 同时提出了 CGAN (Conditional Generative Adversarial Networks) 模型。该模型结构额外为生成器和判别器增加了条件 y (表示希望生成的标签)。D 会生成符合条件 y 的样本, 而 G 会对生成的图像进行判断, 判断是否符合条件 y 和是否具有真实性, 并会对后续的优化目标公式进行相对应的优化改进。

2.3 SGAN: 半监督 GAN

Denton 等人 [6] 在 2016 年提出 SGAN (Semi-supervised learning GAN), 训练时可通过少量带标签的数据配合无标签数据。该模型结构解决了一个半监督分类任务的同时还学习一个生成模型; 在数据集上学习 G 的同时训练一个图像分类器 C; 共享了 D 和 C 之间的部分权重, 一些权重专属于 D,

一些权重专属于 C; 让 GAN 生成带类别标签的样本, 要求 D/C 指派标签。

2.4 InfoGAN: 无监督 GAN

针对传统 GAN 使用高度混杂的模型和随机噪声 z , 但不能显式地表示真实数据空间下的属性或特征的问题, Chen 等人 [7] 提出 InfoGAN 以无监督方式学习隐空间下非混杂方式的真实特征表示。将随机噪声分成不可压缩部分 z 和可解释的隐编码 c 。Generate (G) 表示为 $G(z, c)$, 通过最大化 c 与生成数据 x 之间的信息来防止 GAN 在训练过程中忽略 c 的特征表示能力, 并增加 c 和 $G(z, c)$ 之间的互信息 $I(c; G(z, c))$ 为约束。

2.5 WGAN

Arjovsky 等人 [8] 针对训练梯度不稳定、生成样本单一、模式崩溃、梯度消失等问题提出了四点改进措施: 去除 D 最后一层的 sigmoid, G 和 D 的损失不再取对数, 在更新 D 的参数后截断其绝对值到固定常数 c 下, 使用 RMSProp、SGD 等算法替代基于动量的优化算法。

2.6 SAGAN: 自我注意生成对抗网络

Zhang 等人 [9] 指出先前的模型过分使用卷积来模拟不同图像区域之间的依赖, 导致传统卷积 GAN 存在一些问题, 因此提出 SAGAN。将 self-attention 机制加入传统的卷积 GAN, 该机制在模拟远程依赖、计算和统计时有更加出彩的表现, 能够更好地处理长范围、多层次的依赖, 生成图像时做好每一个位置的细节和远端的协调。此外, D 能够更精准地对全局图像使用几何约束。

2.7 BigGAN: 大型生成对抗网络

针对 GAN 难以实现从复杂数据集中生成高分辨率、多样化样本的缺点, Brock 等人 [10] 提出了 BigGAN, 该模型增大了 BatchSize 和 channel number, 采用“截断技巧”等, 用原有 GAN 的 8 倍 batch size 大小, 将隐藏层变量数量扩充到原有模型的 4 倍进行训练, 获得了很好的效果。

2.8 StyleGAN: 基于风格的生成对抗网络

2017 年 NVIDIA 提出了 ProGAN [11], 该模型采用渐进式训练, 解决了高分辨率图像生成困难的问题, 但控制生成图像的特定特征能力有限。Karras 等人 [12] 提出了 StyleGAN, 每一层视觉特征都可以通过修改该层的输入来控制, 且该过程并不受其他层级的影响。StyleGAN 的模型结构除了采用传统的输入, 将一个可学习的常数作为生成器的初始输入, 还采用 8 个由全连接层所组成的映射网络来实现输出的 w 与输入的 z 大小相同; 引入样式模块 (AdaIN), 将 w 通过卷积层的 AdaIN 输入到生成网络中; 添加尺度化噪声到合成网络的分辨率级上, 然后通过输入层输入到 G 中; 对 G 使用混合正则化, 对训练样本使用样式混合的方式生成图像, 在训

练过程中使用两个随机隐码 z ，而不是一个，在生成图像时会在合成网络上随机地选择一个点，实现一个隐码到另一个隐码的切换，并且加入感知路径长度和线性可分性两种新的量化解耦方法。

2.9 EigenGAN

He 等人^[13]提出了通过明确的维度来控制特定层中语义属性表示的 EigenGAN 模型，可无监督地从不同生成器层挖掘可解释和可控的维度。EigenGAN 将一个具有正交基础的线性空间嵌入到每个生成器层；在对抗训练中，这些子空间会在每一层中发现一组“特征维度”，对应一组语义属性或可解释的变化。

2.10 ICGAN

GAN 所生成图像与训练的数据集密切相关，这导致生成的图像存在局限性。而 Casanova 等人^[14]提出 ICGAN (Instance-Conditioned GAN) 模型，可生成逼真且与训练集图像不同的图像组合。对于每个训练的数据点，采用参数化的混合形式对其周围密度进行建模处理；模型引入了一种非参数化方法对复杂数据集的分布进行建模，使用条件实例和噪声向量作为神经网络的输入，并且隐式地对其局部密度进行建模。ICGAN 中，G 与 D 可以联合训练，都参与了一个两者最小 - 最大博弈，在博弈中找到纳什均衡。

2.11 TransGAN

近年来，Transformers^[15]在 NLP 领域大放异彩，TransGAN 的研究者将 Transformers 结构作为 GAN 的主结构，并完全抛弃了 CNN 结构^[16]。实验表明，采用 Transformers 作为 GAN 模型的主体结构在其性能上几乎能够匹敌目前最好的 GAN 模型。

2.12 OUR-GAN

针对传统的生成模型很难生成视觉上连贯的图像问题，Yoon 等人^[17]提出了 OUR-GAN 模型，用低分辨率生成视觉上连贯的图像，应用无缝衔接的子区域超分辨率来提高分辨率，解决了边界不连续问题，通过向特征图添加垂直位置嵌入来提高视觉连续性和多样性。

2.13 P2GAN

文献 [18] 提出了 P2GAN (Posterior Promoted GAN)，该模型采用 D 生成的后验分布的真实信息来提升 G，并且通过将图像隐射到多元高斯分布来提取出真实的信息；G 则使用 AdaIN 后的真实信息和潜码提升自身。

2.14 Inclusive GAN

文献 [19] 提出了 Inclusive GAN 模型，通过有效融合对抗模型与重建模型提升对少数样本群体的覆盖能力，提升了模型的包容能力。引入 IMLE 重建方法调和 GAN，使其成为 IMLE-GAN，综合了两个模型的优劣，保证了数据的覆盖

性，有效提升模型整体的覆盖能力且更适应了对少数群体的包容能力。

2.15 MSG-GAN: 多尺度梯度生成对抗网络

许多 GAN 及其变体模型很难适应不同的数据集，当真实分布和生成分布的支撑集不够重叠时，判别器反馈给生成器的梯度无法提供有益信息。文献 [20] 提出了 MSG-GAN，可多尺度地提供梯度给 D 和 G。

3 GAN 的优劣

3.1 GAN 的优势

GAN 作为生成式模型，结构简单，不需要类似 GSNs 的马尔科夫链，只是用到了反向传播，其创造的数据样本远远优于其他的生成模型。这表明 GAN 可以训练任何形式下具有非特定目的生成网络；GAN 的学习方式为无监督式学习，该方式被广泛应用于无监督和半监督学习领域，其训练方式不需要推断隐变量；对比 VAEs，GANs 尚未引入决定性偏置，也没有变分下界，因此 GANs 的生成实例相比于 VAEs 更为清晰。如果 D 训练妥善，那么 G 将会得到完美的训练样本分布。GANs 的生成过程是渐进一致的过程，而 VAE 有偏差。

3.2 GAN 的劣势

GAN 普遍存在评价指标和梯度消失、模式崩溃、训练不稳定等问题。对于如何实现 GAN 在训练中达到纳什均衡，目前还没有较好的方法。虽然可以通过梯度下降法实现，但会导致其缺乏稳定性；GAN 不适合处理文本类的离散型数据，因为通常需要将一个词映射为一个高维向量，当 G 向 D 输出结果不同时，D 均会返回同样的结果，导致梯度信息不能很好地传递到 D 中。此外，GAN 损失函数为 JS 散度，并不适用于衡量不相交分布之间的距离。

4 结 语

生成对抗神经网络是当下人工智能领域的前沿。越来越多的研究人员开始在这一领域进行挖掘，并产生越来越多的高效模型。本文从介绍 GAN 模型的生成器和判别器出发，阐述了 GAN 模型的基本原理，并通过典型的 GAN 变体模型展示了 GAN 的发展历程。GAN 目前仍面临着许多挑战，但值得期待。

注：本文通讯作者为赵旭磊。

参 考 文 献

- [1] GONG M, NIU X, ZHANG P, et al. Generative adversarial networks for change detection in multispectral imagery [J]. IEEE geoscience and remote sensing letters, 2017, 14 (12): 2310-2314.
- [2] WEI Q, SONG R, YAN P. Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP [J]. IEEE transactions on neural networks

- & learning systems, 2017, 27 (2): 444-458.
- [3] 傅迎华, 付东翔, 柳樾. 机器学习中的生成式模型和判别式模型比较 [C]// “决策论坛——区域发展与公共政策研究学术研讨会”论文集 (上). 北京:《科技与企业》编辑部, 2016: 314.
- [4] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. (2016-11-01) [2023-01-20]. <https://www.oalib.com/paper/4054347>.
- [5] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J]. Computer science, 2014 (10): 2672-2680.
- [6] DENTON E, GROSS S, FERGUS R. Semi-supervised learning with context-conditional generative adversarial networks [P]. 2016-11-19.
- [7] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets [C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc, 2016.
- [8] ARJOVSKY M, BOTTOU L. Towards principled methods for training generative adversarial networks [EB/OL]. (2017-01-17) [2023-01-20]. <https://arxiv.org/abs/1701.04862>.
- [9] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [EB/OL]. (2018-05-21) [2023-01-20]. <https://doi.org/10.48550/arXiv.1805.08318>.
- [10] BROCK A, DONAHUE J, SIMONYAN K. Large scale gan training for high fidelity natural image synthesis [EB/OL]. (2018-09-28) [2023-01-20]. <https://arxiv.org/abs/1809.11096>.
- [11] KARRAS T, AILA T, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation[EB/OL]. (2017-10-27) [2023-01-20]. <https://arxiv.org/abs/1710.10196>.
- [12] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks [C]// Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 4396-4405.
- [13] HE Z, KAN M, SHAN S. EigenGAN: layer-wise eigen-learning for GANs [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021: 14408-14417.
- [14] CASANOVA A, CAREIL M, VERBEEK J, et al. Instance-conditioned GAN [EB/OL]. (2021-09-10) [2023-01-20]. <https://arxiv.org/abs/2109.05070v2>.
- [15] LIU Y, ZHANG Y, WANG Y, et al. A survey of visual transformers [EB/OL]. (2021-11-11) [2023-01-20]. <https://doi.org/10.48550/arXiv.2111.06091>.
- [16] JIANG Y, CHANG S, WANG Z. TransGAN: two pure transformers can make one strong GAN, and that can scale up [J]. Advances in neural information processing systems, 2021, 34: 1-5.
- [17] YOON D, OH J, CHOI H, et al. OUR-GAN: one-shot ultra-high-resolution generative adversarial networks [EB/OL]. (2021-11-11) [2023-01-20]. <https://doi.org/10.48550/arXiv.2202.13799>.
- [18] ZHANG X, CHENG Z, ZHANG X, et al. Posterior promoted GAN with distribution discriminator for unsupervised image synthesis [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Montreal, QC, Canada: IEEE, 2021: 6519-6528.
- [19] YU N, LI K, ZHOU P, et al. Inclusive GAN: improving data and minority coverage in generative models [C]// European Conference on Computer Vision. Cham: Springer, 2020: 377-393.
- [20] KARNEWAR A, WANG O. MSG-GAN: multi-scale gradients for generative adversarial networks [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle, WA, USA: IEEE, 2020: 7799-7808.

作者简介: 刘鹤丹 (1983—), 女, 锡伯族人, 辽宁人, 博士, 副教授, 主要研究方向为计算机视觉、机器学习。

赵旭磊 (2000—), 男, 浙江人, 本科生, 主要研究方向为计算机视觉、机器学习。

叶汉平 (2001—), 男, 福建人, 本科生, 主要研究方向为计算机视觉、机器学习。

王 健 (1985—), 男, 辽宁人, 硕士, 高级工程师, 主要研究方向为机器学习、船舶噪声。

(上接第 90 页)

- [9] 孙昌会. 物联网在智能家居中的应用研究 [J]. 电脑编程技巧与维护, 2020, 27 (9): 134-135.
- [10] 钟依平. 物联网对智能家居的影响 [J]. 中国公共安全, 2013, 12 (22): 282-283.

作者简介: 刘春池 (2001—), 男, 辽宁沈阳人, 大连理工大学城市学院通信工程专业本科在读。

刘搏飞 (2001—), 男, 辽宁沈阳人, 大连理工大学城市学院通信工程专业本科在读。

邢晓鹏 (2001—), 男, 天津人, 大连理工大学城市学院通信工程专业本科在读。

隋盛誉 (2000—), 男, 辽宁瓦房店人, 大连理工大学城市学院通信工程专业本科在读。

孙嘉成 (2000—), 男, 江苏泰州人, 大连理工大学城市学院通信工程专业本科在读。

李广凯 (2001—), 男, 河南郑州人, 大连理工大学城市学院电子信息工程专业本科在读。

谢印庆 (1980—), 男, 江苏东海人, 大连理工大学城市学院副教授, 研究方向为物联网技术、信号与信息处理。