

基于生成对抗网络的研究综述

葛生国, 范宝杰, 尹哲

(广东工业大学 信息工程学院, 广东 广州 510006)

摘要: 近年来, 随着人工智能的发展, 生成对抗网络 (Generative Adversarial Network, GAN) 已成为生成型深度学习模型中最流行的模型之一, 并且基于 GAN 繁衍了一些改进型的 GAN 模型, 在各个领域进行了研究和应用。首先对 GAN 及其衍生网络进行阐述和对比, 然后对 GAN 的应用场景进行介绍, 最后对 GAN 做了总结和展。

关键词: 人工智能; 生成对抗网络; 深度学习

中图分类号: TP391.4 文献标识码: A

文章编号: 1009-3044(2019)25-0197-02

DOI: 10.14004/j.cnki.ckt.2019.3109

A Review of the Research on Generative Adversarial Networks

GE Sheng-guo, FAN Bao-jie, YIN Zhe

(Faculty of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: In recent years, with the development of artificial intelligence, The Generative Adversarial Networks is one of the most popular models in the generative deep learning model. Firstly, GAN and its derivative network are elaborated and compared. Then, application scenarios of GAN are introduced. Finally, a summary and development of GAN are made.

Key words: artificial intelligence; generative adversarial networks; deep learning



开放科学(资源服务)标识码(OSID):

1 引言

近年来随着深度学习的火热, 大量的神经网络被提出, 而生成对抗网络是目前最为流行的生成型神经网络, 2019 年图灵奖获得者深度学习之父 Yann LeCun 称之为过去十年间, 机器学习领域最让人激动的点子。GAN 模型提出后立即被应用到许多深度学习的领域当中去, 并且针对 GAN 所存在的一些问题如训练不稳定等提出了一些改进型的 GAN 模型。

本文章节安排如下, 在第 2 节中介绍了 GAN 及其改进模型, 在第 3 节中介绍了 GAN 的主要应用, 在第 4 节中做了总结和展望。

2 GAN 及其改进模型

2.1 GAN 模型

2014 年由 Goodfellow 等人^[1]提出了 GAN 模型。GAN 主要有生成网络 G 和判别网络 D 组成。其工作原理如图 1 所示, 首先判别器 D 学习真实样本 y , 当 D 对真实样本 y 有了一定的了解之后用 D 来观察通过加入随机噪声 x 生成的样本 $G(x)$, 然后判断生成样本 $G(x)$ 是否为真实样本 y 。生成模型 G 通过判别模型 D 的判别真假不断提高自己生成接近真实样本 $G(x)$ 的能力, 而判别模型 D 通过不断学习真实样本 y 而不断提高自己判别生成样本 $G(x)$ 真伪的能力, 两个网络通过相互博弈使得各自性能不断提

高, 直到生成模型和判别模型都无法提升自己, 这样生成模型就会成为一个比较完美的模型。

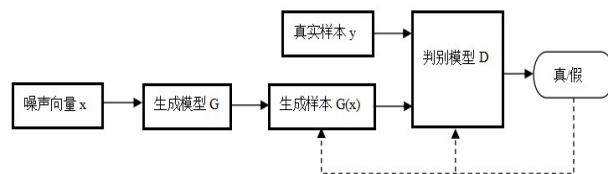


图 1 GAN 工作原理示意图

虽然 GAN 的生成效果很好但也存在一些缺陷如训练不稳定、梯度消失、模式崩溃等问题。因此针对 GAN 的这些缺陷又提出了一些改进的 GAN 模型。

2.2 GAN 的改进模型

(1) DCGAN 模型^[2]。DCGAN 的原理和 GAN 的原理一样, 不过 DCGAN 网络将 GAN 网络的 G 和 D 换成了两个卷积神经网络, 通过对两个卷积神经网络做一些特定的限定使得 DCGAN 网络训练起来更加稳定, 并且使用得到的特征表示来进行图像分类, 得到比较好的效果来验证生成的图像特征表示的表达能力。

(2) WGAN 模型^[3]。Arjovsky 等人提出了 WGAN。WGAN 指出 GAN 使用交叉熵不适合衡量具有不相交部分的分布之间

收稿日期: 2019-04-26

作者简介: 葛生国 (1992—), 男, 甘肃张掖人, 在读硕士, 主要研究方向为深度学习; 范宝杰 (1994—), 男, 安徽安庆人, 在读硕士, 主要研究方向为深度学习; 尹哲 (1993—), 男, 河南周口人, 在读硕士, 主要研究方向为深度学习。

的距离从而导致训练不稳定,因此WGAN使用wasserstein距离去衡量生成数据分布和真实数据分布之间的距离。

(3)WGAN - GP模型^[4]。鉴于WGAN将更新后的权重强制截断到一定范围内导致大多数数值有可能聚集在两个数上,导致梯度消失或梯度爆炸。提出了基于改进WGAN的WGAN - GP网络。WGAN - GP使用梯度惩罚解决了训练梯度消失和梯度爆炸的问题。

(4)LSGAN模型^[5]。LSGAN指出使用JS散度并不能拉近真实分布和生成分布之间的距离,使用最小二乘可以将图像的分布尽可能的接近决策边界,因此LSGAN使用了最小二乘损失函数代替了GAN的损失函数。

(5)BEGAN模型^[6]。BEGAN提出了一种新的判别生成样本真伪的方法,用判别器判别生成样本分布和真实样本分布之间的误差距离,当误差距离较小即生成样本分布与真实样本分布相似时,说明生成器生成的样本接近真实样本。因此可以用多种类型的GAN结构和损失函数做训练。

(6)Info GAN模型^[7]。GAN的生成模型输入的是一个连续的噪声信号并且没有任何约束,导致GAN无法利用这个噪声信号,并将噪声信号的具体维度与数据的语义特征对应起来,因此不是一个可解释的表达。而Info GAN将噪声信号分解为一个不可压缩的噪声和一个可解释的隐变量,使得噪声数据可解释。

(7)CatGAN模型^[8]。CatGAN的判别器模型以较大确信度划分真实样本,而且以较大确信度对真实样本分类到现有的一个类别,但对于生成样本的不确信度却比较大。CatGAN采用数据的熵来作为衡量标准,即用熵值得大小来衡量确信度,熵值越大,不确信度越大,熵值越小,确信度越大。

3 基于GAN的应用

GAN作为一个生成模型最直接的应用就是生成图片,可以对机器学习数据不足的情况下有极大的帮助。除此之外,GAN还被用于迁移学习、强化学习以及图像修复当中。因此基于GAN有着广泛的应用,本节主要结束基于GAN的几个典型应用。

(1)图像超分辨率重建,就是将一个低分辨率的图像,转换为一个高分辨率的图像。传统的图像超分辨率重建方法要对一些高频细节建模,而基于GAN的图像超分辨率重建方法大大简化了这一过程。GAN通过生成模型产生高分辨率的图像来迷惑判别模型,随着判别模型的学习加深,生成模型就能生成更高分辨率的图像。与传统方法比,GAN生成的图像具有更丰富的细节。

(2)图像修复,就是将图像中缺失的部分补全。目前对于图像修复的工作主要是对人脸的修复,现实环境中由于人体姿态转动对人脸的遮挡,光线强弱对人脸成像的影响,需要对人像进行补全。文献[9]使用WGAN网络将图像中遮挡部分生成上下文一致的补全图像,Yeh等人^[10]提出了能够对大范围图像补全的GAN模型。因此GAN对图像修复工作有着很好作用。

(3)迁移学习,就是将所学习到的知识从一个场景迁移到另一个场景。对于一些缺少大量的数据样本或者大量的标注可用数据,可以通过GAN中的对抗方法对齐两个领域的分布,

以拥有大量标注数据的领域对目标域实现迁移学习。文献[11]提出了一种能在多个领域互相生成图像的GAN模型,可以实现人脸属性迁移。

(4)强化学习,就是通过累积悬赏获取到需要完成目标任务的最优方法。Jonathan Ho等人^[12]将GAN引入到强化学习当中,用生成器生成动作序列,判别器用来判别这个动作是否为专家动作。通过GAN的不断训练,使得生成器生成的动作得到不断的优化,从而接近专家动作。基于GAN的强化学习的判别器输出相当于奖赏,避免了人为设定悬赏。

4 总结与展望

GAN作为生成模型被广泛的应用于深度学习中数据的生成、图像的补全、图像的高分辨率重建、迁移学习、强化学习等领域中。除此之外,GAN还被用于文本到图像生成、图像翻译、交互式图像生成等方面。本文对GAN及其衍生模型做了介绍,虽然各个模型有一定的缺陷,但随着以后的改进,GAN凭借其优异的性能一定会应用到更多的场景当中。

参考文献:

- [1] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// International Conference on Neural Information Processing Systems, 2014.
- [2] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer Science, 2015.
- [3] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. 2017.
- [4] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved Training of Wasserstein GANs[J]. 2017.
- [5] Mao X, Li Q, Xie H, et al. Least Squares Generative Adversarial Networks[J]. 2016.
- [6] Berthelot D, Schumm T, Metz L. BEGAN: Boundary Equilibrium Generative Adversarial Networks[J]. 2017.
- [7] Chen X, Duan Y, Houthoofd R, et al. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets[J]. 2016.
- [8] Springenberg J T. Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks[J]. Computer Science, 2015.
- [9] 姚乃明, 郭清沛, 乔逢春, 等. 基于生成式对抗网络的鲁棒人脸表情识别[J]. 自动化学报, 2018, 44(5):100-112.
- [10] Yeh R A, Chen C, Lim T Y, et al. Semantic Image Inpainting with Deep Generative Models[J]. 2016.
- [11] Choi Y, Choi M, Kim M, et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation[J]. 2017.
- [12] Ho J, Ermon S. Generative Adversarial Imitation Learning[J]. 2016.

【通联编辑:代影】