

生成式对抗网络研究综述

文/张倩宇 宋彩芳

摘要

近年来,随着生成式对抗网络的迅速发展,尤其图像处理领域表现突出,本文主要介绍了生成对抗网络的基本模型和衍生模型,以及在图像处理领域的应用,包括图像风格迁移、人机交互领域、视频帧预测、文本与图像的相互生成等应用。

【关键词】 生成式对抗网络 图像风格迁移 人机交互领域 视频帧预测 文本与图像相互生成

1 生成对抗网络模型

1.1 GAN模型

GAN的基本思想就是两个网络互相博弈,一个生成器网络G和一个判别器网络D。生成器网络G和判别器网络D本质上都是函数。生成器G的目标是学习到输入数据的分布从而生成非常真实的图像,而判别器D的目标是正确辨别出真实图片和G生成的图片直接的差异。生成器G和判别器D都有各自的网络结构和不同的输入,其中生成器G的输出,即生成样本也是D的输入之一,而判别器D则会为G提供梯度进行权重的更新。这两个网络通过不断的改善和优化来提高自身网络的生成能力和判别能力。

GAN是作为一种图像生成建模技术,在图像生成任务上有广泛的应用,主要在于:

- (1) 能训练有效的无条件图像生成器;
- (2) 一种能用于不同域之间无监督图像转换的方法;
- (3) 一种有效的感知式图像损失函数(如Pix2pix)。

GAN有明显的优势,固然也有一些关键的劣势:

- (1) GAN很难训练,具体表现包括训练过程非常不稳定、训练容易导致模式崩溃以及对超参数极为敏感;
- (2) GAN有模式丢失问题—只能建模目标分布的某些模式而非所有模式。

1.2 GAN模型的衍化

1.2.1 DCGAN

DCGAN是用一定程度改进的CNN取代了GAN中的生成网络和判别网络,该算法将有监督学习中的CNN和无监督学习中的GAN结合,为GAN的训练提供了很好的网络结构,同时提高了训练过程的稳定性和生成结果的质

量。该模型还实现了通过特征学习或是特征向量计算得到一个稳定的向量来进行特定变换。

1.2.2 WGAN及WGAN-GP

WGAN从损失函数的角度对GAN进行了改进,主要解决了一下三方面的问题:

- (1) 不再需要精心平衡生成器和判别器的训练程度,且效果更稳健,不依赖于过于精心设计的结构;
- (2) 基本解决了模型崩塌问题,确保了生成样本的多样性;
- (3) 训练过程中有了一个像交叉熵、准确率这样的数值来指示训练的进程,这个数值越小代表GAN训练得越好,代表生成器产生的图像质量越高。

WGAN-GP是针对WGAN存在的问题提出来的,使用了梯度惩罚的替换,采用Lipschitz约束方法代替WGAN中的加权剪枝。实验结果表明此方法生成的样本质量高于WGAN,提供稳定的训练,几乎不需要超参数调整,并成功训练各种GAN架构进行生成任务。

1.2.3 EBGAN

EBGAN与WGAN基于样本间的距离度量不同,EBGAN是从能量模型的角度对GAN进行了改进。该模型将判别器看做一个能量函数,在真实数据范围内该能量函数的能量值会减小,在非真实数据(即它认为是生成数据)范围内能量值会增大。这样一来生成器的目标就是产生能量值足够小的样本,而判别器则是以对生成的样本赋高的能量值以及对真实样本赋低的能量值为目的。EBGAN的意义在于它给予GAN一种不同于其他模型通过距离度量定义损失函数的能量模型的定义,也为我们开辟了一条新的道路,GAN模型具有更宽泛的结构和更多样的损失函数类型的训练。

1.2.4 其他模型

Bojanowski等提出的GLO是将训练图像嵌入到一个低维空间中,并在该嵌入向量输入到一个联合训练的深度学习生成器时重建它们。GLO的优势有:

- (1) 无模式丢失地编码整个分布;
- (2) 学习得到的隐含空间能与图像的形义属性相对应,即隐含编码之间的欧几里得距离对应于形义方面的含义差异。

但GLO有一个关键缺点,即没有一种从嵌入空间采用新图像的原则性方法。尽管GLO的提出者建议用一个高斯分布来拟合训练图像的隐编码,但这会导致图像合成质量不高。

Li and Malik提出的IMLE训练生成模型的方式是从一个任意分布采样大量隐含编码,使用一个训练后的生成器将每个编码映射到图像域中并确保对于每张训练图像都存在一张相

近的生成图像。IMLE的采样很简单,而且没有模式丢失问题。类似于其它最近邻方法,具体所用的指标对IMLE影响很大,尤其是当训练集大小有限时。虽然经典的Cover-Hart结果启示我们最近邻分类器的误差率渐进地处于贝叶斯风险的二分之一范围内,但当我们使用有限大小的示例样本集时,选择更好的指标能让分类器的表现更好。当使用L2损失直接在图像像素上训练时,IMLE合成的图像是模糊不清的。

尽管wasserstein距离极大地提升了GAN的效果,但仍在理论上存在训练不稳定和模式丢失的问题。Facebook的两位研究者融合了两种非对抗方法的优势,并提出了一种名为生成式隐含最近邻GLANN(Generative Latent Nearest Neighbors)的新技术。该方法首次使用了GLO来嵌入训练图像,从而克服了IMLE的指标问题。由GLO为隐含空间引入的线性特性能让欧几里得度量在隐含空间Z中具有形义含义。GLANN方法集中了IMLE和GLO的双重优势:易采样、能建模整个分布、训练稳定且能合成锐利的图像。该方法经过使用已确立的指标评估发现其显著优于其它的非对抗方法,同时其表现也比当前的基于GAN的模型更优或表现相当。GLANN也在高分辨率图像生成和3D生成上得到了出色的结果。

2 基于生成对抗网络的图像领域的应用

2.1 图像风格迁移

图像风格迁移就是把图像从一种风格转换到另一种风格。深度学习最早是基于CNN框架尝试进行图像风格迁移。该模型有两大劣势:

- (1) 对训练样本要求较高,需要在成对的数据上训练,实际我们并没有这样完美的成对的训练数据;
- (2) 训练速度慢,每一次生成都相当于重新训练一个模型。后来,GAN在图像风格迁移方面有了突出表现,用较小的数据集完成GAN的无监督训练,然后以此来生成更多与数据样本相类似的图像而不是去耗费人力物力采集样本,可节约大量成本。

2.2 人机交互领域

Santana等人实现了利用GAN的辅助自动驾驶。首先,生成与真实交通场景图像分布一致的图像,然后,训练一个基于循环神经网络的转移模型来预测下一个交通场景。另外,GAN还可以用于对抗神经机器翻译,将神经机器翻译作为GAN的生成器,采用策略梯度

<< 下转 14 页

无线传感器网络的特点和应用

文/潘玉兰 刘广聪

摘要

随着信息化技术的不断发展,无线传感器网络成为最主要的无线网络技术之一,同时也成为了研究热点。无线传感器网络应用非常广,在军事、医疗、交通、家庭等应用领域都有着极高的应用价值。本章首先介绍了无线传感器网络的特点,其次对无线传感器网络的关键技术进行讲解,然后对无线传感器网络的主要应用进行了介绍,最后本文对无线传感器网络的研究进行总结和展望。

【关键词】无线传感器网络 WSN 特点 关键技术 主要应用

无线传感器网络(Wireless Sensor Network, WSN)是一组传感器节点以自组织的方式构成的无线网络,这些传感器节点由处

理器、存储器、接收器、感知单元、电池组成,使得传感器节点具有感知、计算和通信能力。因此无线传感器网络借助于网络中形式多样化的传感器协助地实时感知和采集周边环境众多的信息,并对这些信息进行处理以使人们在任何时候、绝大多数地点和多种环境条件下获取大量详实而可靠的信息。

1 无线传感器网络的特点

相比于其他传统的无线通信网络,本章的无线传感器网络(WSN)有着其自身显著的特点,其主要特点如下所示。

1.1 大规模

在日常应用领域中,比如大面积森林防火监测,为了获取精确的森林监测信息,在森林监测区域会投入大量传感器节点,可能数量达到成千上万,甚至更多。传感器网络大规模分为两种:第一种就是在监测面积空间比较大

的监测领域,需要部署大量的传感器节点;第二种就是在监测面积空间比较小的监测领域,需要密集部署传感器节点。

1.2 自组织性

在无线传感器网络中,通常传感器节点是随机放置在没有任何基础结构的地方,每一个传感器节点的位置事先是不清楚的,节点之间的邻居相互关系也是预先不知道的。例如使用飞机从天空随机抛洒出成千上万的传感器到大面积森林,或随机部署人为不可达或险恶监测区域,这就需要无线传感器节点具有自组织的能力,能够自动配置和管理。通常其实现的方法是通过网络协议和拓扑控制机制自动形成转发监测数据的多跳无线网络系统。

1.3 动态性

通常无线传感器网络的网络拓扑随着如下因素改变,比如由于传感器节点电能耗尽所

<< 上接 13 页

方法训练判别器,通过最小化人类翻译和神经机器翻译的差别生成高质量的翻译。

2.3 视频帧预测

Mathieu 等人首先将 GAN 训练应用于视频帧预测,即生成器根据前面一系列帧,生成视频最后一帧,判别器对该帧进行判断。除最后一帧外的所有帧都是真实的图片,这样的好处是判别器能有效利用时间维度的信息,同时也有助于使生成的帧与前面的所有帧保持一致。实验结果表明,通过对抗训练生成的帧比其他算法更加清晰。

此外,Vondrick[2]等人在视频领域也取得了巨大进展,他们生成 32 帧分辨率为 64×64 的逼真视频,描绘的内容包括高尔夫球场、沙滩、火车站以及新生儿。经过测试,20% 的标记员无法识别这些视频的真伪。

2.4 文本与图像的相互生成

GAN 能够在图像数据集上经过训练生成与真实分布相似的分布,如 GAN 在人脸数据集上经过训练,生成分辨不出真假的人脸图像。这些都是直接学习特征再生成分布。

从文本到图像的生成难度大且限制多,相反,根据给定图像生成文字描述比较容易,

经过训练,机器能够简单描述图像内容,但如何让机器像人类一样看图说话,写出文章?文献[3]提出一种半监督的段落生成框架,它通过对局部语义区域进行推理和利用语言知识合成语义连贯的段落描述。文中提出的循环主题转换生成对抗网络(RTT-GAN)构建了一个结构化段落生成器和多级段落识别器之间的对抗框架。段落发生器通过在每个步骤中引入基于区域的视觉和语言注意机制来循序地产生句子。

3 总结

GAN 由于其强大的生成能力,正被广泛地研究与关注。目前,在学术领域,GAN 训练指标,模式坍塌以及模型的生成能力的可解释性正受广泛的关注。最后,在拓展应用领域,由于生成的图片有较高的噪音,如何提高数据(图片或视频)也是研究的一大热点。此外,GAN 作为一种深度模型,也是解决自然语言处理天然的良好模型。如何将 GAN 应用在 NLP 领域也是下一步要解决的问题。

参考文献

- [1] Yedid Hoshen, Jitendra Malik. Non-Adversarial Image Synthesis

with Generative Latent Nearest Neighbors[EB/OL], 2018-12-21. <https://arxiv.org/abs/1812.08985.pdf>.

- [2] Vondrick C, Torralba A. Generating the future with adversarial Transformers[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2992-3000.

- [3] Liang X D, Hu Z T, Zhang H, et al. Recurrent topic-transition GAN for visual paragraph generation[EB/OL]. 2017-03-23 [2018-01-08]. <https://arxiv.org/pdf/1703.07022.pdf>.

作者简介

张倩宇(1993-),女,山西省吕梁市人。在读硕士。研究方向为图像处理,深度学习。宋彩芳(1980-),女,山西省长治市人。讲师,博士,北京工业大学毕业。研究方向为模式识别,图像处理,深度学习。

作者单位

山西师范大学 数学与计算机科学学院 山西省临汾市 041000