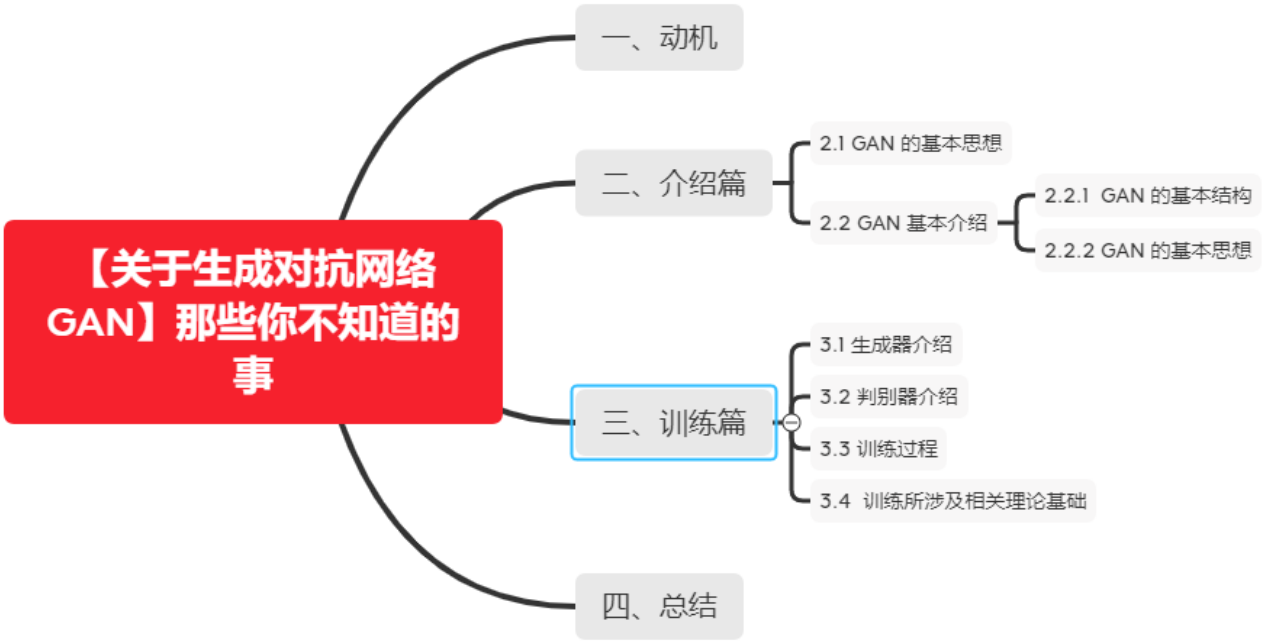
【关于生成对抗网络GAN】那些你不知道的事

作者：杨夕

项目地址：https://github.com/km1994/nlp_paper_study

个人介绍：大佬们好，我叫杨夕，该项目主要是本人在研读顶会论文和复现经典论文过程中，所见、所思、所想、所闻，可能存在一些理解错误，希望大佬们多多指正。



一、动机

之前我们提到玻尔兹曼机(Boltzmann machine)，波尔茨曼机作为一种基于能量函数的概率模型，因为能量函数比较复杂，所以存在较多的限制。虽然受限玻尔兹曼机(Restricted Boltzmann machine) 针对该问题，对能量函数进行进一步简化，即假设网络中仅有隐藏变量与观察变量的连接，而观察变量将没有连接，隐藏变量间也没有连接，且隐藏变量可用 n_h 个二进制随机变量表示，但是仍然存在限制问题。同时，该过程应用了马尔科夫链，导致计算成本较高。针对上述问题，Ian Goodfellow 于 2014 年提出了生成对抗网络（Generative Adversarial Network，GAN）模型，GAN 作为一类在无监督学习中使用的神经网络，有效地避免了马尔科夫链以及减低了波尔茨曼机所存在的限制问题，以至于在按文本生成图像、提高图片分辨率、药物匹配、检索特定模式的图片等任务中 GAN 的研究如火如荼。大牛Yann LeCun甚至评价GAN为 “adversarial training is the coolest thing since sliced bread”。

本文将通过一个简单的例子（发论文问题）向读者深入浅出的介绍 GAN 原理及其应用。

二、介绍篇

2.1 GAN 的基本思想

作为生成模型中的一种，生成对抗网络（Generative Adversarial Network，GAN）模型的训练过程可以被视为两个网络互相博弈的过程。下面我们将举一个简单的例子解释 GAN 的基本思想。

> 假设你是一门研究生，你想尽快地将实验结果写成一篇论文发表。 于是在每一次做完实验并写完初稿之后，都会跟你的导师进行沟通：
>
> 你：boss，我实验结果出来，我想发论文
>
> 导师：（瞄了瞄你的实验结果之后） ... 算了吧
>
> （你通过跟其他论文的实验结果进行比较，发现自己的实验结果还偏低，于是，你又调整了实验参数，重新进行实验）
>
> 你：boss，我实验结果提高了，我想发论文
>
> 导师：...（瞄了瞄你的论文初稿之后）嗯 还有所欠缺
>
> （你通过跟其他论文进行比较，发现自己写的论文初稿在表达方面还有所不足）
>
> ...
>
> 你：boss，我想发论文
>
> 导师：...（仔细看了看你的论文之后）嗯 可以试一试
>
> （通过这样不断的修改和被拒绝，你的论文最终获得了导师的赞赏与肯定）

通过上面的例子，大家应该对 GAN 的思想有一个比较感性的认识了吧，下面我们可以进一步对 GAN 的基本结构和思想进行介绍。

2.2 GAN 基本介绍

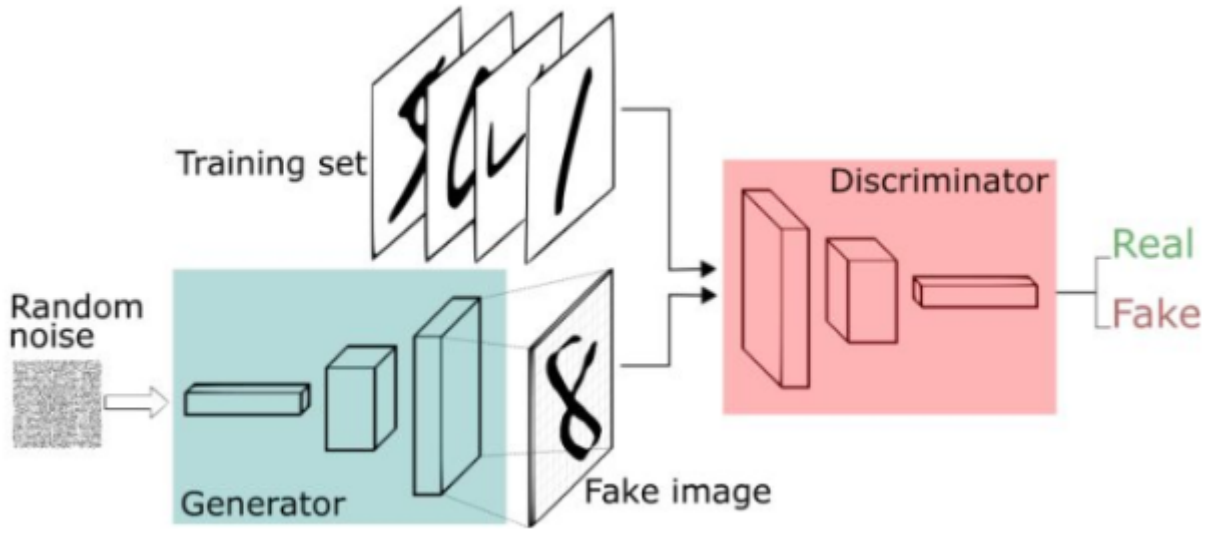
2.2.1 GAN 的基本结构

GAN 的主要结构包括一个生成器 G（Generator）和一个判别器 D（Discriminator）。

在上面的例子中，研究生相对于生成器。在一开始的时候，他只是一个什么都不懂的初学者，为了能让该研究生发出好的 paper，需要给他裴蓓一个导师来指导他做实验写论文，并告诉他 paper 面前的质量，通过反复的修改和被拒绝，paper 最终达到了可以投稿的标准，而这个导师就相当于生成对抗网络 GAN 中的判别器。

2.2.2 GAN 的基本思想

生成对抗网络 GAN 主要包含两个模块：生成器 G（Generator）和一个判别器 D（Discriminator）。生成对抗网络 GAN 中所描述的对抗，其实就是指生成网络与判别网络之间的相互对抗。以下图为例：



生成器模型（上图中蓝色部分 Generator）的主要工作就是学习真实图片集数据，从而使自己生成的图片更加接近与真实图片，以到达“以假乱真”，也就是“欺骗”判别器。

判别器模型（上图中红色部分 Discriminator）的主要工作就是从图片集中找出生成器所生成的图片，并区分该图片与真实图片的差异，以进行真假判别。

在整个迭代过程中，生成器不断的生成越来越逼真的图片，而判别器不断额努力鉴别出图片的真假。该过程可以视为两个网络互相博弈的过程，随着迭代次数的增加，最终两者将会趋于平衡，也就是说生成器能够生成出和真实图片一模一样的的图片，而判别器已经很难从图片集中辨别出生成器所生成的假图片了。也就是说，对于图片集中的每一张图片，判别器都给出接近 0.5 的概率认为该图片是真实的。

三、训练篇

3.1 生成器介绍

生成器模型的任务：首先需要将一个 n 维向量输入生成器模型，然后输出一个图片像素大小的图片（这里，生成器模型可以是任意可以输出图片的模型，如全连接神经网络，反卷积神经网络等）。

> 注：输入向量：携带输出的某些信息，这些信息可以是手写数字为数字几，手写的潦草程度等。由于这里我们对于输出数字的具体信息不做要求，只要求其能够最大程度与真实手写数字相似（能骗过判别器）即可。所以我们使用随机生成的向量来作为输入即可，这里面的随机输入最好是满足常见分布比如均值分布，高斯分布等。

3.2 判别器介绍

判别器模型的任务：主要能够辨别输入的图片的真假都可以作为判别器。

3.3 训练过程

前面分别介绍了生成器和判别器的任务，在这一节，我们将主要介绍生成对抗网络的训练过程，其基本流程如下：

step 1: 初始化：对判别器 D 的参数 θ_d 和生成器 G 的参数 θ_g；

step 2: 生成器"伪造"生成样本：首先，从真实样本中采样 m 个样本 $\left\{x^1, x^2, \ldots x^m\right\}$；然后,从先验分布噪声中采样 m 个噪声样本 $z^1, z^2, \ldots, z^m$；接下去，利用生成器"伪造" m 个新样本 $\left\{\tilde{\boldsymbol{x}}^1, \tilde{\boldsymbol{x}}^2, \ldots, \tilde{\boldsymbol{x}}^m\right\}$；最后，固定生成器 G。

step 3：判别器"鉴别"生成样本：通过对判别器 D 进行训练，以让它尽可能准确的"鉴别"出生成样本。

step 4: "欺骗"判别器：循环更新判别器 k 次之后，再利用较小的学习率来更新一次生成器的参数。使得判别器已经很难从样本集中辨别出生成器所生成的生成样本了。也就是说，对于样本集中的每一个样本，判别器都给出接近 0.5 的概率认为该样本是真实的。

> 注：为什么是先训练判别器再训练生成器呢？
>
> 以上面的导师和学生的例子吧，学生（生成器）要写出一篇好的 paper （生成样本），那么就需要有一个能够较好的区分好 paper （真实样本）和坏 paper （生成样本）的好导师（判别器）之后，才能指导学生（生成器）如何对 paper （生成样本）进行优化。

3.4 训练所涉及相关理论基础

前面已经对生成对抗网络进行介绍，接下去，我们将从理论基础方面介绍生成对抗网络的训练过程。

首先，需要从优化目标函数开始介绍，其表达式如下所示：

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

对于判别式而言，其主要用于区别样本的真伪，所以可以视为是一个二分类问题，上式中所使用的 $V(G, D)$为二分类问题中常见的交叉熵损失。公式如下所示：

$$H(p, q) := -\sum_i p_i \log q_i$$

> p_i 和 q_i 为真实的样本分布和生成器的生成分布。

对于生成器 G 而言，为了尽可能欺骗 D，所以需要最大化生成样本的判别概率 $D(G(z))$，即最小化 $log(1 - D(G(z)))$。

> 注意：$log(D(x))$ 一项与生成器 G 无关，所以可以忽略。

实际训练过程中，生成器和判别器采用交替训练的方式进行。因为对于生成器，其最小化为 $\max_D V(D,G)$，即最小化 $ V(D, G) $的最大值。所以为了保证 $ V(D, G) $ 取得最大值，需要对判别器迭代训练 k 次，然后再训练一次生成器。

当生成器 G 固定时，我们可以对 $V(D,G)$ 求导，求出最优判别器 $D*(x)$：

$$D^*(x) = \frac{p_g(x)}{p_g(x) + p_{data}(x)}$$

把最优判别器代入上述目标函数，可以进一步求出在最优判别器下，生成器的目标函数等价于优化 $p_{data}(x)$, $p_g(x)$ 的 JS 散度（JSD, Jenson Shannon Divergence）。

可以证明，当 G，D 二者的 capacity 足够时，模型会收敛，二者将达到纳什均衡。此时，$p_{data}(x)$= $p_g(x)$，判别器不论是对于 $p_{data}(x)$ 还是 $p_g(x)$ 中采样的样本，其预测概率均为 1/2，即生成样本与真实样本达到了难以区分的地步。

通过上述min max的博弈过程，理想情况下会收敛于生成分布拟合于真实分布。

四、总结

本文首先，通过以一个学生发 paper 的 example 的方式引入了生成对抗网络；然后，并进一步介绍了生成对抗网络的框架和思想，中生成器和判别器；最后，通过介绍生成对抗网络的训练过程，以引入生成对抗网络的训练公式。

参考资料

1. 通俗理解生成对抗网络GAN
2. 白话生成对抗网络 GAN，50 行代码玩转 GAN 模型！【附源码】
3. 万字综述之生成对抗网络（GAN）
4. 生成对抗网络原理与应用：GAN如何使生活更美好
5. 玻尔兹曼机、生成随机网络与自回归网络——深度学习第二十章（二）
6. 生成对抗网络(GAN)相比传统训练方法有什么优势?
7. 火热的生成对抗网络(GAN),你究竟好在哪里