重要概念

生成网络

使用现有数据生成新数据

核心任务是：从随机生成的由数字构成的向量（潜在空间，latent space）中生成数据（图，视音频，文本）

在构建生成网络时需明确该网络目标，例如生成图像

KL散度

又称 相对熵，用于判定两个概率分布之间的相似度，它可以测量一个概率分布p 相对于另一个概率分布q 的偏离

$$D_{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx$$

若 p(x) 和 q(x) 处处相等，则KL散度为0，达到最小值

KL散度具有不对称性，因此不用于测量两个概率分布之间的距离，也不用作距离的度量（metric）

JS散度

又称 信息半径（information radius, IRaD），或 平均值总偏离（total divergence to the average）

可用来测量两个概率分布之间的相似度，它基于KL散度，但具有对称性，可用来测量两个概率分布之间的距离

对 JS散度 开平方即是 JS距离，是一种距离度量

$$D_{JS}(p||q) = \frac{1}{2} D_{KL}(p||\frac{p+q}{2}) + \frac{1}{2} D_{KL}(q||\frac{p+q}{2})$$

纳什均衡

纳什均衡是网络训练中希望达到的状态，它描述了一种在非合作博弈中可以达到的特殊状态

其中每个参与者都试图基于对其他参与者行为的预判，选择使自己获益最多的最佳策略，最终形成一种局面：所有参与者都基于其他参与者的选择，采取对自己来说最佳的策略，此时已经无法通过改变策略来获益

一个著名的例子是囚徒困境

目标函数

生成网络和判别网络各有目标函数，训练过程中分别试图最小化各自大的目标函数

GAN最终的目标函数如下：

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

D(x)：判别网络模型； G(z)：生成网络模型； p(x)：真实数据分布； p(z)：生成数据分布； E：期望输出

评分算法

GAN目标函数不是 均方误差（MSE）或 交叉熵（cross entropy）这样的确定函数，而是在训练过程中习得

1. Inception Score

应用最广泛的GAN评分算法，可用于测量图片质量和多样性

它使用一个在Imagenet上预训练过的Inception V3网络分别提取真实图像和生成图像的特征

$$IS(G) = \exp(\mathbb{E}_{\chi \sim p_g} D_{KL}(p(y|\chi)||p(y)))$$

p_g：一个概率分布； x~p_g：x是该概率分布中的一个抽样； p(y|x)：条件类别分布； p(y)：边缘类别分布

计算步骤：

1. 首先从模型生成的图像抽取N个样本，记作（x^i）
2. 计算边缘类别分布：

$$p(y) = \int_{\chi} p(y|\chi) p_g(\chi)$$

3. 计算KL散度以及期望值，得到IS

IS越高，说明模型质量越好

缺点是：模型对于每个类别只生成一张图片，其IS仍可以很高，但这样的模型缺乏多样性

2. Frechet Inception Distance

FID优于IS之处在于对噪音的抵抗力较好，且可以更好地测量图像的多样性

$$FID = ||\mu_r - \mu_g||^2 + T_r(\sum_r + \sum_g - 2(\sum_r \sum_g)^{\frac{1}{2}})$$

计算步骤：

1. 首先抽取Inception网络的一个中间层的特征映射，构建一个多元正态分布来学习这些特征映射的概率分布
2. 使用该多元正态分布的 均值u 和 协方差∑ 来计算FID

FID越低，说明模型质量越好，其生成多样化、高质量的图像能力就越强，完美的生成模型的FID应该为0

训练GAN的问题

1. 模式塌陷

模式塌陷是指生成网络所生成的样本之间差异不大，生成的所有样本几乎都相同

有一些概率分布是多峰的（multi-modal），构造十分复杂，数据可能是通过不同类型的观测得来的，因此样本中可能会暗含一些细类，每个细类下样本之间比较相似，这会导致数据的概率分布出现多个峰，每个峰对应一个细类，如果数据概率分布是多峰的，GAN可能会出现模式塌陷问题，无法成功构建模型

解决方法：

1. 针对不同的峰训练不同的GAN
2. 使用多样化数据训练GAN

2. 梯度消失

在反向传播过程中，梯度从最后一层反向流动到第一层，并且会越来越小，有时梯度过小会导致前几层的学习速度非常慢，或者根本无法学习，在这种情况下，梯度无法改变前几层的权重值，所以网络的前几层的训练效果没有任何效果，该问题称作为梯度消失

如果使用基于梯度的优化方法（通过计算参数值上的小幅变动对神经网络输出的影响来优化参数）训练更大的神经网络，问题会更严重；使用 sigmoid 和 tanh 激活函数也会存在梯度消失问题

解决方法：

1. 使用ReLU、LeakyReLU、PReLU等激活函数
2. 使用批归一化，对隐藏层接收的输入先进行归一化，然后传递给隐藏层

3. 内部协变量转移

输入数据的概率分布发生改变之后，隐藏层会试图适应新的概率分布，训练速度因此放缓，需要很长时间才会收敛到全局最小值，而神经网络输入数据概率分布和该网络之前接触的数据概率分布差异过大是问题根源

解决方法：

1. 批归一化和其他归一化技术

解决GAN训练稳定性问题

GAN可能会出现训练不稳定的问题，严重时会导致永远无法在数据上收敛

1. 特征匹配

在GAN的训练过程中，判别网络的目标函数需要最大化，生成网络的目标函数需要最小化，这样的目标函数存在一些严重的缺陷，例如没有考虑生成数据和真实数据的分布特征

特征匹配引入了一种新的目标函数来提高GAN的收敛能力，便于生成网络生成在分布特征上和真实数据更为接近的数据

在特征映射技术中，判别网络不再输出二元标签，而改为输出某个中间层对于输入数据的 “激活映射”，也称 “特征映射”，这样可以训练判别网络学习真实数据的分布特征，并且使用这些特征来区分真实数据和虚假数据

f(x)：判别网络某中间层对真实数据的激活映射或者特征映射

f(G(z))：判别网络某中间层对生成网络生成数据的激活映射或者特征映射

新的目标函数：

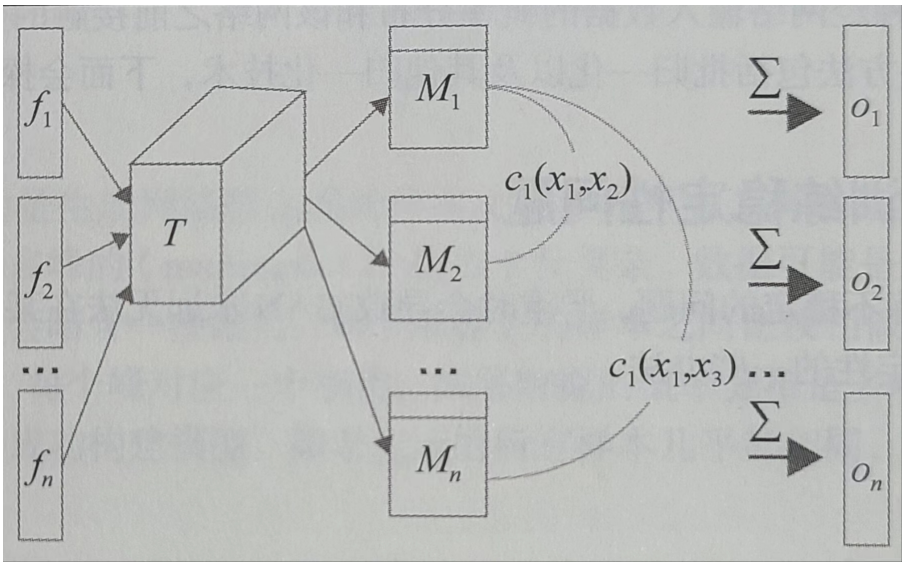
$$|| \mathbb{E}_{x \sim p_{data}} f(x) - \mathbb{E}_{z \sim p_z(z)} f(G(z)) ||_2^2$$

使用该目标函数可以获得更好的结果，但不保证收敛

2. 小批量判别

小批量判别有助于避免峰坍塌，提高训练稳定性：训练GAN过程中，如果判别网络接收的输入图像彼此不相关，它们的梯度之间无法产生联系，判别网络就无法学习如何区别生成网络生成的不同种类的图像，从而导致模式塌陷

小批量判别是一个多步骤过程：



1. 抽取样本的特征映射，然后乘以张量$T \in R^{A \times B \times C}$，得到矩阵$M_i \in R^{A \times B}$
2. 计算M_i到各行之间的L1距离：

$$c_b(x_i, x_j) = \exp(-||M_{i,b} - M_{j,b}||_{L1}) \in \mathbb{R}$$

3. 对于某个输入数据x_i，计算步骤2得到的所有距离之和：

$$o(x_i)_b = \sum_{j=1}^{n} c_b(x_i, x_j) \in \mathbb{R}$$

3. 将$o(x_i)$和$f(x_i)$拼接起来，作为输入传递给神经网络下一层：

$$o(x_i) = [o(x_i)_1, o(x_i)_2, \ldots, o(x_i)_B] \in \mathbb{R}^B$$
$$o(X) \in \mathbb{R}^{n \times B}$$

$f(x_i)$：判别网络某个中间层对于第i个样本的特征映射

$T \in R^{A \times B \times C}$：一个三维张量，用于和$f(x_i)$相乘

$M_i \in R^{A \times B}$：上述两变量相乘生成的矩阵

$o(x_i)$：对于某样本x_i，计算M_i所有行之间L1距离之和

3. 历史平均

历史平均是计算历史参数的平均值，然后将该值分别加入生成网络和判别网络的成本函数中

$$|| \theta - \frac{1}{t} \sum_{i=1}^{t} \theta[i] ||^2$$

θ[i]：全部参数在某一时刻 i 的取值

该方法可以提高GAN的稳定性

4. 单面标签平滑

- 对抗样本：
  - 对抗样本是指在数据集中通过故意添加细微的干扰所形成的输入样本，导致模型以高置信度给出一个错误的输出
  - 在精度达到人类水平的神经网络上通过优化过程故意构造数据点，其上的误差率接近100%，模型在这个输入点 x' 的输出与附近的数据点 x 非常不同，在许多情况下，x' 与 x 非常近似，人类观察者不会察觉原始样本和对抗样本之间的差异，但是网络会作出非常不同的预测
  - 这些对抗样本的主要原因之一是过度线性。如果一个线性函数具有许多输入，那么它的值可以非常迅速地改变
  - 使用对抗训练（在对抗样本上训练模型）可以通过鼓励网络在训练数据附近的局部区域恒定来限制这一高度敏感的局部线性行为

在先前的设定下，GAN分类器的标签只可取0或1，很容易受到对抗样本的问题的影响，对抗样本是一种特殊的输入数据，如果在输入数据中叠加对抗样本，原本可以进行正常分类的神经网络会产生错误的分类结果

而标签平滑技术可以为判别网络提供平滑后的标签，如0.9（真），0.8（真），0.1（假），真图像和假图像的标签值都需要进行平滑处理，标签平滑有助于降低GAN出现对抗样本的风险

5. 批归一化

批归一化技术是将特征向量归一化，使其均值为0，方差为1，该技术可提高学习过程中的稳定性，以及缓解权重值初始化效果差的问题

将该技术作为预处理步骤应用于神经网络的隐藏层，有助于缓解内部协变量转移的问题

批归一化需要应用于所有隐藏层，而不止是输入层

6. 实例归一化

批归一化是一批数据的整体信息进行归一化

实例归一化有所不同，在对一个特征映射进行归一化时只使用该特征映射的信息