

GAN评价指标

1. Inception score

1.1 Inception score 基本思想

基本思想 Inception score 采用图片类别分类器来评估生成质量

对生成图片清晰程度: 对生成图片 x 输入到 Inception Net-v3 产生一个 1000 维输出向量 y .

其中每一维代表 x 属于某一类图像的概率. 对于清晰图片, y 的某一维

应趋近 0, 其余接近 0. 即 $y: P(y|x)$ 的熵很小.

对生成图片多样性: 对于所有生成图片应均匀分布在所有类别中, 即生成 10000 张图片, 对 1000 类别

每一类应生成 10 张图片, 即趋近均匀分布即 $P(y) = \sum P(y|x_i)$ 趋近 1.

1.2 Inception Score 公式

$$IS(G) = \exp(E_{x \sim P} D_{KL}(P(y|x) || P(y)))$$

KL 散度 边缘概率
对于 x 所属类别的概率

25 值越大越好.

1.3 缺陷问题

① 数据集问题: 不能在一个数据集上训练分类模型来评估另一个数据集生成模型

2. FID (Frechet Inception Distance)

2.1 基本思想

直接考虑生成数据和真实数据在 Feature 层次的距离

使用 Inception Net-v3 全连接层前 2048 维作为图片的 feature.

2.2 FID 公式:

FID 为反应生成图片与真实图片的距离, 越小越好 (衡量两个正态分布距离)

$$FID = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}})$$

生成图片特征均值 生成图片协方差矩阵
真实图片特征均值 正 真实图片协方差矩阵

2.3 优缺点

优: 生成模型训练集可 > 与 Inception Net-v3 不同

缺: FID 为衡量多元正态分布的距离但提取图片特征不一定为多元正态分布

3. Perceptual Path Length

主要为 style GAN 提出用 Perceptual path 来衡量生成器能否把不同图片特征分离出来

3.1 基本思想:

给出两个随机噪声 z_1, z_2 , 采用微分思想把 z_1, z_2 插值路径细分为多个小段, 求每小段长度再平均

3.2 公式:

$$PPL = E[\frac{1}{L} \sum_{i=1}^L d(G(\text{slerp}(z_1, z_2, t)), G(\text{slerp}(z_1, z_2, t + \epsilon)))]$$

训练集 VGG 平滑线性插值 $t \sim U(0,1)$ 插值参数服从均匀分布
细分段 生成器

4. MMD (Maximum Mean Discrepancy)

4.1 主要思想:

为希尔伯特空间对两个分布的差别的度量, 可使用 MMD 度量训练集 P_{data} 与生成数据

P_g 的距离作为 GAN 衡量指标:

$$MMD = E_{x, x' \sim P_{data}} [k(x, x')] - 2E_{x \sim P_{data}, y \sim P_g} [k(x, y)] + E_{y, y' \sim P_g} [k(y, y')]$$

核函数: 将两样本映射为实数