

10 BiGAN-QP: 简单清晰的编码&生成模型

Dec By 苏剑林 | 2018-12-10 | 37163位读者

前不久笔者通过直接在对偶空间中分析的思路，提出了一个称为GAN-QP的对抗模型框架，它的特点是可以从理论上证明既不会梯度消失，又不需要L约束，使得生成模型的搭建和训练都得到简化。

GAN-QP是一个对抗框架，所以理论上原来所有的GAN任务都可以往上面试试。前面《不用L约束又不会梯度消失的GAN，了解一下？》一文中我们只尝试了标准的随机生成任务，而这篇文章中我们尝试既有生成器、又有编码器的情况：BiGAN-QP。

BiGAN与BiGAN-QP

注意这是BiGAN，不是前段时间很火的BigGAN，BiGAN是双向GAN（Bidirectional GAN），提出于《Adversarial feature learning》一文，同期还有一篇非常相似的文章叫做《Adversarially Learned Inference》，提出了叫做ALI的模型，跟BiGAN差不多。总的来说，它们都是往普通的GAN模型中加入了编码器，使得模型既能够具有普通GAN的随机生成功能，又具有编码器的功能，可以用来提取有效的特征。把GAN-QP这种对抗模式用到BiGAN中，就得到了BiGAN-QP。

话不多说，先来上效果图（左边是原图，右边是重构）：



这是将256x256x3的图片降维到256维度，然后再重构出来的。可以看到，整体的重构效果是不错的，没有普通自编码器的模糊感。有一些细节缺失，相比IntroVAE是差了一点，不过这是模型架构和调参的问题了，并不是我擅长的。不管怎样，这个效果图应该可以表明BiGAN-QP是可以跑通的，而且效果还行。

本文的内容已经更新到GAN-QP的原论文：<https://arxiv.org/abs/1811.07296>，读者可以在arxiv上下载到最新版本。

BiGAN-QP简明推导

其实相比GAN，BiGAN的推导非常简单，只需要将原来的单输入 x 换成双输入 (x, z) 就行了，同样，有了GAN-QP基础的话，所谓BiGAN-QP，也是非常简单的。具体来说，原来GAN-QP是这样的

$$\begin{aligned}
 T &= \arg \max_T \mathbb{E}_{(x_r, x_f) \sim p(x_r)q(x_f)} \left[T(x_r, x_f) - T(x_f, x_r) - \frac{(T(x_r, x_f) - T(x_f, x_r))^2}{2\lambda d(x_r, x_f)} \right] \\
 G &= \arg \min_G \mathbb{E}_{(x_r, x_f) \sim p(x_r)q(x_f)} [T(x_r, x_f) - T(x_f, x_r)]
 \end{aligned} \quad (1)$$

现在变成了

$$\begin{aligned}
 T &= \arg \max_T \mathbb{E}_{x \sim p(x), z \sim q(z)} \left[\Delta T - \frac{\Delta T^2}{2\lambda d(x, E(x); G(z), z)} \right] \\
 G, E &= \arg \min_{G, E} \mathbb{E}_{x \sim p(x), z \sim q(z)} [\Delta T] \\
 \Delta T &= T(x, E(x); G(z), z) - T(G(z), z; x, E(x))
 \end{aligned} \quad (2)$$

或者简化版直接取 $\Delta T = T(x, E(x)) - T(G(z), z)$ 。理论上就这样行了，这就是BiGAN-QP。

但实际上这样很难学习到一个好的双向映射，因为这相当于从无数可能的映射中自动搜索出一个双向映射，比较困难。所以我们还需要一些“引导项”，我们用两个mse误差作为引导项：

$$\begin{aligned}
 T &= \arg \max_T \mathbb{E}_{x \sim p(x), z \sim q(z)} \left[\Delta T - \frac{\Delta T^2}{2\lambda d(x, E(x); G(z), z)} \right] \\
 G, E &= \arg \min_{G, E} \mathbb{E}_{x \sim p(x), z \sim q(z)} \left[\Delta T + \beta_1 \|z - E(G(z))\|^2 + \beta_2 \|x - G(E(x))\|^2 \right] \\
 \Delta T &= T(x, E(x)) - T(G(z), z)
 \end{aligned} \quad (3)$$

其实生成器的三项loss都很直观， ΔT 是生成的图像更加真实， $\|z - E(G(z))\|^2$ 是希望能重构隐变量空间， $\|x - G(E(x))\|^2$ 是希望能重构显变量空间。后两项不能太大，尤其是最后一项，太大会导致图像的模糊。

其中这两个正则项可以看成是 $G(z)$ 与 z 的互信息、 x 与 $E(x)$ 的互信息的一个上界，因此从信息的角度看，这两个正则项是希望 x, z 之间的互信息越大越好。相关的讨论可以参考InfoGAN论文，这两个正则项代表着它也属于InfoGAN的特里。所以完整来说，这应该是一个Bi-Info-GAN-QP。

互信息项可以在一定程度上稳定GAN的训练过程，减少模型坍缩（mode collapse）的可能性，因为一旦模型坍缩，那么互信息就不会大了。换句话说，如果模型坍缩，那么重构就不大可能了，重构loss会很大。

实验表明，再做一些小的调整，效果会更好。这个小的调整源于：两个mse项耦合起来还是过于强大了（loss的具体值不一定大，但是梯度很大），导致模型还是有生成模糊图像的倾向，所以需要停止掉一半的梯度，变为

$$\begin{aligned}
 T &= \arg \max_T \mathbb{E}_{x \sim p(x), z \sim q(z)} \left[\Delta T - \frac{\Delta T^2}{2\lambda d(x, E(x); G(z), z)} \right] \\
 G, E &= \arg \min_{G, E} \mathbb{E}_{x \sim p(x), z \sim q(z)} \left[\Delta T + \beta_1 \|z - E(G_{ng}(z))\|^2 + \beta_2 \|x - G(E_{ng}(x))\|^2 \right] \\
 \Delta T &= T(x, E(x)) - T(G(z), z)
 \end{aligned} \quad (4)$$

G_{ng} 和 E_{ng} 指的是强行让这部分的梯度为0，一般的框架都有这个算子，直接调用即可。这就是本文最终的BiGAN-QP模型。

代码与效果图 #

代码也已经补充到Github了: <https://github.com/bojone/gan-qp/tree/master/bigan-qp>

再来一些效果图，随机生成的：



重构图（左边是原图，右边是重构）：



可以看到，不管是随机生成还是重构，效果都能让人满意，并没有出现模糊的情况，这表明我们确实成功训练了一个同时具有编码和生成能力的GAN模型。

而且很重要的一个特点是：因为是降维重构，模型并不是（也无法做到）学会了一个逐像素对应的——映射，而是一个整体看上去差不多的清晰的重构结果。比如我们看到第一行的第一张和最后一行的第二张，模型基本上把人重构出来了，但有趣的是眼镜，我们发现模型确实也重构了眼镜但是换了另外一个“款式”的眼镜。我们甚至可以认为，模型已经学到了“眼镜”这个概念，只不过是降维重构，隐变量的表达能力有限，所以尽管模型知道那是眼镜，但不能重构出一模一样的眼镜出来，就只好换一款常见的眼睛了。

这是普通的VAE所要求的“逐点——对应重构”所无法实现的，“逐点——对应重构”也是造成VAE模糊的主要原因了。如果要完全可逆重构，只有像Glow那样的可逆模型才有可能做到了。

另外，又有编码器又有生成器，我们就可以玩玩真实图片的隐变量插值了：



BiGAN-QP真实图片插值（左一、右一是真实图片，左二、右二是重构图片，其余是插值图）

还可以看看BiGAN-QP眼中的相似图片（算出所有真实图片的隐变量，然后用欧氏距离或者cos值算相似度，找出最相似的，下图为欧氏距离的结果）：



BiGAN-QP眼中的相似（第一行为输入，后两行为相似图片）

欢迎使用与分享

前面已经提到，GAN-QP是一个理论完备的对抗框架，理论上所有的GAN任务都可以尝试一下。所以，如果读着您手头上正好有GAN任务，不妨尝试一下，然后你就可以去掉L约束，去掉谱归一化甚至去掉很多正则项，还不用担心梯度消失了。GAN-QP就是笔者致力于去掉GAN各种超参数所得的结果。

如果你有新的基于GAN-QP的应用结果，欢迎在此分享。

转载到请包括本文地址: <https://kexue.fm/archives/6214>

更详细的转载事宜请参考: 《科学空间FAQ》

如果您需要引用本文, 请参考:

苏剑林. (Dec. 10, 2018). 《BiGAN-QP: 简单清晰的编码&生成模型》[Blog post]. Retrieved from <https://kexue.fm/archives/6214>