Semi-supervised learning with GAN – part 2

讨论文章：

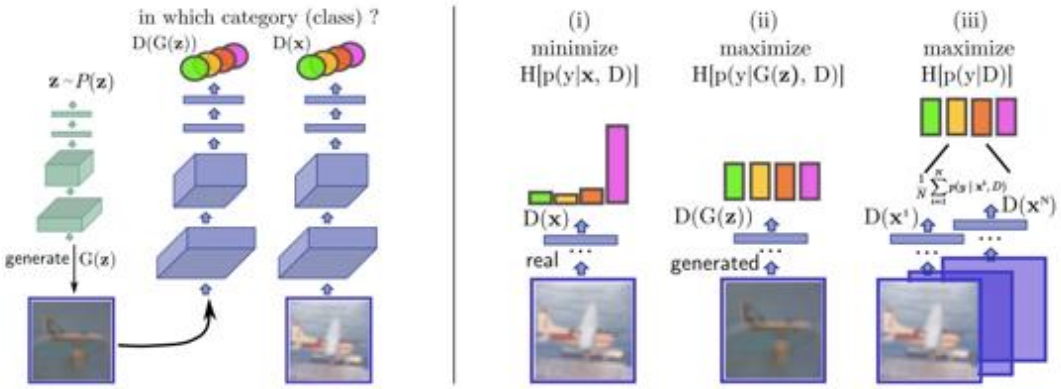
- Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks (arxiv id: 1511.06390, ICLR 2016); a.k.a **CatGAN**
- Triple Generative Adversarial Nets (arxiv id: 1703.02291); a.k.a **TripleGAN**

一、CatGAN

本场讨论由毛豆大佬（郑华滨）主导，分别往深向和纵向展开，讨论水平比较高，能够加入讨论的人不多。这里的总结按照毛豆大佬的思路展开。

深向讨论

- ✓ CatGAN 既可以做无监督，也可以做半监督。
- ✓ **判别器**：利用样本标签，希望它对真样本正确且明确地分类，正确类的概率要高；对假样本模糊分类，每一类的概率都要低。体现在概率分布上，真样本是尖峰分布（而且尖峰得在正确类上），假样本是平坦分布。



- ✓ **生成器**：希望能够在判别器那边获得一个尖峰，随便任何一个尖峰都可以。
- ✓ **Loss**：只要是用互信息。对于判别器，它的目标是最大化真实无标签样本的（样本与标签之间的）互信息，对于有标签样本，它则优化交叉熵，对于生成样本，它要最小化（样本与标签之间的）条件熵以达成平坦分布；对于生成器，它的目标是最大化生成样本的互信息（注意这里的标签是未指定的，因此随便一个尖峰都可以）。
- ✓ **一种可能的解释**：对 CatGAN 的可行性，猫豆大佬指出，可以从标签传播（label propogation）的角度进行解释。
  - ➢ 标签传播是半监督学习的经典范式之一，想法是从少量有标注样本出发，不断地往它们周边临近的样本“传染”自己的标签，慢慢地往外传播，直到所有的无标签样本都被传染到某一个标签。如果某个无标签样本距离有标签样本 a 很近，那它也很可能被“感染”上 a 的标签。
  - ➢ 单从判别器的角度来说，如果现在有一个跟 labelled sample A 非常临近的 unlablled sample B，那么 B 在判别器那边应该会获得一个相对不那么尖峰、但最高峰的类别仍然正确的概率分布。此时我们要求它尖峰更尖，那只会强化它原来的正确概率。在这个过程中，样本 A 的标签就稍微“传染”了临近的样本 B，使得样本 B 对于正确的类别更加“确信”。
  - ➢ 生成器在标签传播意义下的作用：g 建立了一道“隔离带”，阻止 label 跨界传染。

- 标签传播与文中提出的 half shot learning 有相似之处。
  - Half shot learning：比如我现在要做 10 类的分类，先按照非监督的方法聚成 20 簇，然后再人工将每个簇对应到一个类。
  - 标签传播是半监督的，jointly train; half shot learning 属于无监督的，assign 标签和聚类是分开的，它不对应到真实的类别上也能自成一体。

✓ CatGAN 跟 Good SSL requires bad GAN 那篇有很多联系哦！后者从理论上做了一些分析，很有趣，期待下一次 SSL 的讨论吧！提前准备一下，你也能参与进来。

纵向延伸

从判别器对真实有标签样本要求尖峰分布，对生成样本要求平坦分布的角度，可以将 CatGAN 与 CAN（Creative Adversarial Networks Generating "Art" by Learning About Styles and Deviating from Style Norms）做一个对比。（PS：要看懂下面的讨论，请先研究 CatGAN 和 CAN 的 loss 设计。）

✓ CAN
- CAN 想要生成艺术作品，但是又希望有创新性，这是 motivation。
- CAN 里面说的风格就是 label，就是 CatGAN 中的 category。
- 1. 让判别器做两个任务——区分真假和区分图像的艺术风格，看是巴洛克风格还是印象派之类的。
- 2. 让生成器的输出在判别器那边获得真概率，同时获得平坦的多类概率分布。
- CAN 这样要求的原因是：既要像真的，又要无法被归为哪一类艺术风格，既然无法归类，那就说明是创造了一种新的艺术风格。

✓ CAN 的生成器是想要获得平坦的多类概率分布，CatGAN 的生成器想要获得尖峰分布

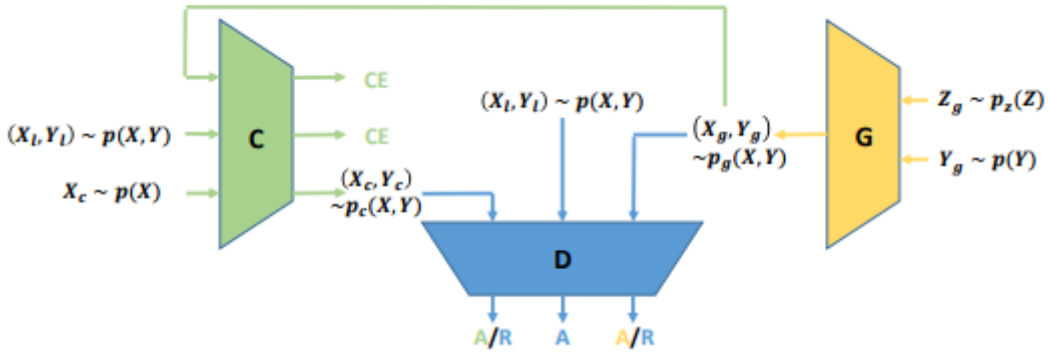
✓ 也正因此，CAN 产生的图像风格有点奇怪。对此，猫豆大佬是这样解释的：艺术家的事情，你等凡人不懂。吃瓜群众表示还是继续吃瓜好了。

二、TripleGAN

✓ TripleGAN 涉及三方，判别器、生成器和分类器。其中，判别器和生成器有对抗；判别器和分类器（在训练前期）有对抗；生成器和分类器有一点的协助作用。可以从斗地主

的角度来看，判别器是地主，生成器和分类器是农名。

- ✓ TripleGAN 还可以这么理解：拆掉分类器，它就是一个 CGAN。拆掉生成器，它就是一个半监督的 GAN。分类器和生成器有协助作用。判别器扮演统筹的角色。



- ✓ 判别器：判断样本与标签是否匹配。D 对于 C 来说，作用就是评估分类性能。
- ✓ 生成器：生成特定类别的图像，使得它尽可能地像真实样本。
- ✓ 分类器：对带标签的真样本要做监督学习（优化交叉熵），对没有标签的真样本要预测 label，然后送给 D 去判断分类好不好。对于生成样本（带标签），当做带标签的真样本做监督学习。实际上分类器只负责监督学习。对于没有标签的样本，它预测 label 以后就把锅甩给判别器了。
- ✓ 分类器把锅甩给判别器并不是一件简单的事，反向传播的时候，梯度无法直接从 label 回传给分类器，因此需要 reinforce algo。此外，作者还提出了另一种解决方案，增加 confidence loss 和 consistency loss。

Since properly leveraging unlabeled data is key to success in SSL, it is necessary to regularize C heuristically as in many existing methods [23, 26, 13, 15] to make more accurate predictions. We consider two alternative losses on the unlabeled data. Confidence loss [26] minimizes the conditional entropy of $p_c(y|x)$ and the cross entropy between $p(y)$ and $p_c(y)$, weighted by a hyperparameter $\alpha_{\mathcal{B}}$, as $\mathcal{R}_{\mathcal{U}} = H_{p_c}(y|x) + \alpha_{\mathcal{B}} E_p\big[-\log p_c(y)\big]$, which encourages C to make predictions confidently and be balanced on unlabeled data. Consistency loss [13] penalizes the network if it predicts the same unlabeled data inconsistently given different noise ϵ, e.g., dropout masks, as $\mathcal{R}_{\mathcal{U}} = E_{x\sim p(x)}||p_c(y|x,\epsilon) - p_c(y|x,\epsilon')||^2$, where $||\cdot||^2$ is the square of the l_2-norm. We use the confidence loss by default except on the CIFAR10 dataset (See details in Sec. 5).

Another consideration is to compute the gradients of $E_{x\sim p(x), y\sim p_c(y|x)}[\log(1-D(x,y))]$ with respect to the parameters θ_c in C, which involves summation over the discrete random variable y. Because directly integrating out y is time-consuming, we use a variant of the REINFORCE algorithm [29], in which the gradients should be $E_{x\sim p(x)} E_{y\sim p_c(y|x)}[\nabla_{\theta_c} \log p_c(y|x) \log(1 - D(x,y))]$. In our experiment, we find the best strategy is to use most probable y instead of sampling one to approximate the expectation over y. The bias is small as the prediction of C is rather confident typically.

- ✓ 生成器对 p(x|y)进行建模，而分类器则对 p(y|x)建模。两者恰好构成对偶！我们来看一下是否真的是对偶学习，对偶学习要求两个方向的建模要得到一致的 p(x,y)，也就是 $p(x)p(y|x) \approx p(y)p(x|y)$。而这个目标实际上是通过统筹者判别器来达成的，判别器恰好是对 p(x,y)进行建模，它接收的样本来源包括：生成器、分类器和带标签的真实样本。一个成功的判别器将使得生成器方向建模的 p(x,y)和分类器方向建模的 p(x,y)达到一致。这是很漂亮的对偶思想！（PS：原来的讨论说不是完整的对偶思想是不对的，在此纠正！）