

译文 | 带条件的 GAN: 在 GAN 中加个 y

作者: Mehdi Mirza 等

翻译: 七月在线 DL 翻译组

译者: 张兴园 路转 管枫

责编: 翟惠良 July

声明: 本译文仅供学习交流, 有任何翻译不当之处, 敬请留言指正。转载请注明出处。

原文: <https://arxiv.org/pdf/1411.1784.pdf>

——前言: 本文展示了如何构建条件生成式对抗网络(GAN), 以及条件生成式对抗网络在 MNIST 手写数字生成以及 MIR Flickr 中 25000 图像的标注生成上的实验结果。

摘要

生成式对抗网络 (GAN) 是一种用来训练生成式模型的新方法。本文中, 我们在 GAN 的基础之上引入条件生成式对抗网络, 它的构建并不复杂, 只需要在生成模型与判别模型的构建中分别输入条件数据 y 。实验结果显示此模型能够在类别标签条件下生成 MNIST 手写体数字。对于多模态模型的训练, 本文也给出了一个生成图像标记的初步示例, 在这个示例中我们演示了如何生成训练标签之外的描述性标签。

简介

生成对抗网络 (GAN) 是最新引入的一种用来训练生成模型的方法, 它可以有效的绕过一些其他方法所很难避免的困难的概率计算。

生成网络中不再需要马尔科夫链, 而只需要用回溯法就可以计算优化步骤中所需要的梯度向量。在模型训练中也不再需要统计推断, 这种方法使得其所训练的模型可以很容易就囊括各种不同类型的信息。

然而, 在非条件生成模型中, 无法控制其生成数据的模式。使用附加信息作为模型的条件变量, 可以引导生成模型的数据生成。这些条件变量可以来自于类别标签, 也可以来自于待修复数据的其他部分, 或者甚至来自于其他模态的数据。

相关工作——多模态学习实现图像内容标记

近年来监督神经网络取得很大成功——尤其是卷积神经网络, 但是对于使用上述模型来预测极大输出类别时仍然是一个挑战。其次, 迄今为止的大部分工作集中于学习从输入到输出的一对一映射。然而, 许多有趣的问题在概率上是一对多的映射问题。例如, 图像标记问题, 对于给定的一副图像, 可能有不同的标签, 同时不同的标记模型可能使用不同的 (但是同义或相关的) 词语来描述同一副图像。

解决第一个问题的一种方式是利用其它模态的额外信息, 例如, 可以使用自然语言语料库来训练标签的向量表示, 并保证向量之间的距离远近可以表示语义上含义的远近。在这样的空间进行预测的一个好处是, 即使预测有偏差预测结果也能跟真实情况比较“接近”(比如“table”和“chair”), 同时这种方法是我们可以预测那些在训练集中没有出现的标签。在引用 3 中提到, 即使一个从图像特征空间到词表示空间的简单的线性映射也能改善分类效果。

解决第二个问题的方式就是使用条件概率生成模型, 在这种模型中输入的是条件变量, 这样一来预测一对多的映射就变成了预测条件分布。

在原文的引用中，有的工作采用了类似的方法来解决上述问题，在 MIR Flickr 25000 数据集上训练了一个多模态深度玻尔兹曼机。此外，原文的引用 12 提到了如何训练一个有监督的多模态神经语言模型，同时能够生成图像的描述性句子。

CGAN——条件生成式对抗网络

生成式对抗网络：

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

——生成式对抗网络由两个“对抗式”模型组成：生成式模型 G 来获取数据分布，判别式模型 D 用来估计一个样本来自训练集而不是来自于 G 的概率。 G 和 D 都是非线性映射函数，例如多层感知器。为了学习数据 \mathbf{x} 的生成式分布 p_g ，生成器构建一个从先验噪声分布 $p_z(\mathbf{z})$ 数据空间的映射函数 $G(\mathbf{z}; \theta_g)$ 。判别器 $D(\mathbf{x}; \theta_d)$ 则输出单个标量来表示 \mathbf{x} 来自训练数据而不是 p_g 的概率。 G 和 D 同时进行训练：我们针对 G 调整参数来最小化 $\log(1 - D(G(\mathbf{z})))$ 同时针对 D 调整参数来最小化 $\log D(\mathbf{x})$ ，如同两个玩家使用如下的价值函数 $V(G; D)$ 玩最小最大游戏。

在引入条件变量 \mathbf{y} 后，生成式对抗网络变成：

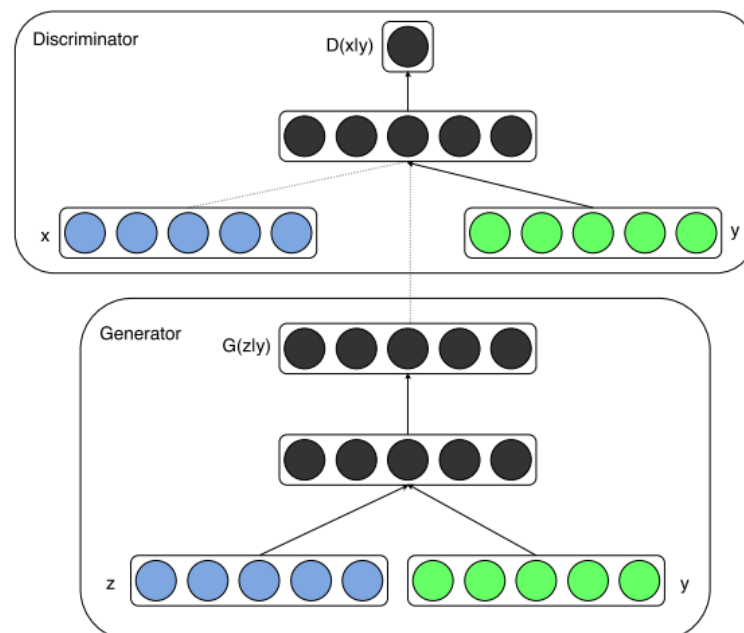
$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x}|\mathbf{y})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z}|\mathbf{y})))].$$

如上面目标函数所示：如果生成器和判别器都共同的额外条件变量 \mathbf{y} ，生成式对抗网络就可以够扩展成条件模型 \mathbf{y} 可以是任何类型的辅助信息，比如类别标签或者其他模态的数据。我们通过将 \mathbf{y} 作为额外输入层导入到判别器和生成器来实现条件模型。

在生成器中，先验输入的噪声 $p_z(\mathbf{z})$ 和 \mathbf{y} 在隐藏层中相结合，这使得对抗网络训练框架在如何构成隐藏层方面具有了很大的灵活性。

在判别器中 \mathbf{x} 和 \mathbf{y} 共同做为判别函数的输入（这里仍然考虑在 MLP 中进行实现的情形）。

典型对抗网络的框架如下：



实验及结果分析

1. 单模态实验

实验方法：使用标签类别作为条件变量 \mathbf{y} ，对编码成 one-hot 向量的 MNIST 图像进行条件对

抗网络的训练。

在生成式网络部分，我们从单位超立方体上的均匀分布中提取 100 维的先验噪音 z 。先通过激活函数 ReLu (Rectified Linear Unit) 把先验噪音 z 和条件变量 y 映射到尺寸分别为 200 和 1000 的隐藏层，然后把这两个再映射到 1200 维的组合 ReLu 隐藏层，最终在输出层使用 sigmoid 单元层来生成 784 维 MNIST 样本。

判别器把 x 映射到具有 240 个单元 5 块的 maxout 层，把 y 映射到具有 50 个单元 5 块的 maxout 层。然后把这两个隐藏层都映射到一个具有 240 个单元 4 块的联合隐藏 maxout 层，最后再把这个层的输出导入 sigmoid 层。(判别器的具体构造并不关键，只要它具有相当的判别能力就可以，我们实践发现 maxout 单元比较适合这个任务) 模型的训练使用随机梯度下降法，其中每个 mini-batch 的大小为 100，学习率初始值为 0.1，然后以 1.00004 的下降率指数下降到 0.000001，同时初始动量为 0.5，最终增长至 0.7。为了防止过度拟合，我们在生成器和判别器中的都使用了概率为 0.5 的 Dropout 层。

Model	MNIST
DBN [1]	138 ± 2
Stacked CAE [1]	121 ± 1.6
Deep GSN [2]	214 ± 1.1
Adversarial nets	225 ± 2
Conditional adversarial nets	132 ± 1.8

Table1: 基于 Parzen window 对 MNIST 进行的对数似然估计，

我们使用了和引文 8 中相同的方法来进行计算

Table1 给出了对于 MNIST 数据集的测试数据做的高斯 Parzen window 对数-似然估计结果。我们从十个类别中每个类别抽取出来 1000 个样本进行了高斯 Parzen window 拟合，然后使用 Parzen window 分布对这个测试集进行对数似然估计 (引文 8 中有关于如何构造这个估计的更详细讨论)。

我们这里使用条件对抗网络得到的结果与一些基于 network 的构架得出的结果相当，但是却不加一些其他的方法 (比如非条件对抗网络)。这里得出的结果更多的是作为一个概念验证，但是相信对超参数以及架构的进一步探索可以使得条件生成网络最终达到或者超过非条件生成网络的水平。

下图 2 给出了一些生成的 MNIST 手写数据样本，每一行代表一个 (分类) 条件标签，而每一列代表一组生成样本。



图 2. 生成的 MNIST 手写数据样本，每一行代表一个标签

2. 多模态实验

试验方法：以图像特征为条件变量，使用条件生成网络来生成标签向量的条件分布。并以此实现图像的多标签自动标注。

类似于 Flickr 这样的照片网站提供了大量的带内容标注的图像数据以及与之相关的用户生成的元数据（UGM），特别是用户标签。

用户生成的元数据不同于“规范的”图像标记，因为他们更具有描述性，同时在语义上与人类使用自然语言而不是识别图像中存在的目标实现对图像进行描述更加接近。UGM 的另一个方面是同义词非常普遍，同时不同的用户可能使用不同的词汇来描述同一个概念。因此，使用有效的方法来标准化这些标签也就变得非常重要。引文 14 提的概念词嵌入方法就很有效，因为这种方法使得相关的概念最终由相似的向量表示。

对于图像特征，我们预先在带有 21000 个标签的完整 ImageNet 数据集上训练一个类似于引文 13 中提到的卷积模型。然后使用该模型特有的带有 4096 个单元的最后一个全连接层对图像特征进行表示。





对于单词表示，首先从 YFCC100M 元数据集中收集了一个混合了用户标签、标题以及描述文本的语料库。经过预处理和文档清理，本文使用大小为 200 的单词向量进行 skip-gram 模型拟合。我们过滤掉出现次数少于 200 次的单词，从而得到一个最终大小为 247465 的单词表。

在训练对抗网络时保持这个卷积模型和语言模型，并把基于这些模型的反向传播算法实验留作今后的工作。

本文的实验对 MIR Flickr 25000 数据集使用如上提到的卷积模型与语言模型提取了图像和标注特征。在实验中过滤掉了没有标注数据的图像，而将附注（annotations）作为额外的标注。实验选取前 150000 个例子作为训练集。有多个标注的图像在训练集中重复出现（每一个标注重复一次）。

作为测试，我们对每个图像生成 100 个样本，并且在每一个样本中使用余弦相似函数选取前 20 个最接近的词语。然后在 100 个样本中选取前 10 个出现最多的词。

下表给出了一些用户标注和附注与模型生成的标注的对比。

	User tags + annotations	Generated tags
	montanha, trem, inverno, frio, people, male, plant life, tree, structures, transport, car	taxi, passenger, line, transportation, railway station, passengers, railways, signals, rail, rails
	food, raspberry, delicious, homemade	chicken, fattening, cooked, peanut, cream, cookie, house made, bread, biscuit, bakes
	water, river	creek, lake, along, near, river, rocky, treeline, valley, woods, waters
	people, portrait, female, baby, indoor	love, people, posing, girl, young, strangers, pretty, women, happy, life

表现最好的模型的生成器使用的是如下配置：以 100 个单位的高斯噪音为先验噪音，然后映射到 500 维的 ReLu 层。把 4096 维的图像特征向量映射到 2000 维的 ReLu 隐藏层。这两个 ReLu 层都映射到一个 200 维的线性混合表示层，而由这一层来输出词语向量。

判别器包含 500 维的词语向量隐藏层和 1200 维的图像特征 ReLu 隐藏层，这两层映射到一个 1000 个单元与 3 块的 maxout 层，然后最终导入到一个单 sigmoid 单元，并输出判断。模型的优化使用的是随机梯度下降法，配置为:100 单元的 mini-batches，初始学习率为 0.1 并以 1.00004 的下降率指数下降到 0.000001，初始动量为 0.5 并最终上升到 0.7。生成器与判别器都使用概率为 0.5 的 Dropout 层。这里，超参数与构架选取中用到了交叉验证、网格查找和人工手动选择。

未来工作

本文展示的都是目前初步的结果，但是他们一方面显示了条件生成网络的潜力，另一方面这些结果本身也是很有趣并具有实用价值的应用。

我们将会在探索更多更复杂的模型，并且对这些模型的表现与特点进行更深入的分析。而且在我们目前的实验中，每一个标签都作为独立的条件变量。在更进一步的研究中，如果能把多个标签联合起来使用，或许会得到更好的结果。

另外将来可以探索的一个方向是为语言模型创建一个混合训练机制。如同引文 12 提到的很多工作都证实可以针对特别的任务训练有针对性的语言模型。

完。

后记

关于我们

七月在线 DL 翻译组是由一群热爱翻译、热爱 DL、英语六级以上的研究生或博士组成，有七月在线的学员，也有非学员。本翻译组翻译的所有全部论文仅供学习交流，宗旨是：汇集顶级内容 帮助全球更多人。目前已经翻译数十篇顶级 DL 论文，详见：

<https://ask.julyedu.com/question/7612>

加入我们

如果你过了英语六级、是研究生或博士、且熟练 DL、热爱翻译，欢迎加入我们翻译组，
微博私信@研究者 July

GAN 课程

为了帮助更多人更好的了解、学习、入门 GAN，今年上半年，我们七月在线亦会开《生成对抗网络班》，从头到尾详解 GAN 的原理及其实战应用，敬请期待。