NVAE: A Deep Hierarchical Variational Autoencoder

Desty Wang

Qiuzhen College of Tsinghua University

2024 年 9 月 12 日

Main idea

1. 目前对 VAE 的改进主要集中于统计方法（减少分布差异、更好的下界、减少梯度噪声等）
2. 本文主要致力于优化训练和生成的整体框架。

Problems of VAE

1. 参数过多
2. 感受野太小，难以处理长程相关性
3. KL 散度无界
4. 层数过多导致不稳定

Model advantages

1. 在所有非自回归似然方法中做到最好，并减小于自回归方法的差距
2. 模型主要添加了深度卷积网络，增加感受野的同时不会大幅增加参数数量

Main contributions

1. 添加深度卷积网络（on generateive model）
2. 近似后验中的 Residual parameterization
3. Spectral regularization
4. 减轻计算负担

Model

1. 借鉴 inverse autoregressive flows（IAF-VAEs）
2. 不同点：网络架构、近似后验如何参数化、扩展到大图片（256x256）
3. 这里的主要问题是如何使用神经网络实现 p（x，z）和 q（z|x）中的条件。为了对生成模型进行建模，自上而下的网络会生成每个条件的参数。从每组采样后，样本与确定性特征图相结合，并传递给下一组（图 2b）。为了推断 q（z|x）中的潜在变量，我们需要一个自下而上的确定性网络来从输入 x 中提取表示。由于潜在变量组的顺序在 q（z|x）和 p（z）中一致，因此我们还需要一个额外的自上而下的网络来逐组推断潜在变量。为了避免额外的自上而下的模型的计算成本，在双向推理中，在生成模型中自上而下的模型中提取的表示被重新用于推断潜在变量（图 2a）。IAF-VAEs 在自上而下和自下而上的模型中都依赖于规则残差网络，没有任何 BN，并且仅在小图像上进行了测试。

Residual cells

1. 使用 hierarchy multi-scale model 构建 VAE：从空间上排列的小型潜在变量 z_1 开始，逐组从层次结构中采样，同时逐渐将空间维度加倍。这种多尺度方法使 NVAE 能够在层次结构的顶部捕获全局长期相关性，并在较低组捕获局部细粒度依赖关系。
2. 增加卷积核的 size 以提高感受野
3. 由于深度卷积网络在使参数数量降低的同时性能优于常规卷积，但表达能力有限。文章提出使用 1x1 卷积核增加通道数并在结束时反卷积还原数据。

Residual cells

1. Batch Normalization：目前一些比较好的 VAE 都采用（WN），省略了 BN，因为批量归一化引入的噪声会损害性能。但由于 BN 的负面影响是在评估期间，而不是训练期间。（在 BN 中使用 running statistics，因此在评估过程中由于积累误差，每个 BN 层的输出可能会略有偏移，从而导致网络输出发生巨大变化。为此，本文调整了 BN 的动量超参数，并对 BN 层中的缩放参数范数进行了正则化，以确保统计中的小错配不会被 BN 放大。
2. Swish activate：$f(x) = x\sigma(\beta x)$，
3. Squeeze and Excitation (SE): 子结构，用于学习 weight，使有效的 feature map 的权重增大，无效的减小
4. For encoder，常规卷积（深度卷积不起作用）。BN-Activation-Conv 的性能优于原始的 Conv-BN-Activation。
5. 使用 Gradient Checkpoint 减少储存：本质是用时间换空间，不保存整个计算图中所有的中间结果进行反传，而是在反传的过程中重新计算中间结果。

Taming unboundness of KL

通常使用两个单独的神经网络来生成分布 p 和 q 的参数。但在存在大量潜在变量组的情况下，很难使两个分布保持和谐，从而 KL 无界。且如果编码器和解码器在训练过程中产生彼此相距甚远的分布，则由 KL 产生的梯度急剧更新会使模型参数不稳定。本文提出了两种改进 KL 优化和稳定训练的方法。

1. Residual Normal Distributions：采用 running（modification）的方法，当前训练的潜变量分布都是之前的潜变量的分布的一个相对修正。
2. Spectral reglarization：上面的 KL 仍然是无界的。本文通过正则化 Lipschitz 常数（encoder 的输入输出），确保编码器预测的潜在代码保持有界，从而实现稳定的 KL 最小化。但 Lipschitz 常数很难计算，因此使用谱正则化来最小化每一层的 Lipschitz 常数。
3. 在 NVAE 中，p 和 q 由组之间的自回归分布和每组独立分布建模, 从而有效地从每个组中并行采样。但其表达能力较差。解决方法：将额外的归一化流应用于 q（z|x）中每个组生成的样本。由于它们仅应用于编码器网络，因此 i）可以依赖逆自回归流（IAF）因为不需要流的显式反转，并且 ii）采样时间不会因为流而增加。

谢谢!