CATEGORICAL REPARAMETERIZATION WITH GUMBEL-SOFTMAX

Eric Jang
Google Brain
ejang@google.com

Shixiang Gu*
University of Cambridge
MPI Tübingen
sg717@cam.ac.uk

Ben Poole*
Stanford University
poole@cs.stanford.edu

【数学】 Gumbel Softmax

张楚珩
清华大学 交叉信息院博士在读

原文传送门

ICLR 2017：Jang, Eric, Shixiang Gu, and Ben Poole. "Categorical reparameterization with gumbel-softmax." arXiv preprint arXiv:1611.01144 (2016).

特色

对于离散变量，最常用的分布就是 categorical 分布，这种分布下需要用 reparameterization trick 来求导的话，就需要用到 Gumbel Softmax 这种方法。在离散版本的 soft actor critic 的实现中需要使用到这种功能技术。

过程

1、Reparameterization Trick

深度学习里面经常会使用神经网络 A 生成一个概率分布，这个分布一般是事先规定好的，神经网络只需要生成这个分布的 statistical parameter D 即可，接下来会从这个概率分布里面采样得到一个样本 S，然后再把这个样本输入到后续的神经网络 B 里面处理，并且计算得到一个可导的损失函数 L。但是由于这个采样步骤的存在，没有办法做到 end-to-end 的训练。我们可以得到 L 关于 S 的导数，也可以得到 D 关于 A 参数的导数；但是一般来说无法得到 S 关于 D 的导数。要得到 S 关于 D 的导数就需要我们使用 reparameterization trick。

$$\frac{\partial}{\partial\theta}\mathbb{E}_{z\sim p_\theta}\left[f(z)\right] = \frac{\partial}{\partial\theta}\mathbb{E}_{\epsilon}\left[f(g(\theta,\epsilon))\right] = \mathbb{E}_{\epsilon\sim p_\epsilon}\left[\frac{\partial f}{\partial g}\frac{\partial g}{\partial\theta}\right] \qquad (4)$$

这里的 f 相当于神经网络 B，z 是采集得到的样本 S，theta 相当于神经网络 A 的参数。这里第一个等式相当于把样本 z 写作了一个固定分布采样得到的样本 epsilon 和与参数有关的函数。对于高斯分布来说 $\epsilon\sim\mathcal{N}(0,1)$ ， $g(\theta,\epsilon)=\mu_\theta+\sigma_\theta\cdot\epsilon$ 。

2、使用 Gumbel-Max Trick 实现样本对于 Statistical Parameter 的求导

可以使用 Gumbel-Max trick 从一个 categorical distribution 中采样，给定每个类的采样概率 $\pi_1,\cdots,\pi_k$ ，采集的样本可以表示为

$$z = \mathrm{one\_hot}\left(\arg\max_i [g_i + \log \pi_i]\right) \quad (1)$$

其中 g_i 是从 Gumbel(0,1) 分布中采集的样本。注意到一个样本表示为一个 k 维的 one-hot 向量。Gumbel(0,1) 的 PDF 函数定义如下

$$p_{\pi,\tau}(y_1, ..., y_k) = \Gamma(k)\tau^{k-1}\left(\sum_{i=1}^{k} \pi_i / y_i^{\tau}\right)^{-k} \prod_{i=1}^{k} \left(\pi_i / y_i^{\tau+1}\right) \quad (3)$$

Gumbel(0,1) 的生成方法如下

[1]The Gumbel$(0,1)$ distribution can be sampled using inverse transform sampling by drawing $u \sim$ Uniform$(0,1)$ and computing $g = -\log(-\log(\mathrm{u}))$.

为了使得样本能够对于 statistical parameter 的求导，还需要解决掉其中不可求到的 arg max 部分。这里使用 softmax 来代替 one_hot(arg max (·))，这样采集的样本可以写作

$$y_i = \frac{\exp((\log(\pi_i) + g_i)/\tau)}{\sum_{j=1}^{k} \exp((\log(\pi_j) + g_j)/\tau)} \qquad \text{for } i = 1, ..., k. \quad (2)$$

其中 τ 是温度参数。如下图所示，它越小温度越低，采样的期望更接近 arg max 的结果，并且采样得到的样本也更接近 one-hot 向量，但是其对应的 gradient estimator 的方差也越大哦；它越大代表更高的温度，采样的期望则更平均，而且采样得到的期望也更不 one-hot，但是其对应的 gradient estimator 的方差会比较小。

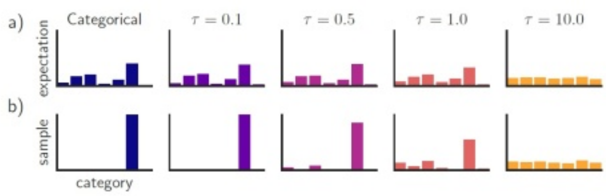

Figure 1: The Gumbel-Softmax distribution interpolates between discrete one-hot-encoded categorical distributions and continuous categorical densities. (a) For low temperatures ($\tau = 0.1, \tau = 0.5$), the expected value of a Gumbel-Softmax random variable approaches the expected value of a categorical random variable with the same logits. As the temperature increases ($\tau = 1.0, \tau = 10.0$), the expected value converges to a uniform distribution over the categories. (b) Samples from Gumbel-Softmax distributions are identical to samples from a categorical distribution as $\tau \to 0$. At higher temperatures, Gumbel-Softmax samples are no longer one-hot, and become uniform as $\tau \to \infty$.

在实际应用中为了平衡这个 trade-off，一般可以在训练过程中逐步减小温度；当然也可以把它作为一个参数来学习。

3、相关工作

考虑这样一个问题，有一个损失函数

$$L(\theta)=\mathbb{E}_{z\sim p_{\theta}(z)}[f(z)]$$

希望求到它对于参数 theta 的导数

$$\nabla_{\theta}\mathbb{E}_{z\sim p_{\theta}(z)}[f(z)]$$

该问题有如下的做法。第一张图表示，如果没有采样的过程，那么可以直接求导。在第二张图中有了一个采样的过程，因此求导无法从样本到分布上。第三张图表示 score function estimator，其方法类似于 REINFOCE。该方法可以表示为

$$\nabla_{\theta}\mathbb{E}_{z}\left[f(z)\right]=\mathbb{E}_{z}\left[f(z)\nabla_{\theta}\log p_{\theta}(z)\right] \tag{5}$$

该方法甚至可以不要求 f 可以求导，直接把 f 的数值作为权重来产生一个无偏估计。该方法的缺点是 variance 大，图片下方括号中给出了在此基础上的一系列降低 variance 的方法。第四张图表示 Straight-Through（ST）方法，即直接认为 z 关于 statistical parameter 的导数为 1，这种功能方法只能适用于 Bernoulli 分布（或者 categorical 分布），因为不难看出，这里至少要求分布的 statistical parameter 的维度和样本 z 的维度相同。最后一张图表示本文中使用的一类方法叫做 path derivative estimator 或者 reparametrization trick。其主要思想是把随机采样这个步骤放到梯度回传的路径之外。

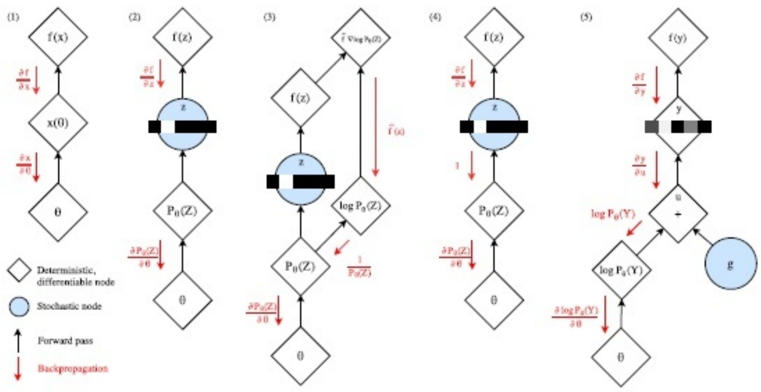

Figure 2: Gradient estimation in stochastic computation graphs. (1) $\nabla_{\theta}f(x)$ can be computed via backpropagation if $x(\theta)$ is deterministic and differentiable. (2) The presence of stochastic node z precludes backpropagation as the sampler function does not have a well-defined gradient. (3) The score function estimator and its variants (NVIL, DARN, MuProp, VIMCO) obtain an unbiased estimate of $\nabla_{\theta}f(x)$ by backpropagating along a surrogate loss $\hat{f}\log p_{\theta}(z)$, where $\hat{f}=f(x)-b$ and b is a baseline for variance reduction. (4) The Straight-Through estimator, developed primarily for Bernoulli variables, approximates $\nabla_{\theta}z\approx 1$. (5) Gumbel-Softmax is a path derivative estimator for a continuous distribution y that approximates z. Reparameterization allows gradients from $f(y)$ to θ. y can be annealed to one-hot categorical variables over the course of training.