Bit-Swap: Recursive Bits-Back Coding for Lossless Compression with Hierarchical Latent Variables

Панаетов Александр

Higher School of Economic

20 февраля 2020 г.

Обзор

Lossless compression

- Есть данные $x = (x_1, ..., x_D)$ из дискретного распределения p_{data}. Хотим отправить их получателю, используя при этом минимальное число бит (в среднем по распределению p_{data}), гарантируя при этом возможность получателю полностью восстановить данные.
- Entropy coding схемы используя вероятностную модель $p_\theta(x)$ требуют $-\log p_\theta(x)$ бит. Если p_θ хорошо приближает p_{data}, то это оптимальный алгоритм.

- ANS позволяет используя $p(x)$ кодировать x используя примерно $-\log p(x)$ бит (константный оверхед около 2 бит)
- Важное свойство - если мы кодируем x в одном порядке, то декодируем в обратном

Latent Variable Model

- $p_\theta(x)$ - приближает $p_{data}(x)$, $q_\theta(z|x)$ приближает $p_\theta(z|x)$. Тогда:

$$p_\theta(x) = \int p_\theta(x,z)dz = \int p_\theta(x|z)p(z)dz$$

$$\log p_\theta(x) = \mathbb{E}_{q_\theta(z|x)} \log \frac{p_\theta(x,z)}{q_\theta(z|x)} + \mathbb{E}_{q_\theta(z|x)} \log \frac{q_\theta(x,z)}{p_\theta(z|x)}$$

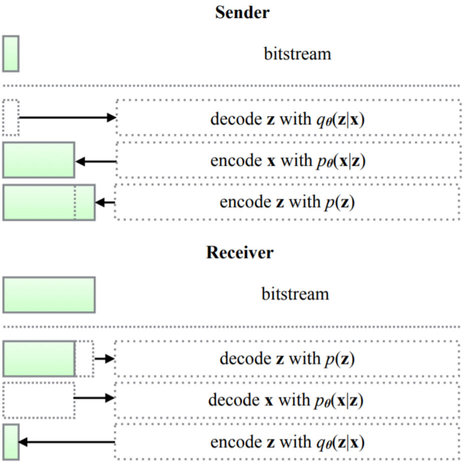
$$L(\theta) = \mathbb{E}_{q_\theta(z|x)}[\log p_\theta(x,z) - \log q_\theta(z|x)]$$

Bits-Back Coding with ANS

- Пусть у нас теперь всем известны $p_\theta(x|z)$, $p(z)$, $q_\theta(z|x)$, хотим отправить x
- Можно сделать так:
  - сэмплируем z из $q_\theta(z|x)$
  - отправляем на стек x используя $p_\theta(x|z)$
  - отправляем на стек z используя $p(z)$
- Лучше также, но z получаем так: используя $q_\theta(z|x)$ забираем $-\log q_\theta(z|x)$ бит из нашего стека и получаем z. Там мы по сути отправляем на это количество бит меньше, чем в предыдущем варианте.
- Итого всего тратим бит:

$$N_{total} - N_{init} = \log q_\theta(z|x) - \log p_\theta(x|z) - \log p(z)$$

Bits-Back Coding with ANS



Hierarchical Latent Variables

- Теперь рассмотрим модель с несколькими латентными переменными, которые образуют марковскую цепь

$$z_L \to z_{L-1} \to ... \to z_1 \to x$$

$$p_\theta(x) = \int p_\theta(x|z_1)p(z_1)dz_1$$

$$p_\theta(z_1) = \int p_\theta(z_1|z_2)p(z_2)dz_2$$

$$...$$

$$p_\theta(z_{L-1}) = \int p_\theta(z_{L-1}|z_L)p(z_L)dz_L$$

$$p_\theta(x) = \int p_\theta(x|z_1)_\theta(z_1|z_2)...p(z_L)dz_{1:L}$$

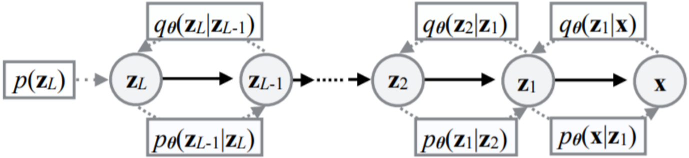
$$L(\theta) = \mathbb{E}_{q_\theta(\cdot|x)}[\log p_\theta(x, z_{1:l}) - \log q_\theta(z_{1:L}|x)]$$

Hierarchical Latent Variables

- Посмотрим теперь на количество бит требуемое для начальной инициализации:

$$N_{init} = \sum -\log q_\theta(z_{i+1}|z_i)$$

- N_{init} достаточно быстро растет с ростом нашей модели

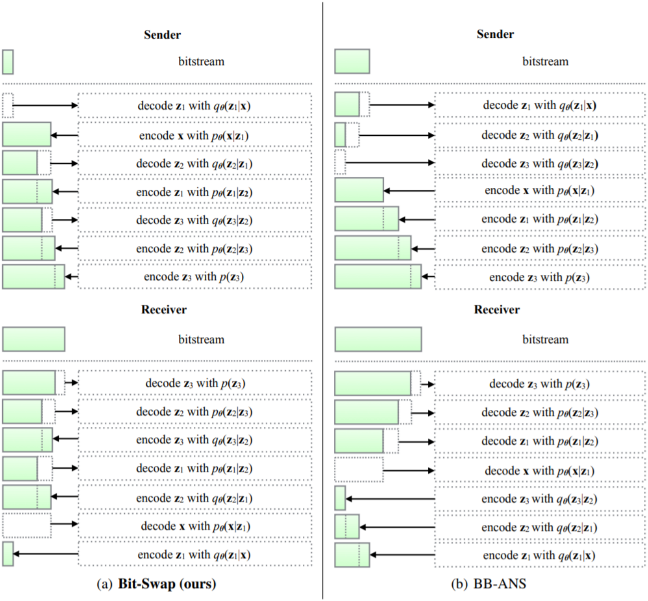


Bit-Swap

- Идея теперь - обобщить прошлый подход с использованием $q_\theta(z|x)$ для того, чтобы забрать из стека информацию, которую потом сможем восстановить.
- Будем не последовательно декодировать все z_i, а сначала только z_1. А потом уже чередуем декодирование новой z_{i+1} с кодированием старой z_i
- Так мы точно потратим меньше бит на N_{init}:

$$N_{init}^{BitSwap} \leq \sum \max(0, \log \frac{p_\theta(z_{i-1}|z_i)}{q_\theta(z_{i+1}|z_i)}) \leq \sum -\log q_\theta(z_{i+1}|z_i) = N_{init}^{BBANS}$$

- Поскольку ничего не понятно, сейчас будут картинки и все станет понятно


Sender
bitstream
decode $\mathbf{z}_1$ with $q_\theta(\mathbf{z}_1|\mathbf{x})$
encode $\mathbf{x}$ with $p_\theta(\mathbf{x}|\mathbf{z}_1)$
decode $\mathbf{z}_2$ with $q_\theta(\mathbf{z}_2|\mathbf{z}_1)$
encode $\mathbf{z}_1$ with $p_\theta(\mathbf{z}_1|\mathbf{z}_2)$
decode $\mathbf{z}_3$ with $q_\theta(\mathbf{z}_3|\mathbf{z}_2)$
encode $\mathbf{z}_2$ with $p_\theta(\mathbf{z}_2|\mathbf{z}_3)$
encode $\mathbf{z}_3$ with $p(\mathbf{z}_3)$
Receiver
bitstream
decode $\mathbf{z}_3$ with $p(\mathbf{z}_3)$
decode $\mathbf{z}_2$ with $p_\theta(\mathbf{z}_2|\mathbf{z}_3)$
encode $\mathbf{z}_3$ with $q_\theta(\mathbf{z}_3|\mathbf{z}_2)$
decode $\mathbf{z}_1$ with $p_\theta(\mathbf{z}_1|\mathbf{z}_2)$
encode $\mathbf{z}_2$ with $q_\theta(\mathbf{z}_2|\mathbf{z}_1)$
decode $\mathbf{x}$ with $p_\theta(\mathbf{x}|\mathbf{z}_1)$
encode $\mathbf{z}_1$ with $q_\theta(\mathbf{z}_1|\mathbf{x})$
(a) **Bit-Swap (ours)**
Sender
bitstream
decode $\mathbf{z}_1$ with $q_\theta(\mathbf{z}_1|\mathbf{x})$
decode $\mathbf{z}_2$ with $q_\theta(\mathbf{z}_2|\mathbf{z}_1)$
decode $\mathbf{z}_3$ with $q_\theta(\mathbf{z}_3|\mathbf{z}_2)$
encode $\mathbf{x}$ with $p_\theta(\mathbf{x}|\mathbf{z}_1)$
encode $\mathbf{z}_1$ with $p_\theta(\mathbf{z}_1|\mathbf{z}_2)$
encode $\mathbf{z}_2$ with $p_\theta(\mathbf{z}_2|\mathbf{z}_3)$
encode $\mathbf{z}_3$ with $p(\mathbf{z}_3)$
Receiver
bitstream
decode $\mathbf{z}_3$ with $p(\mathbf{z}_3)$
decode $\mathbf{z}_2$ with $p_\theta(\mathbf{z}_2|\mathbf{z}_3)$
decode $\mathbf{z}_1$ with $p_\theta(\mathbf{z}_1|\mathbf{z}_2)$
decode $\mathbf{x}$ with $p_\theta(\mathbf{x}|\mathbf{z}_1)$
encode $\mathbf{z}_3$ with $q_\theta(\mathbf{z}_3|\mathbf{z}_2)$
encode $\mathbf{z}_2$ with $q_\theta(\mathbf{z}_2|\mathbf{z}_1)$
encode $\mathbf{z}_1$ with $q_\theta(\mathbf{z}_1|\mathbf{x})$
(b) BB-ANS

Bit-Swap

Bit-Swap

Algorithm 1 BB-ANS for lossless compression with hierarchical latent variables. The operations below show the procedure for encoding a dataset $\mathcal{D}$ onto a bitstream.

Input: data $\mathcal{D}$, depth L, $p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}_{1:L})$, $q_{\boldsymbol{\theta}}(\mathbf{z}_{1:L}|\mathbf{x})$
Require: ANS
Initialize: bitstream
repeat
  Take $\mathbf{x} \in \mathcal{D}$
  decode $\mathbf{z}_1$ with $q_{\boldsymbol{\theta}}(\mathbf{z}_1|\mathbf{x})$
  for $i = 1$ **to** $L - 1$ **do**
    decode $\mathbf{z}_{i+1}$ with $q_{\boldsymbol{\theta}}(\mathbf{z}_{i+1}|\mathbf{z}_i)$
  end for
  encode $\mathbf{x}$ with $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}_1)$
  for $i = 1$ **to** $L - 1$ **do**
    encode $\mathbf{z}_i$ with $p_{\boldsymbol{\theta}}(\mathbf{z}_i|\mathbf{z}_{i+1})$
  end for
  encode $\mathbf{z}_L$ with $p(\mathbf{z}_L)$
until $\mathcal{D} = \varnothing$
Send: bitstream

Algorithm 2 **Bit-Swap (ours)** for lossless compression with hierarchical latent variables. The operations below show the procedure for encoding a dataset $\mathcal{D}$ onto a bitstream.

Input: data $\mathcal{D}$, depth L, $p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}_{1:L})$, $q_{\boldsymbol{\theta}}(\mathbf{z}_{1:L}|\mathbf{x})$
Require: ANS
Initialize: bitstream
repeat
  Take $\mathbf{x} \in \mathcal{D}$
  decode $\mathbf{z}_1$ with $q_{\boldsymbol{\theta}}(\mathbf{z}_1|\mathbf{x})$
  encode $\mathbf{x}$ with $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}_1)$
  for $i = 1$ **to** $L - 1$ **do**
    decode $\mathbf{z}_{i+1}$ with $q_{\boldsymbol{\theta}}(\mathbf{z}_{i+1}|\mathbf{z}_i)$
    encode $\mathbf{z}_i$ with $p_{\boldsymbol{\theta}}(\mathbf{z}_i|\mathbf{z}_{i+1})$
  end for
  encode $\mathbf{z}_L$ with $p(\mathbf{z}_L)$
until $\mathcal{D} = \varnothing$
Send: bitstream

Результаты

	MNIST	**CIFAR-10**	**ImageNet** (32×32)
Uncompressed	8.00	8.00	8.00
GNU Gzip	1.65	7.37	7.31
bzip2	1.59	6.98	7.00
LZMA	1.49	6.09	6.15
PNG	2.80	5.87	6.39
WebP	2.10	4.61	5.29
BB-ANS	1.48	4.19	4.66
Bit-Swap	**1.29**	**3.82**	**4.50**

Источники

- Friso H. Kingma, Pieter Abbeel, Jonathan Ho "Bit-Swap: Recursive Bits-Back Coding for Lossless Compression with Hierarchical Latent Variables"(2019)
- James Townsend, Thomas Bird, David Barber "PRACTICAL LOSSLESS COMPRESSION WITH LATENT VARIABLES USING BITS BACK CODING"(2019)

Вопросы

- Как BB-ANS кодирует x
- В чем заключается отличие Bit-Swap от BB-ANS?
- Верхняя оценка N_{init} у Bit-Swap и нижняя оценка у BB-ANS