Hướng dẫn: Mô hình khuếch tán

Diffusion models

Trần Quốc Long

Viện Trí tuệ nhân tạo - Trường ĐH Công nghệ

ĐHQG Hà Nội

July 1, 2023

Nội dung

Nội dung

1. Giới thiệu
   - Các công bố liên quan
   - Ý tưởng của mô hình khuếch tán
   - Một số kết quả
2. Mô hình khuếch tán
   - Cơ sở
   - Hàm mất mát = Cận dưới biến phân
   - Thuật toán huấn luyện
   - Thuật toán lấy mẫu
3. Các kỹ thuật bổ trợ

Các công bố chính

Related works

- (2015: Cơ sở toán học) Sohl-Dickstein, J. et. al., Deep unsupervised learning using nonequilibrium thermodynamics. In International Conference on Machine Learning (pp. 2256-2265).

Các công bố chính

Related works

- (2015: Cơ sở toán học) Sohl-Dickstein, J. et. al., Deep unsupervised learning using nonequilibrium thermodynamics. In International Conference on Machine Learning (pp. 2256-2265).
- (2020: Áp dụng với kỹ thuật reparameterization) Ho, J. et. al., Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.

Các công bố chính

Related works

- (2015: Cơ sở toán học) Sohl-Dickstein, J. et. al., Deep unsupervised learning using nonequilibrium thermodynamics. In International Conference on Machine Learning (pp. 2256-2265).
- (2020: Áp dụng với kỹ thuật reparameterization) Ho, J. et. al., Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.
- (2020: cải tiến cách lấy mẫu) Song, J. et. al. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502.

Các công bố chính

Related works

- (2015: Cơ sở toán học) Sohl-Dickstein, J. et. al., Deep unsupervised learning using nonequilibrium thermodynamics. In International Conference on Machine Learning (pp. 2256-2265).
- (2020: Áp dụng với kỹ thuật reparameterization) Ho, J. et. al., Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.
- (2020: cải tiến cách lấy mẫu) Song, J. et. al. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502.
- (2021: cải tiến cách chọn tham số) Nichol, A. Q. et. al. Improved denoising diffusion probabilistic models. In International Conference on Machine Learning (pp. 8162-8171).

Mô hình khuếch tán

Ý tưởng

- Quá trình khuếch tán xuôi (Forward diffusion):
  Dần dần phá vỡ cấu trúc của phân bố dữ liệu một cách hệ thống ⇒ sinh dữ liệu có nhãn

Mô hình khuếch tán

Ý tưởng

- Quá trình khuếch tán xuôi (Forward diffusion):
  Dần dần phá vỡ cấu trúc của phân bố dữ liệu một cách hệ thống ⇒ sinh dữ liệu có nhãn
- Quá trình khuếch tán ngược (Reverse diffusion):
  Học cách khôi phục phân bố dữ liệu ⇒ Học tự giám sát (self-supervised)

Mô hình khuếch tán
Ý tưởng

- Quá trình khuếch tán xuôi (Forward diffusion):
  Dần dần phá vỡ cấu trúc của phân bố dữ liệu một cách hệ thống ⇒ sinh dữ liệu có nhãn
- Quá trình khuếch tán ngược (Reverse diffusion):
  Học cách khôi phục phân bố dữ liệu ⇒ Học tự giám sát (self-supervised)
- Mô hình với biến ẩn:
  Hàm mất mát ⇔ Cận dưới biến phân (Variational Lower Bound)

Quá trình khuếch tán xuôi

Forward diffusion process

Dần dần phá vỡ cấu trúc của phân bố dữ liệu một cách hệ thống (bằng nhiễu Gauss)

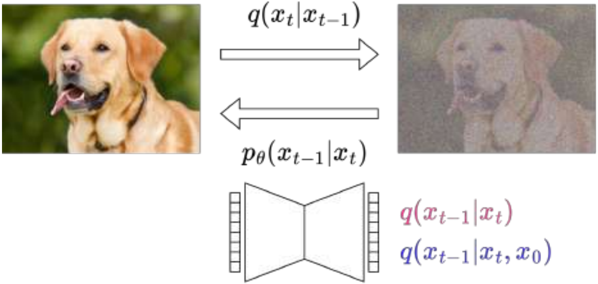
$$q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

Quá trình khuếch tán ngược

Reverse diffusion process

Học cách khôi phục phân bố dữ liệu (loại bỏ nhiễu)

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

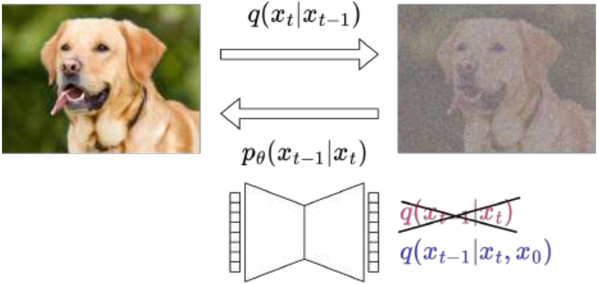


Quá trình khuếch tán ngược

Reverse diffusion process

Học cách khôi phục phân bố dữ liệu (loại bỏ nhiễu)

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$



Bảng xếp hạng

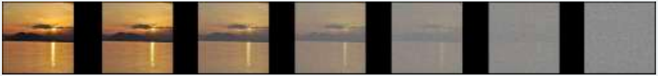
State of the art

Rank	Model	FID ↓	FD	Precision	Recall	sFID	Inception score	Paper	Code	Result	Year	Tags
1	MDT-XL/2	1.79					283.01	Masked Diffusion Transformer is a Strong Image Synthesizer			2023	
2	DiT-XL/2-G++	1.83		0.78	0.64	5.16	281.53	Refining Generative Process with Discriminator Guidance in Score-based Diffusion Models			2022	
3	ViT-XL	2.06						Efficient Diffusion Training via Min-SNR Weighting Strategy			2023	
4	DiT-XL/2	2.27					278.24	Scalable Diffusion Models with Transformers			2022	
5	StyleGAN-XL	2.3						StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets			2022	
6	Poly-INR	2.86						Polynomial Implicit Neural Representations For Large Diverse Datasets			2023	
7	ADM-G++ (FID)	3.18		0.84	0.53	4.53	255.74	Refining Generative Process with Discriminator Guidance in Score-based Diffusion Models			2022	
8	GigaGAN	3.45		0.84	0.61		225.52	Scaling up GANs for Text-to-Image Synthesis			2023	
9	BIGRoC-gt (Guided-Diffusion)	3.63					260.02	BIGRoC: Boosting Image Generation via a Robust Classifier			2021	
10	BIGRoC-pl	[illegible]					[illegible]	BIGRoC: Boosting Image			2021	

Nội dung

Quá trình khuếch tán xuôi

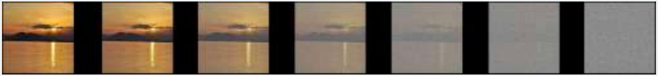
Thêm nhiễu Gauss vào dữ liệu



- Quá trình ngẫu nhiên: Biến đổi phân bố $\mathcal{P}$ chưa biết về phân bố chuẩn.

Quá trình khuếch tán xuôi

Thêm nhiễu Gauss vào dữ liệu



$$x_0 \sim \mathcal{P}$$

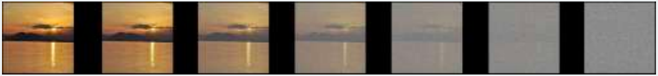
$$x_t = \sqrt{1-\beta_t} x_{t-1} + \sqrt{\beta_t}\epsilon \sim q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t} x_{t-1}, \beta_t I)$$

Với $\beta_t \in (0,1)$ thì $x_t|x_0 \xrightarrow{D} \mathcal{N}(0, I)$.

- Quá trình ngẫu nhiên: Biến đổi phân bố $\mathcal{P}$ chưa biết về phân bố chuẩn.
- Học máy: sinh ra dữ liệu gồm đầu vào x_t và nhãn mới ϵ.

Quá trình khuếch tán xuôi

Thêm nhiễu Gauss vào dữ liệu



$x_0 \sim \mathcal{P}$

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon \sim q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

Với $\beta_t \in (0,1)$ thì $x_t|x_0 \xrightarrow{D} \mathcal{N}(0, I)$.

- Quá trình ngẫu nhiên: Biến đổi phân bố $\mathcal{P}$ chưa biết về phân bố chuẩn.
- Học máy: sinh ra dữ liệu gồm đầu vào x_t và nhãn mới ϵ.
- Nếu có thể học được mô hình (x_t, ϵ) thì có thể lật ngược (undo) quá trình ngẫu nhiên (tính x_{t-1}).

Quá trình khuếch tán xuôi

Bổ đề 1: Lấy mẫu trực tiếp từ x_0 (nice property)

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} \epsilon$$

với $\alpha_t = 1 - \beta_t$ và $\overline{\alpha}_t = \prod_{i=1}^{t} \alpha_t$.

Quá trình khuếch tán xuôi

Bổ đề 1: Lấy mẫu trực tiếp từ x_0 (nice property)

$$\begin{aligned} x_t &= \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon \\ &= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t}\epsilon + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon' \end{aligned}$$

với $\alpha_t = 1 - \beta_t$ và $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t$.

Quá trình khuếch tán xuôi

Bổ đề 1: Lấy mẫu trực tiếp từ x_0 (nice property)

$$\begin{aligned} x_t &= \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon \\ &= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t}\epsilon + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon' \\ &= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t + \alpha_t(1-\alpha_{t-1})}\epsilon'' \end{aligned}$$

với $\alpha_t = 1 - \beta_t$ và $\overline{\alpha}_t = \prod_{i=1}^{t} \alpha_t$.

Quá trình khuếch tán xuôi

Bổ đề 1: Lấy mẫu trực tiếp từ x_0 (nice property)

$$
\begin{aligned}
x_t &= \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon \\
&= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t}\epsilon + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon' \\
&= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t + \alpha_t(1-\alpha_{t-1})}\epsilon'' \\
&= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}}\epsilon''
\end{aligned}
$$

với $\alpha_t = 1 - \beta_t$ và $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t$.

Quá trình khuếch tán xuôi

Bổ đề 1: Lấy mẫu trực tiếp từ x_0 (nice property)

$$
\begin{aligned}
x_t &= \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon \\
&= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t}\epsilon + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon' \\
&= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t + \alpha_t(1-\alpha_{t-1})}\epsilon'' \\
&= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\epsilon'' \\
&= \sqrt{\prod_{i=1}^{t} \alpha_i} x_0 + \sqrt{1-\prod_{i=1}^{t}\alpha_i}\epsilon_t \\
&= \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t \quad \sim \mathcal{N}(\sqrt{\bar{\alpha}_t} x_0, (1-\bar{\alpha}_t)I)
\end{aligned}
$$

với $\alpha_t = 1 - \beta_t$ và $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t$.

Quá trình khuếch tán ngược

Xác suất liên hợp

$$q(x_0, x_1, \ldots, x_T) = q(x_0) \prod_{t=1}^{T} q(x_t | x_{t-1}) \leftarrow \text{(xuôi)}$$

Quá trình khuếch tán ngược

Xác suất liên hợp

$$q(x_0, x_1, \ldots, x_T) = q(x_0) \prod_{t=1}^{T} q(x_t|x_{t-1}) \leftarrow \text{(xuôi)}$$

- $q(x_0) = \mathcal{P} \rightarrow$ chưa biết, nhưng có dữ liệu
- $q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I) \rightarrow$ đã biết

Quá trình khuếch tán ngược

Xác suất liên hợp

$$\begin{aligned} q(x_0, x_1, \ldots, x_T) &= q(x_0) \prod_{t=1}^{T} q(x_t | x_{t-1}) \leftarrow \text{(xuôi)} \\ &= q(x_T) \prod_{t=1}^{T} q(x_{t-1} | x_t) \leftarrow \text{(ngược)} \end{aligned}$$

- $q(x_0) = \mathcal{P} \rightarrow$ chưa biết, nhưng có dữ liệu
- $q(x_t | x_{t-1}) = \mathcal{N}(\sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \rightarrow$ đã biết

Quá trình khuếch tán ngược

Xác suất liên hợp

$$\begin{aligned} q(x_0, x_1, \ldots, x_T) &= q(x_0) \prod_{t=1}^{T} q(x_t | x_{t-1}) \leftarrow \text{(xuôi)} \\ &= q(x_T) \prod_{t=1}^{T} q(x_{t-1} | x_t) \leftarrow \text{(ngược)} \end{aligned}$$

- $q(x_0) = \mathcal{P} \rightarrow$ chưa biết, nhưng có dữ liệu
- $q(x_t | x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t} x_{t-1}, \beta_t I) \rightarrow$ đã biết
- $q(x_{t-1} | x_t) \rightarrow$ chưa biết

Quá trình khuếch tán ngược

Xác suất liên hợp

$$
\begin{aligned}
q(x_0, x_1, \ldots, x_T) &= q(x_0) \prod_{t=1}^{T} q(x_t | x_{t-1}) \leftarrow \text{(xuôi)} \\
&= q(x_T) \prod_{t=1}^{T} q(x_{t-1} | x_t) \leftarrow \text{(ngược)} \\
&= q(x_T | x_0) \prod_{t=1}^{T} q(x_{t-1} | x_t, x_0) q(x_0) \leftarrow \text{(ngược)}
\end{aligned}
$$

- $q(x_0) = \mathcal{P} \rightarrow$ chưa biết, nhưng có dữ liệu
- $q(x_t | x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t} x_{t-1}, \beta_t I) \rightarrow$ đã biết
- $q(x_{t-1} | x_t) \rightarrow$ chưa biết

Quá trình khuếch tán ngược

Xác suất liên hợp

$$
\begin{aligned}
q(x_0, x_1, \ldots, x_T) &= q(x_0) \prod_{t=1}^{T} q(x_t | x_{t-1}) \leftarrow (\text{xuôi}) \\
&= q(x_T) \prod_{t=1}^{T} q(x_{t-1} | x_t) \leftarrow (\text{ngược}) \\
&= q(x_T | x_0) \prod_{t=1}^{T} q(x_{t-1} | x_t, x_0) q(x_0) \leftarrow (\text{ngược})
\end{aligned}
$$

- $q(x_0) = \mathcal{P} \rightarrow$ chưa biết, nhưng có dữ liệu
- $q(x_t | x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t} x_{t-1}, \beta_t I) \rightarrow$ đã biết
- $q(x_{t-1} | x_t) \rightarrow$ chưa biết
- $q(x_T | x_0) \approx \mathcal{N}(0, I) \rightarrow$ đã biết
- $q(x_{t-1} | x_t, x_0) \rightarrow$ có thể tính được nhờ bổ đề 1

Quá trình khuếch tán ngược

Quá trình ngược $q(x_{t-1}|x_t)$ chưa biết nhưng nếu biết x_0 thì

Bổ đề 2: $q(x_{t-1}|x_t, x_0)$ (reparameterization trick)

$$q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1})q(x_{t-1}|x_0)}{q(x_t|x_0)} =$$

Quá trình khuếch tán ngược

Quá trình ngược $q(x_{t-1}|x_t)$ chưa biết nhưng nếu biết x_0 thì

Bổ đề 2: $q(x_{t-1}|x_t, x_0)$ (reparameterization trick)

$$q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1})q(x_{t-1}|x_0)}{q(x_t|x_0)} = \mathcal{N}(\tilde{\mu}_t, \tilde{\beta}_t I)$$

số mũ có dạng $-\frac{1}{2\tilde{\beta}_t}(x_{t-1} - \tilde{\mu}_t)^2$ (completing the squares) với

Quá trình khuếch tán ngược

Quá trình ngược $q(x_{t-1}|x_t)$ chưa biết nhưng nếu biết x_0 thì

Bổ đề 2: $q(x_{t-1}|x_t, x_0)$ (reparameterization trick)

$$q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1})q(x_{t-1}|x_0)}{q(x_t|x_0)} = \mathcal{N}(\tilde{\mu}_t, \tilde{\beta}_t I)$$

số mũ có dạng $-\frac{1}{2\tilde{\beta}_t}(x_{t-1} - \tilde{\mu}_t)^2$ (completing the squares) với

$$\tilde{\mu}_t = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0$$

Quá trình khuếch tán ngược

Quá trình ngược $q(x_{t-1}|x_t)$ chưa biết nhưng nếu biết x_0 thì

Bổ đề 2: $q(x_{t-1}|x_t, x_0)$ (reparameterization trick)

$$q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1})q(x_{t-1}|x_0)}{q(x_t|x_0)} = \mathcal{N}(\tilde{\mu}_t, \tilde{\beta}_t I)$$

số mũ có dạng $-\frac{1}{2\tilde{\beta}_t}(x_{t-1} - \tilde{\mu}_t)^2$ (completing the squares) với

$$\tilde{\mu}_t = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} \quad \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_t\right)$$

Quá trình khuếch tán ngược

Quá trình ngược $q(x_{t-1}|x_t)$ chưa biết nhưng nếu biết x_0 thì

Bổ đề 2: $q(x_{t-1}|x_t, x_0)$ (reparameterization trick)

$$q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1})q(x_{t-1}|x_0)}{q(x_t|x_0)} = \mathcal{N}(\tilde{\mu}_t, \tilde{\beta}_t I)$$

số mũ có dạng $-\frac{1}{2\tilde{\beta}_t}(x_{t-1} - \tilde{\mu}_t)^2$ (completing the squares) với

$$\tilde{\mu}_t = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} \quad \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_t\right)$$

$$= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t\right)$$

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

Ước lượng hợp lý cực đại

Maximum likelihood estimation - MLE

- Chọn mô hình $p_\theta(x_{0:T}) = p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)$ để xấp xỉ quá trình ngược
- Cần tìm θ để cực đại hoá độ hợp lý

$$\ell(\theta) = \log p_\theta(x_0) = \int_{x_{1:T}} p_\theta(x_{0:T})dx_{1:T}$$

trong đó x_0 (dữ liệu) là biến quan sát được, còn $x_1, x_2, \ldots, x_T$ là các biến ẩn (không quan sát được).

Cận dưới biến phân

Evidence lower bound - ELBO

$$\log p_\theta(x_0) \geq \log p_\theta(x_0) - D_{KL}(q(x_{1:T}|x_0) \| p_\theta(x_{1:T}|x_0))$$

Cận dưới biến phân

Evidence lower bound - ELBO

$$\begin{aligned}\log p_\theta(x_0) &\geq \log p_\theta(x_0) - D_{KL}\left(q(x_{1:T}|x_0) \| p_\theta(x_{1:T}|x_0)\right) \\ &= \log p_\theta(x_0) - \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{1:T}|x_0)} \frac{p_\theta(x_0)}{p_\theta(x_0)}\right]\end{aligned}$$

Cận dưới biến phân

Evidence lower bound - ELBO

$$
\begin{aligned}
\log p_\theta(x_0) &\geq \log p_\theta(x_0) - D_{KL}(q(x_{1:T}|x_0) \| p_\theta(x_{1:T}|x_0)) \\
&= {\color{blue}\log p_\theta(x_0)} - \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{1:T}|x_0)} \frac{{\color{blue}p_\theta(x_0)}}{p_\theta(x_0)}\right] \\
&= -\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] = \text{ELBO}
\end{aligned}
$$

Cận dưới biến phân

Evidence lower bound - ELBO

$$
\begin{aligned}
\log p_\theta(x_0) &\geq \log p_\theta(x_0) - D_{KL}(q(x_{1:T}|x_0) \| p_\theta(x_{1:T}|x_0)) \\
&= \log p_\theta(x_0) - \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{1:T}|x_0)} \frac{p_\theta(x_0)}{p_\theta(x_0)}\right] \\
&= -\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] = \text{ELBO}
\end{aligned}
$$

Hàm mất mát cross-entropy

$$
\mathcal{L}_{CE}(\theta) = -\mathbb{E}_{q(x_0)}[\log p_\theta(x_0)]
$$

Cận dưới biến phân

Evidence lower bound - ELBO

$$
\begin{aligned}
\log p_\theta(x_0) &\geq \log p_\theta(x_0) - D_{KL}(q(x_{1:T}|x_0) \| p_\theta(x_{1:T}|x_0)) \\
&= \color{blue}{\log p_\theta(x_0)} \color{black}{} - \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{1:T}|x_0)} \frac{\color{blue}{p_\theta(x_0)}}{p_\theta(x_0)}\right] \\
&= -\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] = \text{ELBO}
\end{aligned}
$$

Hàm mất mát cross-entropy

$$
\begin{aligned}
\mathcal{L}_{CE}(\theta) &= -\mathbb{E}_{q(x_0)}[\log p_\theta(x_0)] \\
&\leq \mathbb{E}_{q(x_0)}\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right]
\end{aligned}
$$

Cận dưới biến phân

Evidence lower bound - ELBO

$$
\begin{aligned}
\log p_\theta(x_0) &\geq \log p_\theta(x_0) - D_{KL}(q(x_{1:T}|x_0)\|p_\theta(x_{1:T}|x_0)) \\
&= \color{blue}{\log p_\theta(x_0)} \color{black} - \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{1:T}|x_0)} \frac{\color{blue}{p_\theta(x_0)}}{p_\theta(x_0)}\right] \\
&= -\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] = \mathrm{ELBO}
\end{aligned}
$$

Hàm mất mát cross-entropy

$$
\begin{aligned}
\mathcal{L}_{CE}(\theta) &= -\mathbb{E}_{q(x_0)}[\log p_\theta(x_0)] \\
&\leq \mathbb{E}_{q(x_0)}\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] \\
&= \color{red}{\mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] = \mathcal{L}_{VLB}(\theta)}
\end{aligned}
$$

Hàm mất mát

Evidence lower bound - ELBO

$$\begin{aligned}\mathcal{L}_{\mathrm{CE}}(\theta) &= -\mathbb{E}_{q(x_0)}[\log p_\theta(x_0)] \\ &\le \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] = \mathcal{L}_{\mathrm{VLB}}(\theta)\end{aligned}$$

- $-\log p_\theta(x_0)$: khó tính được
- $\log \dfrac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$: tính được, là cận trên của $-\log p_\theta(x_0)$

Hàm mất mát

$$\mathcal{L}_{\mathrm{VLB}}(\theta) = \mathbb{E}_{\mathrm{q}(\mathrm{x}_{0:\mathrm{T}})}\left[\log \frac{\mathrm{q}(\mathrm{x}_{1:\mathrm{T}}|\mathrm{x}_0)}{\mathrm{p}_\theta(\mathrm{x}_{0:\mathrm{T}})}\right]$$

Hàm mất mát

$$
\begin{aligned}
\mathcal{L}_{\mathrm{VLB}}(\theta) &= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_T|x_0)}{p_\theta(x_T)} + \log \frac{q(x_{T-1}|x_T, x_0)}{p_\theta(x_{T-1}|x_T)} + \dots \right. \\
&\qquad \left. + \log \frac{q(x_1|x_2, x_0)}{p_\theta(x_1|x_2)} - \log p_\theta(x_0|x_1)\right]
\end{aligned}
$$

Hàm mất mát

$$
\begin{aligned}
\mathcal{L}_{\mathrm{VLB}}(\theta) &= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_T|x_0)}{p_\theta(x_T)} + \log \frac{q(x_{T-1}|x_T, x_0)}{p_\theta(x_{T-1}|x_T)} + \ldots \right. \\
&\qquad \left. + \log \frac{q(x_1|x_2, x_0)}{p_\theta(x_1|x_2)} - \log p_\theta(x_0|x_1)\right] \\
&= \mathcal{L}_T + \mathcal{L}_{T-1} + \ldots + \mathcal{L}_1 + \mathcal{L}_0
\end{aligned}
$$

Hàm mất mát

$$\begin{aligned}\mathcal{L}_{\mathrm{VLB}}(\theta) &= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] \\ &= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_T|x_0)}{p_\theta(x_T)} + \log \frac{q(x_{T-1}|x_T, x_0)}{p_\theta(x_{T-1}|x_T)} + \ldots \right. \\ &\qquad \left. + \log \frac{q(x_1|x_2, x_0)}{p_\theta(x_1|x_2)} - \log p_\theta(x_0|x_1)\right] \\ &= \mathcal{L}_T + \mathcal{L}_{T-1} + \ldots + \mathcal{L}_1 + \mathcal{L}_0\end{aligned}$$

- $\mathcal{L}_T = D_{KL}(q(x_T|x_0) \| p_\theta(x_T)) \approx \mathrm{const}$

Hàm mất mát

$$\begin{aligned}
\mathcal{L}_{\mathrm{VLB}}(\theta) &= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_T|x_0)}{p_\theta(x_T)} + \log \frac{q(x_{T-1}|x_T, x_0)}{p_\theta(x_{T-1}|x_T)} + \ldots \right. \\
&\qquad \left. + \log \frac{q(x_1|x_2, x_0)}{p_\theta(x_1|x_2)} - \log p_\theta(x_0|x_1)\right] \\
&= \mathcal{L}_T + \mathcal{L}_{T-1} + \ldots + \mathcal{L}_1 + \mathcal{L}_0
\end{aligned}$$

- $\mathcal{L}_T = D_{KL}(q(x_T|x_0) \| p_\theta(x_T)) \approx \text{const}$
- $\mathcal{L}_{t-1} = D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t))$: phần chính của hàm mất mát

Hàm mất mát

$$
\begin{aligned}
\mathcal{L}_{\mathrm{VLB}}(\theta) &= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_T|x_0)}{p_\theta(x_T)} + \log \frac{q(x_{T-1}|x_T, x_0)}{p_\theta(x_{T-1}|x_T)} + \ldots \right. \\
&\qquad \left. + \log \frac{q(x_1|x_2, x_0)}{p_\theta(x_1|x_2)} - \log p_\theta(x_0|x_1)\right] \\
&= \mathcal{L}_T + \mathcal{L}_{T-1} + \ldots + \mathcal{L}_1 + \mathcal{L}_0
\end{aligned}
$$

- $\mathcal{L}_T = D_{KL}(q(x_T|x_0)\|p_\theta(x_T)) \approx \mathrm{const}$
- $\mathcal{L}_{t-1} = D_{KL}(q(x_{t-1}|x_t, x_0)\|p_\theta(x_{t-1}|x_t))$: phần chính của hàm mất mát
- $\mathcal{L}_0 = -\log p_\theta(x_0|x_1)$ có thể bỏ qua khi mạng được huấn luyện tốt

Hàm mất mát

- Chọn $p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta(x_t, t), \sigma_t^2 I)$

$$\mathcal{L}_t = D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t))$$

là khoảng cách KL giữa 2 phân bố chuẩn

Hàm mất mát

- Chọn $p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta(x_t, t), \sigma_t^2 I)$
$$\mathcal{L}_t = D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t))$$
là khoảng cách KL giữa 2 phân bố chuẩn
- Có thể xấp xỉ bằng khoảng cách giữa kì vọng $\|\tilde{\mu}_t - \mu_\theta(x_t, t)\|^2$ với $\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t\right)$

Hàm mất mát

- Chọn $p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta(x_t, t), \sigma_t^2 I)$

$$\mathcal{L}_t = D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t))$$

  là khoảng cách KL giữa 2 phân bố chuẩn
- Có thể xấp xỉ bằng khoảng cách giữa kì vọng $\|\tilde{\mu}_t - \mu_\theta(x_t, t)\|^2$ với $\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t\right)$
- Chọn $\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right)$ thì có xấp xỉ

$$\mathcal{L}_t^{\text{simple}} = \|\epsilon_t - \epsilon_\theta(x_t, t)\|^2$$

Hàm mất mát

- Chọn $p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta(x_t, t), \sigma_t^2 I)$

$$\mathcal{L}_t = D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t))$$

là khoảng cách KL giữa 2 phân bố chuẩn

- Có thể xấp xỉ bằng khoảng cách giữa kì vọng $\|\tilde{\mu}_t - \mu_\theta(x_t, t)\|^2$ với $\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t\right)$
- Chọn $\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right)$ thì có xấp xỉ

$$\mathcal{L}_t^{\text{simple}} = \|\epsilon_t - \epsilon_\theta(x_t, t)\|^2$$

- $\epsilon_\theta(x_t, t) \Leftrightarrow$ mạng nơ-ron xấp xỉ nhiễu ϵ_t đã "làm hỏng" dữ liệu gốc x_0 và tạo ra dữ liệu nhiễu

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t$$

Thuật toán huấn luyện

Bài toán huấn luyện (training)

- Đầu vào: bộ dữ liệu thuộc phân bố $\mathcal{P}$
- Đầu ra: mô hình nhiễu $\epsilon_\theta(x_t, t)$

Một bước (step) huấn luyện

1. Lấy mẫu $x_0 \sim \mathcal{P}$
2. Lấy mẫu $t \sim \mathcal{U}[1 \ldots T]$
3. Sinh nhiễu $\epsilon_t \sim \mathcal{N}(0, I)$
4. Lấy mẫu $x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t} \epsilon_t$
5. Xuống đồi bằng đạo hàm $\nabla_\theta \|\epsilon_t - \epsilon_\theta(x_t, t)\|^2$

Thuật toán lấy mẫu

Bài toán lấy mẫu (sampling)

- Đầu vào: mô hình nhiễu $\epsilon_\theta(x_t, t)$
- Đầu ra: mẫu $x_0 \sim \mathcal{P}$

Lấy mẫu

1. Lấy mẫu $x_T \sim \mathcal{N}(0, I)$
2. For $t = T, T-1, \ldots, 1$:
3. Lấy mẫu $z \sim \mathcal{N}(0, I)$ nếu $t > 1$, ngược lại $z = 0$
4. Tính

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z$$

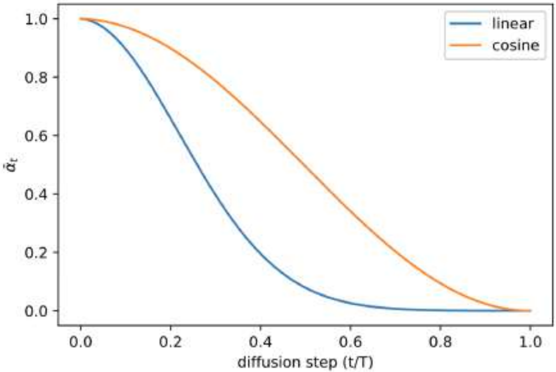
Nội dung

Lựa chọn lịch trình β_t

$$f(t) = \cos\left(\frac{t/T + s}{1+s}\frac{\pi}{2}\right)^2$$

$$\overline{\alpha}_t = \frac{f(t)}{f(0)}$$

$$\beta_t = 1 - \alpha_t = 1 - \frac{\overline{\alpha}_t}{\overline{\alpha}_{t-1}}$$



Lựa chọn lịch trình β_t

Tuyến tính

Cosine

Lấy mẫu nhanh

Song, J. et.al . Denoising diffusion implicit models (DDIM). ICLR 2020

Ý tưởng: sử dụng chuỗi "không Markov"

- Cho phép lấy mẫu từ tập con $\{\tau_1, \ldots, \tau_S\}$ của các bước $\{1, 2, \ldots, T\}$
- Dùng chung hàm mất mát như DDPM
- Quá trình sinh ảnh đơn định
- Cho phép nội suy từ nhiều ảnh nhiễu ban đầu

Sinh dữ liệu có điều kiện $q(x|y)$

Ý tưởng: chuyển mọi công thức với điều kiện y

- $p_\theta(x_{t-1}|x_t) \longrightarrow p_\theta(x_{t-1}|x_t, y)$ hoặc $p_\theta(x_{t-1}|x_t, \tau_\theta(y))$
- $\epsilon_\theta(x_t, t) \longrightarrow \epsilon_\theta(x_t, t, y)$ hoặc $\epsilon_\theta(x_t, t, \tau_\theta(y))$

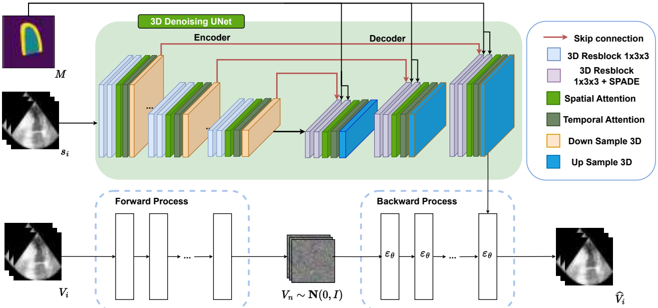
Ví dụ về nhãn y

- (Khôi phục ảnh) Ảnh đen trắng, ảnh bị mất một phần, ảnh độ phân giải thấp
- (Text-to-X) mô tả bằng văn bản của ảnh, âm thanh, video

Sinh video siêu âm tim

Ý tưởng

- Sử dụng phân tách Attention theo chiều không gian và thời gian riêng biệt.



Sinh video siêu âm tim