

Denoising Diffusion Probabilistic Models

Рецензия

Лысенко Иван 193

Outline

- 1 Paper Overview
- 2 Denoising Diffusion Probabilistic Models (DDPM) Overview
- 3 Predecessor: Diffusion Probabilistic Model (DPM)
- 4 Successors
- 5 Плюсы работы
- 6 Минусы работы
- 7 Материалы для погружения

Paper Overview

- **Название:** Denoising Diffusion Probabilistic Models
- **Дата:** 16 декабря 2020
- **Конференция:** NeurIPS 2020 (Part of Advances in Neural Information Processing Systems 33 (NeurIPS 2020))
- **Авторы:**
 - ▶ Jonathan Ho (UC Berkeley)
 - ★ Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding (Saharia et al. (2022))
 - ★ Variational Diffusion Models (Kingma et al. (2021))
 - ▶ Ajay Jain (UC Berkeley)
 - ★ Dreamfusion: Text-to-3d using 2d diffusion (Poole et al. (2022))
 - ▶ Pieter Abbeel (UC Berkeley)

Denoising Diffusion Probabilistic Models (DDPM)

Overview

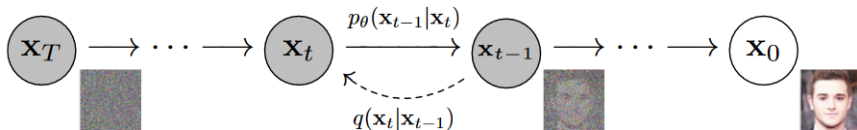


Figure 2: The directed graphical model considered in this work.

Algorithm 1 Training

```

1: repeat
2:    $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ 
3:    $t \sim \text{Uniform}(\{1, \dots, T\})$ 
4:    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
5:   Take gradient descent step on
      $\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)\|^2$ 
6: until converged
    
```

Algorithm 2 Sampling

```

1:  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
2: for  $t = T, \dots, 1$  do
3:    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  if  $t > 1$ , else  $\mathbf{z} = \mathbf{0}$ 
4:    $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$ 
5: end for
6: return  $\mathbf{x}_0$ 
    
```

Predecessor: Diffusion Probabilistic Model

Берем модель из Sohl-Dickstein et al. (2015), но:

- Используем нормальное распределение на каждом шаге
- Фиксируем линейное расписание с константными (малыми) β_1, \dots, β_T
- Фиксируем дисперсию $\Sigma_\theta(x_t, t) = \sigma_t^2 I, \sigma_t = \text{const}(t)$ из параметризации обратного диффузионного процесса
- Репараметризуем $\mu_\theta(x_t, t)$, чтобы предсказывать добавленный шум, для упрощения функции потерь и процесса обучения
- Используем в качестве модели U-Net (Ronneberger et al. (2015))

Successors

- Improved Denoising Diffusion Probabilistic Models (Nichol et al. (2021))
- Variational Diffusion Models (Kingma et al. (2021))

Плюсы работы

- Есть исходный код:
`https://github.com/hojonathanho/diffusion`
- Есть подробные выводы используемых формул
- Есть ablation study по различным вариантам параметризации обратного диффузионного процесса
- В appendix подробно указаны гиперпараметры и конкретная используемая архитектура модели




Минусы работы

- Нет экспериментов на разнообразных выборках типа ImageNet, соответственно нельзя сказать, как изменяется качество генерации с ростом diversity
- Нет ablation study для многих фиксированных параметров DDM (например, распределения или модели)
- Нет исследования/объяснения, почему правдоподобия для DDPM хуже, чем у других генеративных моделей
- Новизна не сильная, так как статья сильно базируется на более старой (Sohl-Dickstein et al. (2015))
- Введение в диффузионные модели представлено довольно коротко, и его сложно понять, не разбираясь в теме. Из-за этого текст довольно сложно читать.

Материалы для погружения

- Простое введение в статью с примерами кода:
<https://huggingface.co/blog/annotated-diffusion>
- Пайплайн в Hugging Face Diffusers: <https://huggingface.co/docs/diffusers/api/pipelines/ddpm>

Источники (1)

-  Sohl-Dickstein, Jascha and Weiss, Eric A. and Maheswaranathan, Niru and Ganguli, Surya. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In Proceedings of the 32nd International Conference on Machine Learning, 2015.
-  Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015.
-  Kingma, Diederik P. and Salimans, Tim and Poole, Ben and Ho, Jonathan. Variational Diffusion Models. In Advances in Neural Information Processing Systems 34 (NeurIPS 2021), 2021.

Источники (2)



Saharia, Chitwan and Chan, William and Saxena, Saurabh and Li, Lala and Whang, Jay and Denton, Emily and Ghasemipour, Seyed Kamyar Seyed and Ayan, Burcu Karagol and Mahdavi, S. Sara and Lopes, Rapha Gontijo and Salimans, Tim and Ho, Jonathan and Fleet, David J and Norouzi, Mohammad. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. 2022.



Poole, Ben and Jain, Ajay and Barron, Jonathan T. and Mildenhall, Ben. DreamFusion: Text-to-3D using 2D Diffusion. 2022.



Nichol, Alex and Dhariwal, Prafulla. Improved Denoising Diffusion Probabilistic Models. 2021.