

DDPM Рецензия

Поздеев Дмитрий

В рецензии будет дан обзор статьи [Ho et al. \[2020\]](#) "Denoising Diffusion Probabilistic Models"(DDPM). Статья была опубликована в 2020 на NIPS группой авторов, состоящей из J Ho, A Jain, P Abbeel, университет UC Berkeley. J Ho, A Jain, P Abbeel. Последний является главой лаборатории, в которой на момент публикации работали авторы. На текущий момент J Ho и A Jain продолжают работу в области диффузионных моделей и опубликовали множество узнаваемых статей.

В статье применяются диффузионные вероятностные модели, предложенные в [Sohl-Dickstein et al. \[2015\]](#). Авторы реализовали предложенную раннюю модель на современный лад, возродив интерес к диффузионным моделям. Основной вклад работы использование в качестве архитектуры UNet [Ronneberger et al., 2015](#) с self-attention, также авторы ввели упрощенную функцию потерь и показали связь введенного подхода с предыдущими работами, основанными на обучении градиента плотности распределения. Скорее всего авторы наткнулись на предыдущую работу [Sohl-Dickstein et al. \[2015\]](#), и решили улучшить качество, используя более современные инженерные подходы. DDPM заимствует вероятностную модель, а также способ минимизации ошибки, путем перехода к нижней вариационной оценке.

Статья является основой для современных диффузионных моделей, получивших широкое распространение за последнее время ([репозиторий](#)). Особый интерес к диффузионным моделям появился после выхода модели DALL-E-2 [\[Ramesh et al., 2022\]](#) от OpenAI для генерации картинок по тексту.

Стоит отметить, что между предыдущей работой про диффузионные модели и DDPM прошло 5 лет, в это время [Янг Сонг](#) активно развивал альтернативный подход, основанный на обучении градиента функции плотности. В DDPM показывается связь с данным подходом, при использовании более простой функции потерь. Далее Сонг и коллабораторы в [Song et al. \[2020b\]](#) показали, что данные подходы являются дискретизацией стохастических дифференциальных уравнений.

Сильные стороны Сильными сторонами работы являются значимость и актуальность для исследовательского сообщества. DDPM послужила большим толчком для будущих улучшений и применений диффузионных моделей, кроме того введение более простой функции потерь соединило предыдущие работы по обучению градиентов плотности. Текст хоть и является полным и строгим, но может вызывать затруднения для неподготовленного читателя. К плюсам данной работы можно добавить аккуратно проделанные эксперименты - статья сравнивается с основными подходами для генерации изображений, включая модели основанные на оптимизации правдоподобия. Результаты воспроизводимы, детали по реализации указаны в аппендиксе, ссылка на репозиторий указана в статье.

Слабые стороны Статья не предлагает новые решения, а реализовывает предыдущую работу [Sohl-Dickstein et al. \[2015\]](#). К новизне можно только добавить инженерные решения, а также связь вывода с предыдущими работами. Кроме того авторы используют решения, оптимальность которых, не исследуют: фиксирование линейного расписания

на дисперсию прямого процесса, не обучаемая дисперсия обратного процесса, фиксирование числа шагов в процессе. Все они будут исследованы в следующих работах. Также предложенная модель имеет низкое правдоподобие по сравнению с остальными моделями из семейства лог-правдоподобных моделей, о чем авторы честно признаются. Также в работе не была затронута одна из основных проблем диффузионных моделей - скорость генерации изображений.

Улучшения статьи нашли своих исследователей [Nichol and Dhariwal, 2021], [Ramesh et al., 2022], [Song et al. [2020a] так, например, расписание лучше брать косинусное, а не линейное, параметры дисперсий обратного процесса лучше учить, а скорости можно добиться путем пропуска шагов, смены моделей на немарковскую и введение латентных пространств. Интересно было бы посмотреть на результаты диффузии, имеющего в качестве прямого процесса не только нормальное распределение.

Автору рецензии интересно применение диффузии для связи двух сложных распределений (по аналогии с Schrödinger Bridge Problem). Так, например, мост между исходной и целевой задачами могла бы помочь решить проблемы трансферного обучения, например, катастрофическое забывание и искажение признаков при дообучении.

Список литературы

- J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.
- A. Q. Nichol and P. Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.
- A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022.
- O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.
- J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.
- Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.