Denoising Diffusion Probabilistic Models

Бобков Денис 192

October 2022

1 Авторы

Jonathan Ho: основные интересы: RL и unsupervised learning, статья является его PHD работой, после этого также выпускал достаточно много статей по дифузионным моделям. Его страничка: http://www.jonathanho.me

Ajay Jain: занимается в основном генеративным моделями: дифузионными и в 3d. Его страничка: https://www.ajayjain.net

Pieter Abbeel: изучаем различные RL, Apprenticeship Learning и Meta-Learning подходы, использующиеся в ИИ для роботов. Его страничка: https://people.eecs.berkeley.edu/ pabbeel/

Все авторы из калифорнийского университета Berkeley из лаборатории, специализирующейся на ии для роботов (которой руководит Pieter Abbeel).

2 Предметная область

Предметная область статьи - генерация изображений. Ранние подходы представляли из себя примерно следующее: **PixelCNN** - создавали изображения неплохого качества, но работали очень долго, и к тому же не получали никакого латентного пространства. **VAE** - автоэнкодеры, быстро обучаются, но создают очень заблюренные изображения. Ранние модели **GAN**'ов, имели достаточно

хорошее качество, но процесс обучения был довольно нестабилен, и работал в основном с изображениями небольшого разрешения. Были также и гибридные методы, объединяющие особенности всех предыдущих, но их качество зачастую было меньше обычных GAN моделей.

Витком развития области стала эволюция GAN моделей, а именно выход статей про Big GAN, в которой предлагается очень большая GAN версия, являвшаяся в своё время sota методом на ImageNet, а также методы стабилизации обучения; семейство статей про Style GAN (Style GAN, Style GAN 2, Style GAN 2 ADA) также стали очень популярны, их архитектуры показывали очень хорошее качество и стали основой для многих моделей.

3 Диффузионные модели

Одной из первых статей в области диффузионных моделей была Deep Unsupervised Learning using Nonequilibrium Thermodynamics, в которой как раз описывается последовательный диффузионный процесс зашумления/раззушумления данных. Именно эта статья стала основой для нашей статьи, и породила один из основных подходов в дифузионных моделях.

Другой подход, также часто используемый в диффузионныз моделях использует некую score function, и данамику Ланжевена для генерации изображений. Если кратко описывать метод, то суть его в следующем: пусть у нас есть какое то истинное распределние данных $p(x)$, хотим построить для него модель $p_\theta(x)$, которая по изображению на вход будет выдавать вероятность того, что изображения настоящее. Зададим $p_\theta(x)$ как $p_\theta(x) = \frac{e^{-f_\theta(x)}}{Z_\theta}$, где $f_\theta(x)$ - это некий функционал энергии (к примеру нейросеть), а $Z_\theta = \int e^{-f_\theta(x)}dx$ - нормировочная константа, необходимая для того, чтобы $0 \leq p_\theta(x) \leq 1$.

Подобные статистические модели зачастую обучаются максимизацией логарифма правдоподобия, но если расписать это для $p_\theta(x)$, то получим $\mathbb{E} - \log(p_\theta(x)) = \mathbb{E}(f_\theta(x) + \log Z_\theta)$, и т.к. Z яв-

ляется интегралом, то считать это очень сложно. Чтобы обойти этот момент, и найти веса θ вводят специальную **score function** $\mathrm{s(x)} = \nabla_x \log p(x)$. Для $p_\theta(x)$ $s_\theta(x) = -\nabla_x f_\theta(x)$, то есть если можно оценить реальную $p(x)$, то можно оптимизировать именно s_θ на функционале $\mathbb{E}\|s_\theta(x) - \nabla_x p(x)\|_2^2$. А как оценивать $p(x)$ - описывают различные статьи.

Но остаётся один вопрос, как при имеющейся $p_\theta(x)$ генерировать новые изображения? Предлагается использовать динамику Ланжевена, а именно: $x_{i+1} = x_i + \varepsilon \nabla_x \log p(x) + \sqrt{2\varepsilon} z_i, i = 0, 1, ..., K$, где $z_i \sim \mathcal{N}(0, I)$, и $K \to \infty$, то есть мы с каждым шагом направляемся в сторону наибольшего возрастания плотности с поправкой на случайное направление z_i.

Основным конкурентом нашей статьи, использующий вышеописанный метод была статься NCSN.

4 Резюме

Подводя итоги, в области диффузионных моделей есть две основных ветки развития: DDPM (наша статья) и score matchinng + Langevin dynamic (NCSL), обе ветки на данный момент успешно развиваются, но DDPM является несколько более популярной. Как основные преимущества дифузионных моделей можно выделить достаточно высокую стабильность обучения и очень хорошее качество генерируемых изображений. Однако для таких результатов моделям необходимо потратить множество времени на обучение.

Область в настоящий момент является крайне прогрессивной, и развивается практически во всех направлениях. Авторы выпустили также продолжение статьи - Improved DDPM, в которой предлагают более точные оценки. Примеры работы диффузионных моделей можно найти и для генерации видео, и для генерации изображений, и для генерации 3d моделей, и для аудио, и во множестве других областей.