Практика

Диффузионные модели

Tуем Гислен

11-04-2024

Содержание

Список иллюстраций

Список таблиц

# 1 Введение

\*\*Актуальность темы\*\*:   
Диффузионные модели революционизировали генеративный ИИ, превзойдя GAN в качестве синтеза изображений (Stable Diffusion, Imagen). Их применение распространяется на медицину, NLP и физику.  
Наряду с GAN, модели диффузии (diffusion models) или диффузионные модели являются одним из наиболее влиятельных и действенных методов генеративного моделирования для генерации изображений, которые были предложены за последнее десятилетие. Во многих тестах диффузионные модели теперь превосходят наиболее совершенные GAN и быстро становятся предпочтительным выбором для практиков генеративного моделирования, особенно для визуализации (например, DALL.E 2 от компании OpenAI и ImageGen от компании Google для генерации изображений по тексту). В последнее время произошел взрывной рост числа диффузионных моделей, применяемых для решения широкого круга задач, что напоминает распространение GAN, которое имело место в период 2017–2020 годов.  
  
Многие из основных идей, лежащих в основе диффузионных моделей, имеют сходство с более ранними типами генеративных моделей, которые мы уже рассматривали (например, автоэнкодеры с шумоподавлением, энергетические модели). Действительно, название «диффузия» вдохновлено хорошо изученным свойством термодинамической диффузии – была установлена важная связь между этой чисто физической областью и глубоким обучением.  
  
Основная идея диффузионных моделей основана на наблюдении, что шум трудно преобразовать в структурированные данные, но достаточно легко преобразовать структурированные данные в шум. В частности, мы можем использовать прямой (forwards) диффузионный процесс для постепенного преобразования наблюдаемых данных 𝐱0  
 в зашумленную версию 𝐱𝑇  
 , пропуская данные через 𝑇  
 шагов стохастического кодировщика 𝑞(𝐱𝑡∣𝐱𝑡−1)  
 . После достаточного количества шагов мы (приближенно) получаем стандартное многомерное нормальное распределение 𝐱𝑇∼(0,𝐈)  
 или какое-либо другое удобное известное распределение. Затем мы изучаем обратный процесс, чтобы отменить выполненные кодировщиком шаги, пропуская шум через 𝑇  
 шагов декодера 𝑝𝜃(𝐱𝑡−1∣𝐱𝑡)  
 , пока мы не сгенерируем первоначальные данные 𝐱0.

**Цель работы**:  
Систематизировать математические основы, архитектурные решения и прикладные аспекты диффузионных моделей.

# 2 Теоретические основы

## 2.1 Определение

Диффузионные модели — это генеративные модели, обучающиеся через последовательное зашумление и восстановление данных по формуле:  
  
$$ q(x\_t|x\_{t-1}) = \mathcal{N}(x\_t; \sqrt{1-\beta\_t}x\_{t-1}, \beta\_t\mathbf{I}) $$  
  
## Связь с другими методами  
| Метод | Отличия |  
|----------------|----------------------------------|  
| VAE | Используют энкодер-декодер |  
| GAN | Опираются на состязательность |  
| Score-Based | Общие теоретические корни |  
  
# Принцип работы  
  
## Прямой процесс  
1. Итеративное добавление гауссова шума  
2. Марковская цепь с параметрами $\beta\_t$  
  
## Обратный процесс  
$$ p\_\theta(x\_{t-1}|x\_t) = \mathcal{N}(\mu\_\theta(x\_t,t), \Sigma\_\theta(x\_t,t)) $$  
  
# Применение

# 3 Пример генерации в PyTorch

from diffusers import StableDiffusionPipeline pipe = StableDiffusionPipeline.from\_pretrained(“stabilityai/stable-diffusion-2”) image = pipe(“кошка в шляпе”).images[0] ```

# 4 Заключение

Диффузионные модели демонстрируют: 1. Более стабильное обучение vs GAN 2. Высокое качество генерации 3. Широкие перспективы в мультимодальных приложениях