2025 年 12 月 8 日至 12 月 14 日周报

何瑞杰
中山大学, 大湾区大学

目录

速 览

论文方面，本周阅读了 DDPM 后的经典工作 DDIM 的理论部分，DDIM 将 DDPM 中的 Markov 模型扩展为增广的 Markov 过程或是 $\boldsymbol{x}_0$ 条件下的一族 Markov 过程，并利用这一改变带来的性质同时达到兼容 DDPM 的训练成果和加速生成的目的。

项目方面，本周对先前搭建的简易规则推断进行了测试，在测试的 6 个规则中有五个规则预测正确。

其他自主学习方面，阅读了《非线性动力学与混沌》的第三章的部分内容，观看学习了《测度论》简易教程，了解了 σ-代数、测度、测度空间、Lebesgue 测度的定义。

1. 项目进展

1.1. 使用神经网络学习生命游戏的演化动力学

本周对先前搭建的简易规则推断进行了测试, 在测试的 6 个规则中有五个规则预测正确。但由于目前的代码位于另一台机器中，无法获取，将在下周完善可视化的工作。

本周的规划工作重心为向 Overleaf 在线文章中同步至今的所有结果。

2. 文献阅读

2.1. Demoising Diffusion Implicit Models [1]

ICLR 2021 | Jiaming Song et al.

在 DDPM 论文中，加噪过程为一个 Markov 过程，有状态转移分布 $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ 和对应的反向转移分布 $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$。这代表加噪过程中进行的步数和模型参与的去噪步骤数是相同的。另外，在假设状态转移分布为 Gauss 分布时，到达加噪过程终点时的分布 $\boldsymbol{x}_N$ 的分布理论上趋于标准正态分布 $N(0,I)$，当 N 较小时，$\boldsymbol{x}_N$ 的分布与标准正态分布不再近似，而去噪过程又是从标准正态分布 $N(0,I)$ 开始的，这就造成分布不对应。

本文为解决这一问题引入了成为隐式去噪扩散模型（denoising diffusion implicit models）。对于第一个问题，DDIM 引入了一个增广 Markov 加噪和去噪过程，使得生成过程中可以跳步，同时也一定程度缓解了第二个问题：在加噪过程中可以引入任意大的加噪步数，使得 $\boldsymbol{x}_N$ 的分布趋于标准正态分布。除此之外，DDIM 还将噪声项趋于零，这表示加噪和去噪动力学从 SDE 变成了 ODE，从而获取更高的求解速度和更优的数值性能。

2.1.1. 增广 Markov 前向过程

DDIM 引入的去噪过程的转移分布为 $p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)$，并令其为一个简单的 Gauss 分布 $q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)=N(\boldsymbol{\mu}(\boldsymbol{x}_t,\boldsymbol{x}_0),\sigma_t^2\boldsymbol{I})$。由重参数化技巧，$\boldsymbol{x}_{t-1}$ 可以写为

$$\begin{aligned}\boldsymbol{x}_{t-1} &\sim \boldsymbol{\mu}(\boldsymbol{x}_t,\boldsymbol{x}_0)+\sigma_t^2\boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon}\sim N(\boldsymbol{0},\boldsymbol{I})\\ &\sim a\boldsymbol{x}_t+b\boldsymbol{x}_0+\sigma_t^2\boldsymbol{\varepsilon}\end{aligned} \tag{1}$$

其中 a 和 b 是实数。DDIM 引入的来自 DDPM 的假设为 $\boldsymbol{x}_t\sim N\big(\sqrt{\alpha_t}\boldsymbol{x}_0,(1-\alpha_t)\boldsymbol{I}\big)$，因此用重参数化的语言来写，就是

$$\begin{gathered}\sqrt{\alpha_{t-1}}\boldsymbol{x}_0+\sqrt{1-\alpha_{t-1}}\varepsilon\sim a\Big(\sqrt{\alpha_t}\boldsymbol{x}_0+\sqrt{1-\alpha_t}\varepsilon'\Big)+b\boldsymbol{x}_0+\sigma_t\varepsilon''\\ N(\sqrt{\alpha_{t-1}}\boldsymbol{x}_0,(1-\alpha_{t-1})\boldsymbol{I})=N(a\sqrt{\alpha_t}\boldsymbol{x}_0+b\boldsymbol{x}_0,(a^2(1-\alpha_t)+\sigma_t^2)\boldsymbol{I})\end{gathered} \tag{2}$$

分布参数对应相等，就得到

$$\begin{aligned}a&=\sqrt{\frac{1-\alpha_{t-1}-\sigma_t^2}{1-\alpha_t}}\\ b&=\sqrt{\alpha_{t-1}}-\sqrt{\alpha_t}\cdot\sqrt{\frac{1-\alpha_{t-1}-\sigma_t^2}{1-\alpha_t}}\end{aligned} \tag{3}$$

这样就得到逆向过程转移概率应该有的形式：

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)=N\left(\sqrt{\alpha_{t-1}}\boldsymbol{x}_0+\sqrt{1-\alpha_{t-1}-\sigma_t^2}\cdot\frac{\boldsymbol{x}_t-\sqrt{\alpha_t}\boldsymbol{x}_0}{\sqrt{1-\alpha_t}},\sigma_t^2\boldsymbol{I}\right) \tag{4}$$

对应的加噪过程可以由 Bayes 公式给出

$$q_\sigma(\boldsymbol{x}_t\mid\boldsymbol{x}_{t-1},\boldsymbol{x}_0)=\frac{q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)q_\sigma(\boldsymbol{x}_t|\boldsymbol{x}_0)}{q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)} \tag{5}$$

然而事实上，由于上面推导中的 $t-1$ 时间步和 t 时间步并没有联系，我们可以将两个时间步视作 s,t，满足 $0<s<t$，这样一来跳步的去噪过程可以写成

$$q(\boldsymbol{x}_s|\boldsymbol{x}_t,\boldsymbol{x}_0) = N\left(\sqrt{\alpha_s}\boldsymbol{x}_0 + \sqrt{1-\alpha_s-\sigma_{s,t}^2}\cdot\frac{\boldsymbol{x}_t-\sqrt{\alpha_t}\boldsymbol{x}_0}{\sqrt{1-\alpha_t}}, \sigma_{s,t}^2\boldsymbol{I}\right) \tag{6}$$

此时 $\sigma_{s,t}$ 需要做相应的调整。或者将 t 和 $t-1$ 的下标视作序列 $\{t\}_{t=0}^N$ 的一个子列 $\{t_k\}_{k=1}^N$，相应地方差的下标就变成 σ_{t_k}。

然而需要注意的是，对 $q_\sigma(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)$ 的拆解依然使用了 Markov 性:

$$q_\sigma(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0) = \prod_{k=1}^T q_\sigma(\boldsymbol{x}_k|\boldsymbol{x}_{1:k-1},\boldsymbol{x}_0) = \prod_{k=1}^T q_\sigma(\boldsymbol{x}_k|\boldsymbol{x}_{k-1},\boldsymbol{x}_0) \tag{7}$$

因此与其说 DDIM 使用了非 Markov 链的加噪/去噪模型，不如说它使用的是一个相比原版**增广 (augmented) 的 Markov 链**，其中状态 $\hat{x}_t$ 可理解为 $(\boldsymbol{x}_t,\boldsymbol{x}_0)$。

2.1.2. 兼容 DDPM

注意 DDPM 中的目标可以写成

$$L_\gamma(\varepsilon_\theta) = \sum_{t=1}^T \gamma_t\mathbb{E}_*\left\|\varepsilon_\theta^{(t)}\left(\sqrt{\alpha_t}\boldsymbol{x}_0+\sqrt{1-\alpha_t}\varepsilon_t\right)-\varepsilon_t\right\|_2^2 \tag{8}$$

其目标所对齐或模拟的是 $q_\sigma(\boldsymbol{x}_0|\boldsymbol{x}_t)$。我们假设通过 DDPM 训练目标训练好的模型 ε_θ^t 给出的是 $\boldsymbol{x}_0$ 的一个估计。就可以利用该估计计算逆向转移概率。令

$$f_\theta^{(t)}(\boldsymbol{x}_t) = \frac{\boldsymbol{x}_t-\sqrt{1-\alpha_t}\varepsilon_\theta^{(t)}(\boldsymbol{x}_t)}{\sqrt{\alpha_t}} \tag{9}$$

则逆向转移概率可以写成 $p_{\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)} = q_\sigma\left(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, f_\theta^{(t)}(\boldsymbol{x}_t)\right)$。自然地，我们需要对齐模型估计的去噪序列和真实的序列:

$$J_\sigma(\varepsilon_\theta) = \mathbb{E}_{\boldsymbol{x}_{0:T}\sim q_\sigma(\boldsymbol{x}_{0:T})}[\log q_\sigma(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0) - \log p_\theta(x_{0:T})] \tag{10}$$

可以证明，该目标与 DDPM 的目标等价。

在生成时，给定 $\boldsymbol{x}_t$，可以根据下式预测 $\boldsymbol{x}_s$

$$\boldsymbol{x}_s = \underbrace{\sqrt{\alpha_s}\left[\frac{x_t-\sqrt{1-\alpha_t}\sigma_\theta^{(t)}(\boldsymbol{x}_t)}{\sqrt{\alpha_t}}\right]}_{x_0\text{ 的预测值}} + \underbrace{\sqrt{1-\alpha_s-\sigma_{s,t}^2}\varepsilon_\theta^{(t)}(\boldsymbol{x}_t)}_{\text{指向 } x_{t-1} \text{ 的调节项}} + \sigma_{s,t}\varepsilon_{s,t} \tag{11}$$

如果对任意 s,t，都令 $\sigma_{s,t}\to 0$，则去噪动力学的随机性逐渐消失，该离散动力学变成了下面的 ODE 的 Euler 离散化版本

$$\mathrm{d}\bar{\boldsymbol{x}} = \varepsilon_\theta^{(t)}\left[\frac{\bar{\boldsymbol{x}}}{\sqrt{\sigma^2+1}}\right]\mathrm{d}\sigma, \tag{12}$$

这还对应着 Song 等人 [2]论文中的 VE-SDE 的概率流 ODE 的形式。由于其没有随机性，可以应用更快速的 ODE 数值解法。

为更好理解论文的理论，本文需要参照 DDPM 和 DDIM 的代码。

3. 下周计划

论文阅读

1. 生成模型
   - DiT
2. 机器学习理论
   - Reconciling modern machine learning practice and the bias-variance trade-off

项目进度

1. 使用神经网络学习生命游戏的演化动力学
   - 更新在线 Overleaf 文档

理论学习

1. 随机过程课程
   - 开始总复习
2. 随机微分方程
   - 第五章完成
3. 非线性动力学与混沌
   - 第三章、第四章
4. Stein 实分析
   - 第一章
5. Sakurai 现代量子力学
   - 第一章

参考文献

[1] J. Song, C. Meng, and S. Ermon, "Denoising Diffusion Implicit Models," *CoRR*, 2020, [Online]. Available: https://arxiv.org/abs/2010.02502

[2] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-Based Generative Modeling through Stochastic Differential Equations," *CoRR*, 2020, [Online]. Available: https://arxiv.org/abs/2011.13456