生成扩散模型漫谈（十七）：构建ODE的一般步骤（下）

23 Feb By 苏剑林 | 2023-02-23 | 76464位读者 引用

历史总是惊人地相似。当初笔者在写《生成扩散模型漫谈（十四）：构建ODE的一般步骤（上）》（当时还没有“上”这个后缀）时，以为自己已经搞清楚了构建ODE式扩散的一般步骤，结果读者 @gaohuazuo 就给出了一个新的直观有效的方案，这直接导致了后续《生成扩散模型漫谈（十四）：构建ODE的一般步骤（中）》（当时后缀是“下”）。而当笔者以为事情已经终结时，却发现ICLR2023的论文《Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow》又给出了一个构建ODE式扩散模型的新方案，其简洁、直观的程度简直前所未有，令人拍案叫绝。所以笔者只好默默将前一篇的后缀改为“中”，然后写了这个“下”篇来分享这一新的结果。

直观结果

我们知道，扩散模型是一个$\boldsymbol{x}_T \to \boldsymbol{x}_0$的演化过程，而ODE式扩散模型则指定演化过程按照如下ODE进行：

$$\frac{d\boldsymbol{x}_t}{dt} = \boldsymbol{f}_t(\boldsymbol{x}_t) \tag{1}$$

而所谓构建ODE式扩散模型，就是要设计一个函数$\boldsymbol{f}_t(\boldsymbol{x}_t)$，使其对应的演化轨迹构成给定分布$p_T(\boldsymbol{x}_T)$、$p_0(\boldsymbol{x}_0)$之间的一个变换。说白了，我们希望从$p_T(\boldsymbol{x}_T)$中随机采样一个$\boldsymbol{x}_T$，然后按照上述ODE向后演化得到的$\boldsymbol{x}_0$是$\sim p_0(\boldsymbol{x}_0)$的。

原论文的思路非常简单，随机选定$\boldsymbol{x}_0 \sim p_0(\boldsymbol{x}_0), \boldsymbol{x}_T \sim p_T(\boldsymbol{x}_T)$，假设它们按照轨迹

$$\boldsymbol{x}_t = \boldsymbol{\varphi}_t(\boldsymbol{x}_0, \boldsymbol{x}_T) \tag{2}$$

进行变换。这个轨迹是一个已知的函数，是我们自行设计的部分，理论上只要满足

$$\boldsymbol{x}_0 = \boldsymbol{\varphi}_0(\boldsymbol{x}_0, \boldsymbol{x}_T), \quad \boldsymbol{x}_T = \boldsymbol{\varphi}_T(\boldsymbol{x}_0, \boldsymbol{x}_T) \tag{3}$$

的连续函数都可以。接着我们就可以写出它满足的微分方程：

$$\frac{d\boldsymbol{x}_t}{dt} = \frac{\partial \boldsymbol{\varphi}_t(\boldsymbol{x}_0, \boldsymbol{x}_T)}{\partial t} \tag{4}$$

但这个微分方程是不实用的，因为我们想要的是给定$\boldsymbol{x}_T$来生成$\boldsymbol{x}_0$，但它右端却是$\boldsymbol{x}_0$的函数（如果已知$\boldsymbol{x}_0$就完事了），只有像式(1)那样右端只含有$\boldsymbol{x}_t$的ODE（单从因果关系来看，理论上也可以包含$\boldsymbol{x}_T$，但我们一般不考虑这种情况）才能进行实用的演化。那么，一个直观又“异想天开”的想法是：**学一个函数$\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$尽量逼近上式右端！**为此，我们优化如下目标：

$$\mathbb{E}_{\boldsymbol{x}_0\sim p_0(\boldsymbol{x}_0),\boldsymbol{x}_T\sim p_T(\boldsymbol{x}_T)}\left[\left\Vert \boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \frac{\partial \boldsymbol{\varphi}_t(\boldsymbol{x}_0, \boldsymbol{x}_T)}{\partial t}\right\Vert^2\right] \tag{5}$$

由于$\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$尽量逼近了$\frac{\partial \boldsymbol{\varphi}_t(\boldsymbol{x}_0, \boldsymbol{x}_T)}{\partial t}$，所以我们认为将方程(4)的右端替换为$\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$也是成立的，这就得到实用的扩散ODE：

$$\frac{d\boldsymbol{x}_t}{dt} = \boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) \tag{6}$$

简单例子

作为简单的例子，我们设$T=1$，并设变化轨迹是直线

$$\boldsymbol{x}_t = \boldsymbol{\varphi}_t(\boldsymbol{x}_0, \boldsymbol{x}_1) = (\boldsymbol{x}_1 - \boldsymbol{x}_0)t + \boldsymbol{x}_0 \tag{7}$$

那么

$$\frac{\partial \boldsymbol{\varphi}_t(\boldsymbol{x}_0, \boldsymbol{x}_T)}{\partial t} = \boldsymbol{x}_1 - \boldsymbol{x}_0 \tag{8}$$

所以训练目标(5)就是：

$$\mathbb{E}_{\boldsymbol{x}_0\sim p_0(\boldsymbol{x}_0),\boldsymbol{x}_T\sim p_T(\boldsymbol{x}_T)}\left[\left\Vert \boldsymbol{v}_{\boldsymbol{\theta}}\big((\boldsymbol{x}_1 - \boldsymbol{x}_0)t + \boldsymbol{x}_0, t\big) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\right\Vert^2\right] \tag{9}$$

或者等价地写成

$$\mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_t\sim p_0(\boldsymbol{x}_0)p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[\left\Vert\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \frac{\boldsymbol{x}_t - \boldsymbol{x}_0}{t}\right\Vert^2\right] \tag{10}$$

这就完事了！结果跟《生成扩散模型漫谈（十四）：构建ODE的一般步骤（中）》的“直线轨迹”例子是完全一致的，也是原论文主要研究的模型，被称为“Rectified Flow”。

从这个直线例子的过程也可以看出，通过该思路来构建扩散ODE的步骤只有寥寥几行，相比之前的过程是大大简化了，简单到甚至让人有种“颠覆了对扩散模型的印象”的不可思议之感。

证明过程

然而，迄今为止前面“直观结果”一节的结论只能算是一个直观的猜测，因为我们还没有从理论上证明优化目标(5)所得到的方程(6)的确实现了分布$p_T(\boldsymbol{x}_T)$、$p_0(\boldsymbol{x}_0)$之间的变换。

为了证明这一结论，笔者一开始是想证明目标(5)的最优解满足连续性方程：

$$\frac{\partial p_t(\boldsymbol{x}_t)}{\partial t} = -\nabla_{\boldsymbol{x}_t}\cdot\big(p_t(\boldsymbol{x}_t)\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\big) \tag{11}$$

如果满足，那么根据连续性方程与ODE的对应关系（参考《生成扩散模型漫谈（十二）：“硬刚”扩散ODE》、《测试函数法推导连续性方程和Fokker-Planck方程》），方程(6)确实是分布$p_T(\boldsymbol{x}_T)$、$p_0(\boldsymbol{x}_0)$之间的一个变换。

但仔细想一下，这个思路似乎有点迂回了，因为根据文章《测试函数法推导连续性方程和Fokker-Planck方程》，连续性方程本身就是由ODE通过

$$\mathbb{E}_{\boldsymbol{x}_{t+\Delta t}}\left[\phi(\boldsymbol{x}_{t+\Delta t})\right] = \mathbb{E}_{\boldsymbol{x}_t}\left[\phi(\boldsymbol{x}_t + \boldsymbol{f}_t(\boldsymbol{x}_t)\Delta t)\right] \tag{12}$$

推出的，所以按理说(12)更基本，我们只需要证明(5)的最优解满足它就行。也就是说，我们想要找到一个纯粹是$\boldsymbol{x}_t$的函数$\boldsymbol{f}_t(\boldsymbol{x}_t)$满足(12)，然后发现它正好是(5)的最优解。

于是，我们写出（简单起见，$\boldsymbol{\varphi}_t(\boldsymbol{x}_0, \boldsymbol{x}_T)$简写为$\boldsymbol{\varphi}_t$）

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_{t+\Delta t}}\left[\phi(\boldsymbol{x}_{t+\Delta t})\right] =&\, \mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_T}\left[\phi(\boldsymbol{\varphi}_{t+\Delta t})\right] \\
=&\, \mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_T}\left[\phi(\boldsymbol{\varphi}_t) + \Delta t\,\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\cdot\nabla_{\boldsymbol{\varphi}_t}\phi(\boldsymbol{\varphi}_t)\right] \\
=&\, \mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_T}\left[\phi(\boldsymbol{x}_t)\right] + \Delta t\,\mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_T}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\cdot\nabla_{\boldsymbol{x}_t}\phi(\boldsymbol{x}_t)\right] \\
=&\, \mathbb{E}_{\boldsymbol{x}_t}\left[\phi(\boldsymbol{x}_t)\right] + \Delta t\,\mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_T}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\cdot\nabla_{\boldsymbol{x}_t}\phi(\boldsymbol{x}_t)\right]
\end{aligned}\tag{13}$$

其中第一个等号是因为式(2)，第二个等号是泰勒展开到一阶，第三个等号同样是式(2)，第四个等号就是因为$\boldsymbol{x}_t$是$\boldsymbol{x}_0, \boldsymbol{x}_T$的确定性函数，所以关于$\boldsymbol{x}_0, \boldsymbol{x}_T$的期望就是关于$\boldsymbol{x}_t$的期望。

我们看到，$\frac{\partial \boldsymbol{\varphi}_t}{\partial t}$是$\boldsymbol{x}_0, \boldsymbol{x}_T$的函数，接下来我们再做一个假设：**式(2)关于$\boldsymbol{x}_T$是可逆的。** 这个假设意味着我们可以从式(2)中解出$\boldsymbol{x}_T = \boldsymbol{\psi}_t(\boldsymbol{x}_0, \boldsymbol{x}_t)$，这个结果可以代入$\frac{\partial \boldsymbol{\varphi}_t}{\partial t}$，使它变为$\boldsymbol{x}_0, \boldsymbol{x}_t$的函数。所以我们有

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_{t+\Delta t}}\left[\phi(\boldsymbol{x}_{t+\Delta t})\right] =&\, \mathbb{E}_{\boldsymbol{x}_t}\left[\phi(\boldsymbol{x}_t)\right] + \Delta t\,\mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_T}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\cdot\nabla_{\boldsymbol{x}_t}\phi(\boldsymbol{x}_t)\right] \\
=&\, \mathbb{E}_{\boldsymbol{x}_t}\left[\phi(\boldsymbol{x}_t)\right] + \Delta t\,\mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_t}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\cdot\nabla_{\boldsymbol{x}_t}\phi(\boldsymbol{x}_t)\right] \\
=&\, \mathbb{E}_{\boldsymbol{x}_t}\left[\phi(\boldsymbol{x}_t)\right] + \Delta t\,\mathbb{E}_{\boldsymbol{x}_t}\left[\underbrace{\mathbb{E}_{\boldsymbol{x}_0|\boldsymbol{x}_t}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\right]}_{\boldsymbol{x}_t\text{的函数}}\cdot\nabla_{\boldsymbol{x}_t}\phi(\boldsymbol{x}_t)\right] \\
=&\, \mathbb{E}_{\boldsymbol{x}_t}\left[\phi\left(\boldsymbol{x}_t + \Delta t\,\mathbb{E}_{\boldsymbol{x}_0|\boldsymbol{x}_t}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\right]\right)\right]
\end{aligned}\tag{14}$$

其中第二个等号是因为$\frac{\partial \boldsymbol{\varphi}_t}{\partial t}$已经改为$\boldsymbol{x}_0, \boldsymbol{x}_t$的函数，所以第二项期望的随机变量改为$\boldsymbol{x}_0, \boldsymbol{x}_t$；第三个等号则是相当于做了分解$p(\boldsymbol{x}_0, \boldsymbol{x}_t) = p(\boldsymbol{x}_0|\boldsymbol{x}_t)p(\boldsymbol{x}_t)$，此时$\boldsymbol{x}_0, \boldsymbol{x}_t$不是独立的，所以要注明$\boldsymbol{x}_0|\boldsymbol{x}_t$，即$\boldsymbol{x}_0$是依赖于$\boldsymbol{x}_t$的。注意$\frac{\partial \boldsymbol{\varphi}_t}{\partial t}$原本是$\boldsymbol{x}_0, \boldsymbol{x}_t$的函数，现在对$\boldsymbol{x}_0$求期望后，剩下的唯一自变量就是$\boldsymbol{x}_t$，后面我们会看到它就是我们要找的纯粹是$\boldsymbol{x}_t$的函数！第四个等号，就是利用泰勒展开公式将两项重新合并起来。

现在，我们得到了

$$\mathbb{E}_{\boldsymbol{x}_{t+\Delta t}}\left[\phi(\boldsymbol{x}_{t+\Delta t})\right] = \mathbb{E}_{\boldsymbol{x}_t}\left[\phi\left(\boldsymbol{x}_t + \Delta t\,\mathbb{E}_{\boldsymbol{x}_0|\boldsymbol{x}_t}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\right]\right)\right] \tag{15}$$

对于任意测试函数ϕ成立，所以这意味着

$$\boldsymbol{x}_{t+\Delta t} = \boldsymbol{x}_t + \Delta t\,\mathbb{E}_{\boldsymbol{x}_0|\boldsymbol{x}_t}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\right] \quad\Rightarrow\quad \frac{d\boldsymbol{x}_t}{dt} = \mathbb{E}_{\boldsymbol{x}_0|\boldsymbol{x}_t}\left[\frac{\partial \boldsymbol{\varphi}_t}{\partial t}\right] \tag{16}$$

就是我们要寻找的ODE。根据

$$\mathbb{E}_{\boldsymbol{x}}[\boldsymbol{x}] = \mathop{\text{argmin}}_{\boldsymbol{\mu}} \mathbb{E}_{\boldsymbol{x}}\left[\Vert\boldsymbol{x} - \boldsymbol{\mu}\Vert^2\right] \tag{17}$$

式(16)的右端正好是训练目标(5)的最优解，这就证明了优化训练目标(5)得出的方程(6)的确实现了分布$p_T(\boldsymbol{x}_T)$、$p_0(\boldsymbol{x}_0)$之间的变换。

读后感受

关于“直观结果”中的构建扩散ODE的思路，原论文的作者还写了篇知乎专栏文章《[ICLR2023] 扩散生成模型新方法：极度简化，一步生成》，大家也可以去读读。读者也是在这篇专栏中首次了解到该方法的，并深深为之震惊和叹服。

如果读者读过《生成扩散模型漫谈（十四）：构建ODE的一般步骤（中）》，那么就会更加体会到该思路的简单直接，也更能理解笔者为何如此不吝赞美之词。不怕大家笑话，笔者在写“中篇”（当时的“下篇”）的时候，是考虑过式(2)所描述的轨迹的，但是在当时的框架下，根本没法推演下去，最后以失败告终，当时完全想不到它能以一种如此简捷的方式进行下去。所以，写这个扩散ODE系列真的让人有种“人比人，气死人”的感觉，“中篇”、“下篇”就是自己智商被一次次“降维打击”的最好见证。

读者可能想问，还会不会有更简单的第四篇，让笔者再一次经历降维打击？可能有，但概率真的很小了，真的很难想象会有比这更简单的构建步骤了。“直观结果”一节看上去很长，但实际步骤就只有两步：1、随便选择一个渐变轨迹；2、用$\boldsymbol{x}_t$的函数去逼近渐变轨迹对t的导数。就这样的寥寥两步，还能怎么再简化呢？甚至说，“证明过程”一

节的推导也是相当简单的了，虽然写得长，但本质就是求个导，然后变换一下求期望的分布，比前两篇的过程简单了可不止一丁半点。总而言之，亲自完成过ODE扩散的前两篇推导的读者就能深刻感觉到，这一篇的思路是真的简单，简单到让我们觉得已经无法再简单了。

此外，除了提供构建扩散ODE的简单思路外，原论文还讨论了Rectified Flow跟最优传输之间的联系，以及如何用这种联系来加速采样过程，等等。但这部分内容并不是本文主要关心的，所以等以后有机会我们再讨论它们。

文章小结

本文介绍了Rectified Flow一文中提出的构建ODE式扩散模型的一种极其简单直观的思路，并给出了自己的证明过程。

如果您需要引用本文，请参考：

苏剑林. (Feb. 23, 2023). 《生成扩散模型漫谈（十七）：构建ODE的一般步骤（下）》[Blog post]. Retrieved from https://spaces.ac.cn/archives/9497

```
@online{kexuefm-9497,
    title={生成扩散模型漫谈（十七）：构建ODE的一般步骤（下）},
    author={苏剑林},
    year={2023},
    month={Feb},
    url={\url{https://spaces.ac.cn/archives/9497}},
}
```