生成式AI与随机微分方程

An Introduction to Flow and Diffusion Models

笔记整理：Gemini

MIT IAP 2025 | Jan 21, 2025

2025 年 7 月 25 日

目录

1 核心思想："生成"就是"采样"

1.1 什么是生成式AI？

生成式AI (Generative AI) 是能够创造全新内容的人工智能系统。我们日常接触的产物包括：

- **文本生成** (例如：ChatGPT)
- **艺术图像** (例如：Midjourney, Stable Diffusion)
- **逼真视频** (例如：Sora)

图 1: 生成式AI可以创造文本、艺术图像和视频等新内容 (源自幻灯片第3页)。

1.2 如何用数学语言描述"生成"？

为了让计算机理解"生成"这一概念，我们需要将其过程形式化。

1.2.1 第一步：万物皆向量

我们想要生成的任何对象，无论是图片、视频还是蛋白质分子，都可以被表示为一个高维度的数字向量 $z \in \mathbb{R}^d$。

1.2.2 第二步：引入"数据分布" (p_{data})

在所有可能的向量中，只有极少数对应着"有意义"的内容。我们可以想象存在一个**数据分布** p_{data}，它为"好的"、"真实的"向量赋予了高概率。

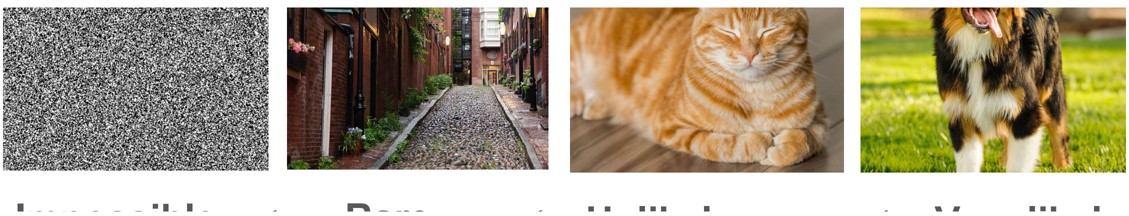

图 2: 从无意义的噪声到有意义的图像，其在数据分布中的概率逐渐升高 (源自幻灯片第10页)。

1.2.3 第三步：生成 = 从数据分布中采样

有了数据分布的概念后，“生成一个对象”这个任务就可以被精确地定义为从数据分布 p_{data} 中采样一个向量 z。

$$z \sim p_{\text{data}}$$

这是一个核心的观念转变：**创造性任务被转化为了一个概率采样问题**。

1.3 生成模型的终极目标

我们并不知道 p_{data} 的确切形式，但我们可以通过一个庞大的数据集（例如，互联网上的所有图片）来近似它。生成模型的终极目标是训练一个模型，它能将一个来自简单分布（如高斯噪声）的输入，转化为一个看起来像是从真实数据分布中采样出来的高质量输出。

图 3: 生成模型将左侧的随机噪声图转换为右侧高质量的真实图片 (源自幻灯片第14页)。

2 实现路径：流模型与扩散模型

我们如何构建一个能将简单噪声分布 p_{init} 变换为复杂数据分布 p_{data} 的模型呢？答案是利用**微分方程**来定义一个从噪声到数据的连续变换路径。

2.1 流模型 (Flow Models) 与常微分方程 (ODE)

流模型构建的是一个**确定性**的变换过程。

2.1.1 核心直觉：向量场与流

想象一个平稳流动的河流，河中任何一个点的水流速度和方向都是确定的。这个描述水流的“速度场”就是一个**向量场 (Vector Field)** $u_t(x)$，它为空间中的每个点 x 和每个时刻 t 指定了一个速度向量。

一个从某点 x_0 出发的粒子，会沿着向量场指定的方向和速度运动，其形成的轨迹被称为一条**流 (Flow)**。在生成模型中，我们的目标就是学习一个向量场，它能引导来自初始噪声分布的点，“流向”目标数据分布。

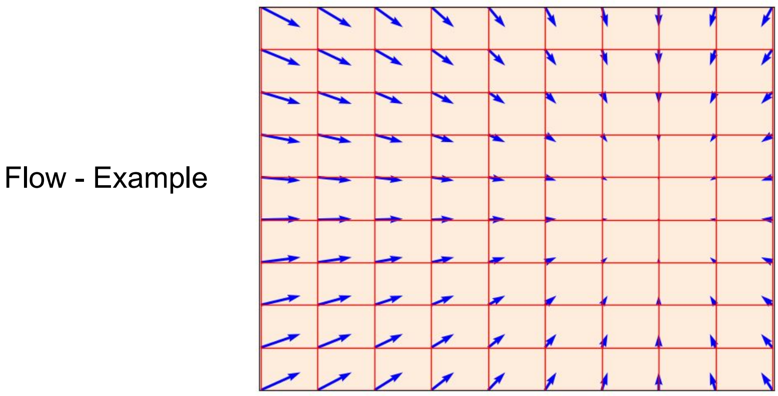

图 4: 一个向量场的可视化，箭头表示粒子在该点的运动方向 (源自幻灯片第17页)。

2.1.2 数学形式化：常微分方程 (ODE)

粒子在向量场中的运动轨迹，可以用一个**常微分方程 (Ordinary Differential Equation, ODE)** 来精确描述。

给定一个初始点 $X_0 = x_0$，其在向量场 u_t 中的运动轨迹 X_t 满足：

$$\frac{d}{dt}X_t = u_t(X_t)$$

这个方程的含义是：粒子在任意时刻 t 的瞬时速度 $\frac{d}{dt}X_t$，等于它所在位置 X_t 处的向量场值 $u_t(X_t)$。

定理 2.1 (Picard-Lindelöf, 解的存在性与唯一性). *如果向量场 $u_t(x)$ 是连续可微且其导数有界（或更一般地，满足Lipschitz条件），那么对于任意给定的初始点 x_0，上述ODE存在唯一的解。*

重要启示：这个定理为我们提供了理论保障。只要我们学习到的神经网络向量场足够“平滑”，那么从任意一个初始噪声点出发，都存在一条唯一的、确定的路径通往最终的生成样本。

2.1.3 详细推导：线性ODE的解析解

让我们通过一个简单的例子来理解ODE的求解。考虑一个线性向量场：$u_t(x) = -\theta x$ (其中 $\theta > 0$ 为常数)。其对应的ODE为：

$$\frac{d}{dt}X_t = -\theta X_t$$

求解过程推导： 这是一个可分离变量的一阶线性微分方程。我们将变量分离：

$$\frac{dX_t}{X_t} = -\theta dt$$

对两边从初始时刻 0 到任意时刻 t 进行积分：

$$\int_{X_0}^{X_t} \frac{1}{x} dx = \int_0^t -\theta d\tau$$

计算积分：

$$\ln|x|\Big|_{X_0}^{X_t} = -\theta\tau\Big|_0^t$$
$$\ln|X_t| - \ln|X_0| = -\theta t$$
$$\ln\left|\frac{X_t}{X_0}\right| = -\theta t$$

取指数，并考虑到 X_t 和 X_0 符号相同：

$$\frac{X_t}{X_0} = e^{-\theta t}$$

最终得到解析解，即流的映射 $\psi_t(x_0)$:

$$X_t = \psi_t(X_0) = X_0 e^{-\theta t}$$

这个解表明，任何初始点 X_0 都会沿着指数衰减的路径趋向于原点0。

2.1.4 数值模拟：欧拉法

在复杂的模型中，ODE通常没有解析解，我们需要使用数值方法近似求解。最简单的方法是**欧拉法 (Euler Method)**。其思想是将连续的路径离散化为一系列小步长。

Algorithm 1 使用欧拉法模拟ODE

需要: 向量场 u_t, 初始条件 x_0, 时间区间 $[0, 1]$, 步数 N
$h \leftarrow 1/N$ ▷ 计算步长
$X \leftarrow x_0$ ▷ 初始化位置
$t \leftarrow 0$
for $i = 0$ to $N - 1$ **do**
    $X \leftarrow X + h \cdot u_t(X)$ ▷ 沿着向量场方向前进一小步
    $t \leftarrow t + h$
end for
return X ▷ 返回在 t=1 时的最终位置

2.2 扩散模型 (Diffusion Models) 与随机微分方程 (SDE)

扩散模型在流模型的基础上，为变换过程增加了**随机性**，使得路径更加灵活。

2.2.1 核心直觉：布朗运动

想象一下悬浮在水中的花粉颗粒，它的运动轨迹是完全无规则、随机的。这种现象被称为**布朗运动 (Brownian Motion)**，它是由大量水分子对花粉颗粒不均衡的、随机的碰撞引起的。在数学上，我们用维纳过程 (Wiener Process) W_t 来描述它。

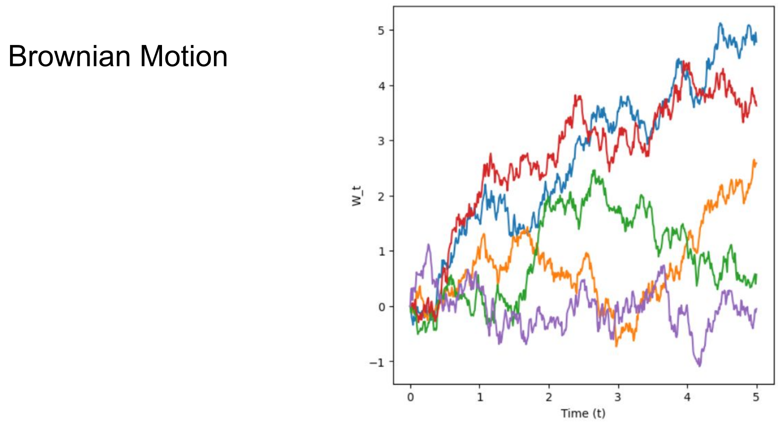

图 5: 布朗运动的随机路径示例 (源自幻灯片第24页)。

2.2.2 数学形式化：随机微分方程 (SDE)

扩散模型的路径由一个**随机微分方程 (Stochastic Differential Equation, SDE)** 描述。

$$dX_t = \underbrace{u_t(X_t)dt}_{\text{漂移项 (Drift)}} + \underbrace{\sigma_t dW_t}_{\text{扩散项 (Diffusion)}}$$

- **漂移项 (Drift)**: $u_t(X_t)dt$ 与ODE中的向量场作用相同，是确定性的部分，指明了主要的运动方向。
- **扩散项 (Diffusion)**: $\sigma_t dW_t$ 是随机的部分，引入了一个强度为 σ_t 的布朗运动，为路径增加随机扰动。

定理 2.2 (SDE解的存在性与唯一性). *与ODE类似，在向量场 u_t 和扩散系数 σ_t 满足一定平滑性条件时，SDE同样存在唯一的解。*

2.2.3 经典示例：Ornstein-Uhlenbeck 过程

这是线性ODE的随机版本，其SDE形式为：

$$dX_t = -\theta X_t dt + \sigma dW_t$$

它描述了一个粒子在被拉向原点的力（漂移项）和随机碰撞（扩散项）共同作用下的运动。随着时间推移，它会收敛到一个以原点为中心的正态分布。

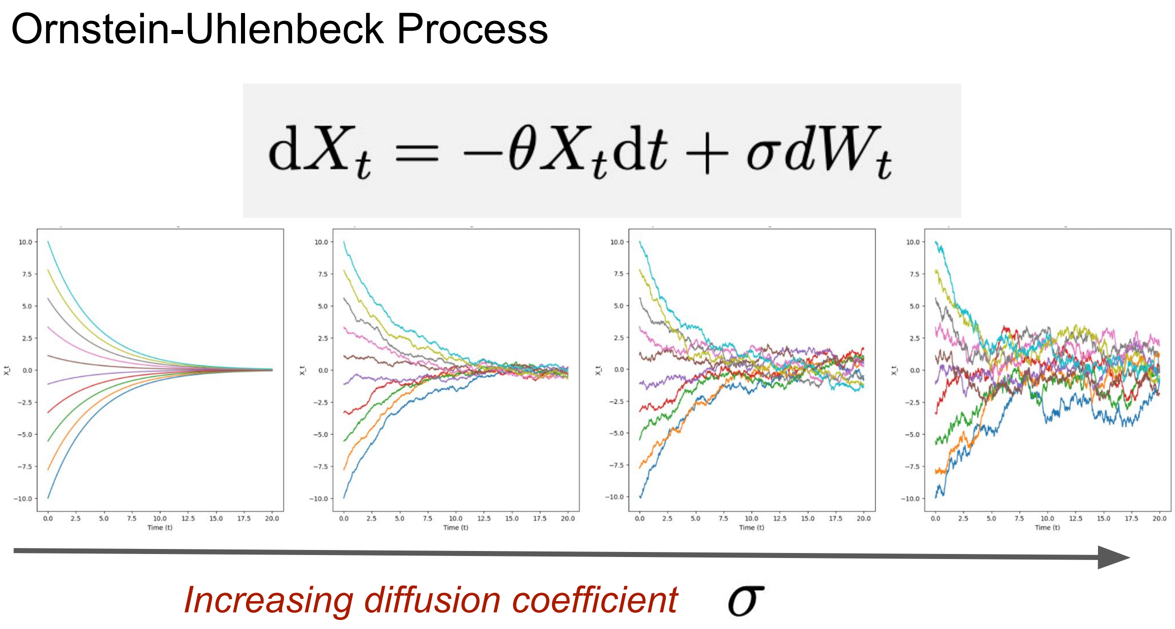

图 6: Ornstein-Uhlenbeck过程的演化图，展示了扩散系数对随机性的影响 (源自幻灯片第27页)。

2.2.4 数值模拟：欧拉-丸山法

求解SDE需要使用**欧拉-丸山法 (Euler-Maruyama Method)**。它在欧拉法的基础上，为每一步都增加了一个随机项。

Algorithm 2 使用欧拉-丸山法模拟SDE

需要: 向量场 u_t, 扩散系数 σ_t, 初始条件 x_0, 时间区间 $[0,1]$, 步数 N
$h \leftarrow 1/N$
$X \leftarrow x_0$
$t \leftarrow 0$
for $i = 0$ to $N-1$ **do**
  $\epsilon \sim \mathcal{N}(0, I_d)$ ▷ 从标准正态分布中采样一个随机向量
  $X \leftarrow X + h \cdot u_t(X) + \sigma_t \sqrt{h} \cdot \epsilon$ ▷ 增加随机扰动项
  $t \leftarrow t + h$
end for
return X