Dinh 在《Nice: Non-linear independent components estimation》提出了一种重要的生成模型——流模型 NICE，这篇论文被视为流模型的开山鼻祖。相较于其他生成模型，理论上，流模型能够求出原始数据的一个近似的密度函数显式表达。

NICE 的基本思想是：通过神经网络，寻找一个连续几乎处处可微的可逆非线性双射转移函数 f，将原始数据 x 转换到服从指定先验分布（作者推荐各维度独立的标准正态分布和标准逻辑分布）的数据$h = f(x)$，借助

$$p_X(x) = p_H(h)\left|det\frac{\partial f(x)}{\partial x}\right| \quad （1）$$

找到原始数据一个简单的易于建模的概率分布形式。通常，这样构造的困难是变换对应的雅克比行列式难以计算和变换的逆难以获得。

若矩阵经过适当的分块之后，是三角形式的矩阵，那么矩阵行列式就等于对角线上矩阵行列式之积。因此，NICE 将变量维度一分为二，一部分变量$\mathrm{x}_{\mathrm{I}_1}$保持不变，对另一部分变量$\mathrm{x}_{\mathrm{I}_2}$进行变换，构造如下变换：

$$\begin{cases} y_{I_1} = x_{I_1} \\ y_{I_2} = g\left(x_{I_2}, m(x_{I_1})\right) \end{cases}$$

将 m 称为耦合函数（NICE 中是一个深度神经网络），g 称为耦合法则，变换称为耦合层。可以发现，这样构造的变换对应的雅可比行列式可以更简单的求得。使用加法耦合法则，就可以构造出易于求得逆变换的结构，而且一般可以取得数值稳定的效果。

$$\begin{cases} y_{I_1} = x_{I_1} \\ y_{I_2} = x_{I_2} + m(x_{I_1}) \end{cases} \xrightarrow{\text{求逆}} \begin{cases} x_{I_1} = y_{I_1} \\ x_{I_2} = y_{I_2} - m(y_{I_1}) \end{cases}$$

如果$f_1, f_2, \ldots, f_L$都是这样的加法耦合层，那么对于$f = f_L \circ \ldots \circ f_2 \circ f_1$

这样的分层复合变换，这样良好的性质仍然不会消失，f的雅可比行列式和逆变换也可以简单地求得。将原始数据按变量维度一分为二后，在不同层对这两部分数据进行交替，从而完成学习，至少需要三个这样的耦合层才能让所有维度能互相影响，作者建议使用四个。

通常使用似然函数准则作为训练准则，原始数据的似然函数可以表示为：

$$log\big(p_X(x)\big) = log\big(p_H(h)\big) + log\,(|det\frac{\partial f(x)}{\partial x}|)$$

使用加法耦合法则时， $|det\frac{\partial f(x)}{\partial x}|$恒为 1，理论上，似然函数值会无限制的增加。作者提出的解决办法是：添加一个对角缩放层，对数据进行缩放变换，对应雅克比矩阵对角线上元素为正数，其他位置上的数据均为 0。通过加入的缩放变换，既可以解决似然函数无限制增加的可能性，也允许模型在一些维度上给出更多的权重，衡量分量的重要性。

生成模型的一大作用是生成数据，通过神经网络训练f，得到四个耦合函数和对角放缩层后，就能够根据公式（1）进行密度估计，将先验分布的随机数输入f^{-1}，得到采样后的数据。此外，也可以将 NICE 用于缺失值预测、图像修复、近似白化等领域。

作者：张岐彪

审阅：王洪桥