

大自然的计算：从伊辛模型到生成学习

王磊^{1,†} 张潘^{2,††}

(1 中国科学院物理研究所 北京 100190)

(2 中国科学院理论物理研究所 北京 100190)

2025-01-10收到

† email: wanglei@iphy.ac.cn

†† email: panzhang@itp.ac.cn

DOI: 10.7693/wl20250102

CSTR: 32040.14.wl20250102

The computation of nature: from the Ising model to generative learning

WANG Lei^{1,†} ZHANG Pan^{2,††}

(1 Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China)

(2 Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China)

摘要 2024年诺贝尔物理学奖授予约翰·霍普菲尔德和杰弗里·辛顿，这对很多人来说是出乎意料的。文章将从统计物理的视角，从伊辛模型出发，逐步介绍霍普菲尔德和辛顿的主要贡献，其中包括Hopfield模型、玻尔兹曼机、非监督学习，以及现代生成模型。还将回顾统计物理和机器学习在20世纪末期的精彩合作历程，并对未来物理与机器学习交互领域的发展方向进行简单展望。

关键词

玻尔兹曼机, Hopfield模型, 伊辛模型, 机器学习, 统计物理, 生成模型

Abstract The 2024 Nobel Prize in Physics was awarded to John Hopfield and Geoffrey Hinton, which came as a surprise to many people. In this introductory article, we will explore their major contributions from the perspective of statistical physics, starting with the Ising model. This includes the Hopfield model, Boltzmann machines, unsupervised learning, and modern generative models. The article will also review the exciting collaboration between statistical physics and machine learning at the end of the last century and provide a brief outlook on the future directions of the interdisciplinary field between physics and machine learning.

Keywords Boltzmann machine, Hopfield model, Ising model, machine learning, statistical physics, generative model

1 引言

2024年诺贝尔物理学奖的颁奖词是“for foundational discoveries and inventions that enable machine learning with artificial neural networks” (因奠定了使用人工神经网络进行机器学习的基础性发现和发明), “they trained artificial neural networks using physics” (他们使用物理学训练了人工

神经网络)^[1]。言下之意是颁发给“Science for AI”，这正好和诺贝尔化学奖^[2]的“AI for Science”形成呼应。

诺奖颁奖词还强调了两个名词，霍普菲尔德网络(Hopfield network)和玻尔兹曼机(Boltzmann machine)，这两个模型可以被视为生成学习的开端。生成学习的任务是从没有标签的数据中学习数据变量的联合分布概率，并对学习到的概率

分布进行采样从而生成新的学习样本。从物理的角度上看，这就是费曼说过的 “What I can not create, I do not understand”，其中 create 意味着生成，understand 意味着学习和理解。

生成学习的数据是多种多样的，对应的生成模型可以创造出各种各样的信息。数据是图像时，Midjourney^[3]可以画出新的画卷；数据是音符时，Wavenet^[4]可以生成新的乐章；当数据是自然语言时，ChatGPT^[5]可以写出华丽的文章。生成模型面对的最大挑战是如何学习到在一个巨大空间中的概率分布。以 28×28 像素的黑白图片为例，全部可能的图片数目是 2^{784} ；对于一首长度为 20 个汉字的诗，即使每个汉字只有 3000 种可能性，所有可能出现的诗句数目为 3000^{20} 。这都是个天文数字，不可能简单地通过穷举每个可能的图片或者诗句这种方式来描述像素或者汉字的联合分布概率。在 20 世纪，由于缺乏数据、计算能力，以及具体的应用，这种类型的高维概率分布的建模问题在各个领域还不是很普遍，但统计物理在其建立之初已经习惯于研究类似的高维概率分布问题了。熟悉统计物理的读者可能会立刻意识到，描述高维空间中概率分布的性质是统计物理的主要研究对象。例如，在统计物理中最常见的概率分布之一是玻尔兹曼分布：

$$P(s) = \frac{1}{Z} e^{-\beta E(s)},$$

其中 $E(s)$ 是能量函数， $Z = \sum_s e^{-\beta E(s)}$ 是配分函数，即玻尔兹曼分布的归一化因子。如果要问每个构型 s 严格的概率，我们需要知道配分函数的数值。直接的配分函数计算需要对 2^n 个构型的玻尔兹曼权重求和，这在大的系统中也是个天文数字，无法直接计算。因此我们可以看到，在机器学习和统计物理中都面对着非常相似的计算困难，也都需要有效的算法，以及强大的计算能力。相比于统计物理中那些自然界给定的哈密顿量和分布，在机器学习中描述概率分布是更加困难的问题，因为人们甚至没有一个“给定”的模型来描述数据分布。正如本文将向大家介绍的，在生成学习的开端，受到统计物理的启发，杰弗里·辛顿

(Geoffrey Hinton)确实是用自然界的分布——玻尔兹曼分布，来参数化数据的概率分布的。

再者，因为许多数据都可以看作是离散变量的组合，因此可以用统计物理中的自旋来进行表示。最简单的例子是黑白的图片：白色的像素可以映射为向上的自旋，黑色的像素可以映射为向下的自旋，这样整张黑白图片就可以映射为一个伊辛构型；对于多值的数据来说也是同样的，只不过映射成了 Potts 构型。基于这两个原因，早期的生成学习受到了统计物理的启发，而生成模型也自然地源于统计物理中的伊辛模型和玻尔兹曼分布。

2 历史

在自然界中，玻尔兹曼分布是司空见惯的。例如我们都很熟悉的冰和液态水，它们由同样的水分子构成，但水分子的构型在不同温度下遵从的玻尔兹曼分布不同。在不同温度下对玻尔兹曼分布进行采样，所观察到的水分子的构型就不相同了。这造就了在低温下常常看到冰这种形态，而在高温下一般看到的是水这种形态。而这些样本是大自然从玻尔兹曼分布中采样所生成的，采样的能力来自于自然界。

在这里我们可以加以类比，把图 1 左边的冰

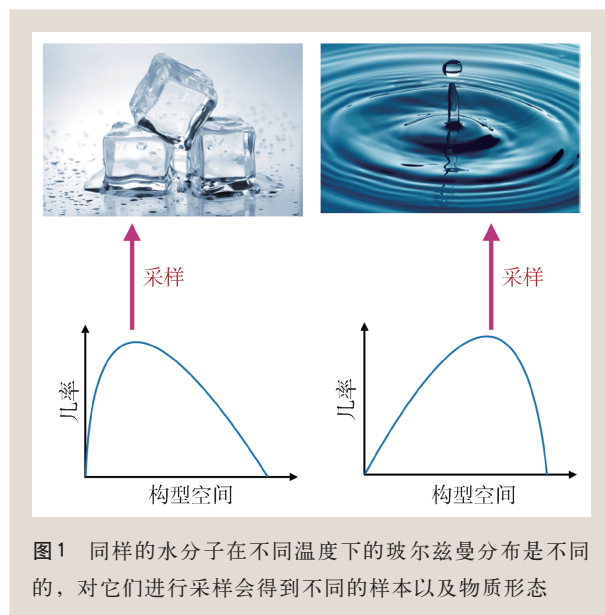


图1 同样的水分子在不同温度下的玻尔兹曼分布是不同的，对它们进行采样会得到不同的样本以及物质形态

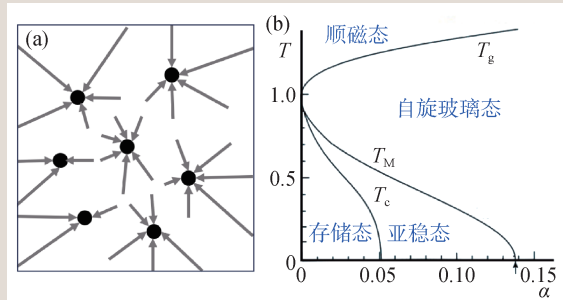


图2 (a) 数据作为动力学的吸引子；(b) Hopfield 模型的相图^[9]

当作一张数据，右边的水也当作一张数据。这样低温时的玻尔兹曼分布编码了冰这个数据，而高温时的玻尔兹曼分布编码了水这个数据——这是大自然的生成模型。

另外一个例子是统计物理中著名的伊辛模型^[6]。伊辛模型的变量被称为自旋变量，每个变量只能取+1和-1两种状态。 n 个自旋变量状态的组合被称为一个构型 $s = \{1, -1\}^n$ 。最简单的铁磁伊辛模型的能量函数 $E(s) = -\sum_{(ij)} s_i s_j$ ，即自旋之间的耦合参数 $J_{ij} = 1$ 。可以看到，能量是由每一组 (ij) 对所贡献的，相当于第 i 个自旋和第 j 个自旋都有相互作用。它们取相同的值会使这个自旋对的能量最小。所有可能的构型共 2^n 个，每个构型 s 的概率取决于能量，也取决于温度 $P(s) = \frac{1}{Z} e^{\beta \sum_{(ij)} s_i s_j}$ 。在非常高温的时候 β 几乎为0，所有构型的概率都一样。在非常低温 $\beta \rightarrow \infty$ 下，只有那些能量最低的构型，即所有自旋都取+1，或者都取-1的那些构型才有足够高的概率，其他构型的概率几乎是0。如果从低温的伊辛模型中进行采样，所得到的(或者说生成)的构型就只有两种。基于前面所介绍的从图片到伊辛构型的类比，不同的温度对应不同的生成模型，而低温下的铁磁伊辛模型可以认为是学习了全部都是黑色和全部都是白色像素的两幅图片。这几乎是最简单的图片生成模型了，学习到的概率分布只有两张图片的概率值非零，其他构型概率均为零，因此只能记住两张图。

Hopfield 模型^[7]可以认为是铁磁伊辛模型的推

广，它是个伊辛模型，仍然用玻尔兹曼分布来表达数据变量的分布。但受到神经生物学的Hebb规则^[8]所启发，耦合参数是从数据的关联得到的 $J_{ij} = \frac{1}{an} \sum_{\mu=1}^{an} \xi_i^\mu \xi_j^\mu$ 。其中一共存储了 an 个数据， μ 是数据的索引， i 和 j 是自旋指标。如果只存储了一个数据，即 $an = 1$ ，可以看到 $J_{ij} = \xi_i \xi_j$ ，仍然是一个自旋变量。这个模型也被称为 Mattis 玻璃，其实和铁磁伊辛模型只相差一个将+1映射到随机 ± 1 的变换，本质上是一样的，只存储了两个数据 $\{\xi_i^\mu\}$ 和它的镜像。当我们运行一种动力学，例如马尔可夫链蒙特卡罗法对系统进行采样时，在零温下只能采样到存储的数据，在低温下可以采样到和存储的数据相似的构型。当 $an > 1$ 时，情况就不同了。为了方便理解，我们考虑使用并行的Glauber动力学对系统进行采样，给定系统的状态 s^t ，对每个自旋 i 来说，下一个时刻取 s_i^{t+1} 的概率为 $P(s_i^{t+1}) \propto e^{\beta \sum_j J_{ij} s_i^{t+1} s_j^t}$ 。在存储的数据比较少时，这些数据都会表现为动力学的吸引子。当网络状态已经在存储的一个数据时，下一个时刻取值概率为

$$P(s_i^{t+1} = \xi_i^1) \propto e^{\frac{\beta}{n} + \frac{\beta}{m} \sum_{\mu=1}^m \sum_{j=1}^n \xi_j^\mu \xi_i^1 \xi_j^1}.$$

可以看到，这个概率有两部分构成，第一部分是信号项，第二部分是由很多项数据变量乘在一起的噪音项。信号项倾向于待在第一个数据，而噪音项驱使动力学离开第一个数据。当存储的数据很多时，噪音会很大，使得每一个数据都不能再成为动力学的吸引子，从而无法记住任何一个数据。Hopfield模型的平衡态分析在20世纪80年代是非常重要的问题，在1985年时，三位以色列科学家利用统计物理的复本方法(replica method)求解Hopfield模型的相图(图2)^[9]。相图分为三个区域：在高温时系统为顺磁态；在低温且 α 小时，系统可以记住训练的数据，成为记忆态；在低温且 α 大时，系统忘记了所有训练数据的态，称为自旋玻璃态。在零温时 Hopfield 模型最多可以存储 $0.138n$ 的数据。

在20世纪70到90年代，对Hopfield等模型的平均场理论研究是统计物理的一个重要研究方

向。当时的一些统计物理学家认为神经网络的分析可以被当成统计力学的反问题(inverse problem)来处理^[10]。正问题是给定系统的能量函数,求解玻尔兹曼函数的性质,以及获得玻尔兹曼分布无偏的样本。而反问题(也称反伊辛模型)则认为玻尔兹曼分布的样本已经获取了,就是用来训练的数据,要从这些样本推断(或者学习)出玻尔兹曼分布的参数,即能量函数中的参数。

从历史的角度看,统计物理与机器学习在20世纪确实有着很长一段时间的交互和互相发展阶段。基于个人观点,我们总结了一个对照图,如图3所示,大致分为4个阶段。在19世纪时,图3统计物理这边可以看到一些著名的统计物理学家的名字,而在机器学习中最小二乘法、贝叶斯定理被建立和使用。在1900年到1950年左右的第二阶段,统计物理在伊辛模型方面、采样等方向逐渐发展起来,特别是在1957年Jaynes提出了最大熵原理,建立了统计物理与统计推断的联系。在机器学习方面,信息与计算的基础理论逐步建立起来,感知器(perceptron)^[11]也在20世纪50年代被发明并引发了人工智能的第一波研究热潮。然而到了1969年,Minsky和Papert^[12]指出感知器甚至都解决不了XOR分类问题,因此大家开始重新审视人工智能,人工智能也陷入了第一个寒冬。在图3中的第三个阶段,统计物理在自旋玻璃领域获得了进展。自旋玻璃是具有随机耦合参数的伊辛模型,能量图景非常丰富。2021年诺贝尔物理学奖得主Parisi教授为Sherrington—Kirkpatrick自旋玻璃^[13]建立了全阶副本破缺的渐进严格解^[14],展示了其能量图景具有自相似的层级结构。研究自旋玻璃使用的方法,例如副本方法,也可以用来研究神经网络模型,这其中的著名工作包括Gardner和Derrida关于感知器的容量^[15],Krauth和Mezard关

于二元感知器容量^[16]的统计物理研究。在神经网络领域,这个阶段建立了基础的监督与非监督模型,例如卷积网络的前身Necognitron、玻尔兹曼机、反向传播算法,以及强化学习的开端。然而需要注意的是,统计物理的方法在当时并不能成功理解更复杂的卷积网络、玻尔兹曼机等神经网络的运行机理,因为无法分析多层的网络,也很难为实际的数据建立模型。然而在1995年时,支持向量机(support vector machine)^[17]的出现使得人工智能领域的注意力从神经网络转移了出来,因为支持向量机当时比多层神经网络(MLP)的结果更好,而且更容易解释。这又使得神经网络的研究进入了第二个寒冬。在2000年之后的发展中,统计物理方法被应用于更加广泛的领域中,这包含组合优化、统计推断问题等,但由于仍然很难分析多层神经网络的性质,统计物理与神经网络领域的发展逐渐分道扬镳,并最终错过了机器学习在2012年之后产生的深度学习时代的序幕。

在机器学习方面,生成学习的确是在20世纪80年代时沿着玻尔兹曼分布展开的。Hopfield模型虽然在理论神经科学、联想记忆、机器学习领域都具有里程碑似的意义,但它最多能存储 $0.138n$ 个数据^[9],且要求数据之间不能存在关联,

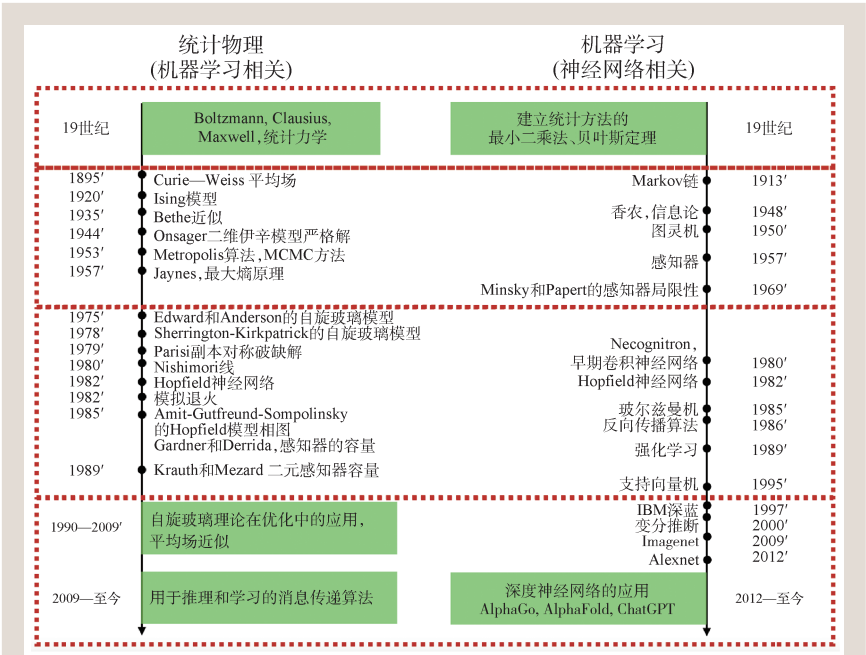


图3 与机器学习相关的统计物理和神经网络发展的对照表格

这给 Hopfield 模型的实际应用带来了很多困难。对 Hopfield 模型的改进是非常自然的：可以不用通过 Hebb 规则构造伊辛模型的耦合参量 J_{ij} ，逐步学习这些耦合参量。在统计物理中这种做法被称为反伊辛问题(inverse Ising problem)。在反伊辛问题中，损失函数 $\mathcal{L}(\mathcal{D})$ 是负对数似然度(negative log-likelihood)。对它优化的目标是找到一组 J_{ij} ，使得玻尔兹曼分布在生成所有训练数据 $\mathcal{D}$ 时的概率最大化，也等价于最小化训练数据的经验分布与玻尔兹曼分布之间的 Kullback—Leibler (KL) 散度。通常的优化方法是计算出损失函数相对于参数的梯度，然后利用梯度信息优化损失函数。对于反伊辛问题^[18]，它的梯度为

$$\frac{\partial \mathcal{L}(\mathcal{D})}{\partial J_{ij}} = \langle s_i s_j \rangle_{\text{Data}} - \langle s_i s_j \rangle_{\text{Model}}.$$

其中 $\langle s_i s_j \rangle_{\text{Data}}$ 是数据中第 i 个变量与第 j 个变量的关联，而 $\langle s_i s_j \rangle_{\text{Model}}$ 是在模型中的关联。这个梯度的表达式从侧面说明了，伊辛模型是给定数据一阶矩和二阶矩的最大熵模型，作为指数族函数，它的充分统计量(sufficient statistics)是一阶矩和二阶矩，这也意味着它很难刻画数据中的高阶关联。另外，虽然数据关联是容易计算的，但模型关联的计算与配分函数的计算同属于 #P 难问题，不存在多项式的严格算法。除了梯度计算的困难之外，反伊辛模型的表达能力也不足，因为所有参数一共只有 $n^2/2$ 个，很难扩展。

辛顿与合作者发明的玻尔兹曼机(Boltzmann

machine, BM)^[19]可以解决反伊辛模型中对关联的表述，以及模型的整体表达能力问题。在玻尔兹曼机中， n 个显变量的构型 $\mathbf{v} = \{v_1, v_2, \dots, v_n\}$ 以及 m 个隐变量的构型 $\mathbf{h} = \{h_1, h_2, \dots, h_m\}$ 的联合分布概率满足玻尔兹曼分布：

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h})}.$$

而显变量(数据变量)的联合分布概率通过对隐变量的求和来得到：

$$P(\mathbf{v}) = \sum_{\mathbf{h}} P(\mathbf{v}, \mathbf{h}).$$

通过引入多个隐变量，使得所有变量都连接在一起。这样一方面增加了系统的参数数目，因为隐变量和显变量之间的耦合参量的数目随着隐变量的增加而增加；另一方面也通过与所有显变量连接的隐变量诱导了显变量之间的高阶关联。虽然解决了表述能力的问题，但玻尔兹曼机没有解决反伊辛模型的梯度计算问题，实际上这个问题变得更严重了。例如，对数似然度相对于显变量 i 和隐变量 a 之间的耦合参量 W_{ia} 的梯度可以写成：

$$\frac{\partial \mathcal{L}(\mathcal{D})}{\partial J_{ij}} = \langle s_i s_j \rangle_{\text{Data} + \text{Model}} - \langle s_i s_j \rangle_{\text{Model}}.$$

不仅第二项的梯度计算没有变得简单，连第一项的计算也变得更为复杂。为了解决这个问题，辛顿与合作者提出了受限玻尔兹曼机(restricted Boltzmann machine, RBM)^[20]。相比于玻尔兹曼机，RBM 去除了显变量之间的连接以及隐变量之间的连接。因此它是一个定义在只有显变量和隐变量之间有连接的二分图上。虽然损失了一些参数和表达能力，RBM 的优点是可以快速地进行采样，然后利用样本计算 $\langle s_i s_j \rangle_{\text{Data} + \text{Model}}$ 以及 $\langle s_i s_j \rangle_{\text{Model}}$ 。这个性质完全来自于二分图，这使得给定隐变量之后，显变量的概率分布可以严格地分解为连乘的形式：

$$P(\mathbf{v} | \mathbf{h}) = \prod_i P(v_i | \mathbf{h}).$$

当给定隐变量的构型以后，显变量可以非常有效地进行采样。根

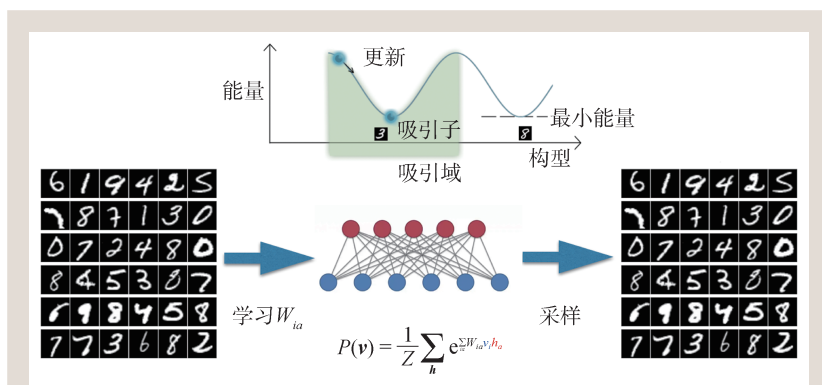


图4 从训练数据中学习得到 RBM 的参数 W_{ia} ，使得在 RBM 的能量图景中，数据构型能量低而随机构型能量高。从 RBM 的分布中采样可以生成新的数据

据同样的道理，给定显变量构型(即数据构型)，隐变量的概率分布也具有连乘的形式，因此隐变量也可以迅速采样。通过这个性质来计算RBM梯度并进行学习的方法被称为对比散度算法，如图4所示。训练用的数据是一组二值的手写字符图片，每个图片被视为一组自旋构型。通过最大化对数似然的方法，定义在二分图上的RBM参数 W_{ia} 可以被学习到。在学习结束之后，RBM的能量图景具有图中的形式，即训练图片以及和它们相近的图片的能量比较低，而其他那些随机构型图片的能量都比较高。这样如果从能量高的构型出发做采样，就可以生成和训练图片相似的图片。

RBM中的隐变量不仅可以为显变量提供多体相互作用，也可以作为数据变量的表示：如果我们把 $P(\mathbf{v}, \mathbf{h})$ 对显变量求和，就能得到隐变量的边际分布。如果隐变量比较少，这个分布可以认为是数据分布的低维表示。这和物理中的重正化群有相似之处。可以想象一个二维伊辛模型，如果每个隐变量连接到二维格点上相邻的4个自旋上，那么把自旋变量求和就类似做了一次重正化群，提取了系统粗粒化的信息。辛顿与他的学生 Ruslan Salakhutdinov 在2006年发表了一篇 *Science* 论文^[21]，提出可以将多个RBM堆叠起来，将下一层的RBM的隐变量当作上一层的显变量，隐变量数目逐层减少，最终堆成一个金字塔形状的深度玻尔兹曼机。将这个金字塔倒着扣过来，和之前的玻尔兹曼机形成一个沙漏形状的神经网络，辛顿将它称为自编码器。自编码器输入和输出的维度是一样的，如果可以训练它使得输入和输出的结果几乎一致，就意味着在最中间那个维度最低的隐层中找到了一个由隐变量所承载的低维表示。这个训练的方式被称为非监督预训练，如今已经在以GPT^[5]为代表的大语言模型中广泛采用。其

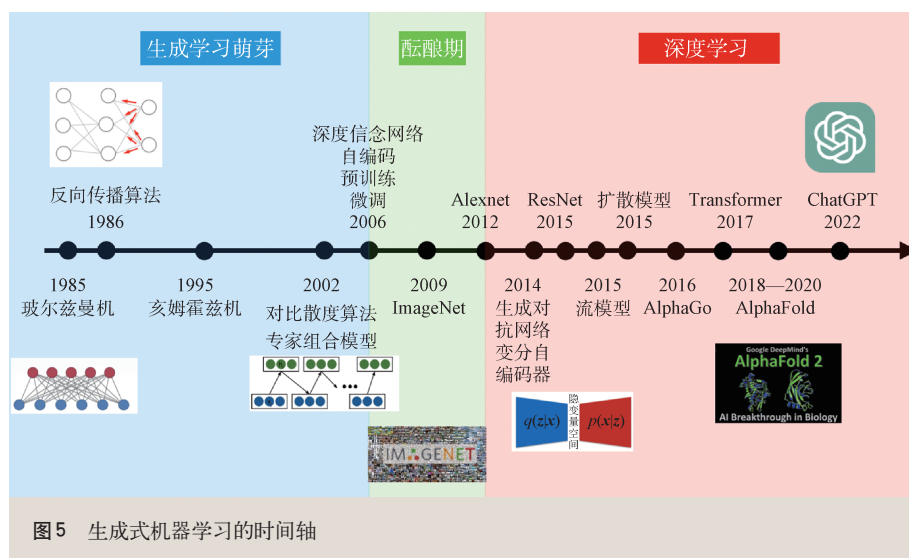


图5 生成式机器学习的时间轴

中GPT的含义是generative pretrained transformer，即生成式预训练Transformer。对于深度玻尔兹曼机和自编码器来说，预训练之后的模型可以做好很多事情，例如在图像分类任务中可以工作得比支持向量机更好。至此，也拉开了新一轮神经网络研究热潮的序幕。

3 当下

自2006年以后，虽然由深度玻尔兹曼机启发的自编码器、非监督预训练具有很好的应用前景，但玻尔兹曼机的发展却陷入了瓶颈，似乎这条路线走向了一条死胡同。按照本文作者的理解，主要原因有两个，一是玻尔兹曼机的配分函数计算困难，这使得模型无法严格计算一个数据的概率值和似然度；二是虽然有RBM的隐变量—显变量交替采样，但高效的采样仍然是一个难题。这两个问题不仅困扰玻尔兹曼机，实际上也是统计物理对于玻尔兹曼分布性质进行研究时的主要困难。虽然当时大家尝试了包括退火重要采样(annealed importance sampling, AIS, 几乎是Jarzynski equality的具象化算法)^[22]在内的多种近似采样和计算方法，效果都不是太显著。

给这个领域带来真正突破性进展的事件来自于2009年和2012年(图5)。在2009年，带着“Data will redefine how we think about models”的理念，华人计算机科学家李飞飞在WordNet的基

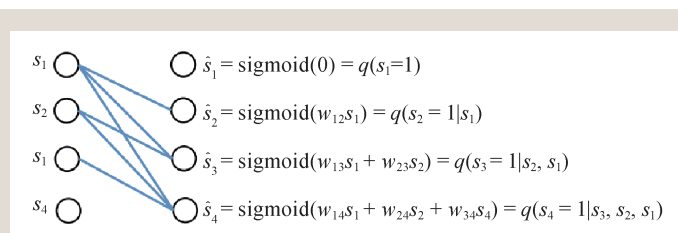


图6 有4个变量的自回归模型的一个简单实现

基础上，利用 Amazon 的众包服务为一个具有 1200 万张图片的数据集打上了人工的标签，并组织了每年一度的 ImageNet^[23] 挑战赛。在 2012 年时，辛顿带领学生 Alex Krizhevsky 和 Ilya Sutskever 用 Alexnet^[24] 深度神经网络以极大的优势战胜了当时所有的竞争对手。Alexnet 的横空出世可谓是天时地利人和，将深度卷积神经网络这个模型，利用当时先进的 GPU 计算能力，应用到大量数据 Imagenet 上，才获得了成功。这使得深度神经网络一战成名，正式宣布深度学习时代的到来。2012 年之后，在生成学习领域也出现了快速的发展，2014 年的 Generative Adversarial Network^[25]，Variational Autoencoder^[26]，2015 年的 Flow Models，Diffusion Models，2017 年的 Self-Attention 机制和 Transformer^[27] 接踵出现，最终催生了 2022 年的 ChatGPT^[5]。

可以看到在这场现代生成模型的浪潮中，玻尔兹曼机的的身影逐渐消失了。具体的原因是现代的生成模型从模型的角度上解决了玻尔兹曼机的配分函数计算以及采样这两个难题。这里我们以现代生成模型之一的自回归神经网络模型(可参阅作者 2024 年第 6 期在本刊发表的科普文章《写给物理学家的生成模型》)为例具体解释是如何解决这个问题。在自回归模型中，联合概率分布具有条件概率连乘的形式 $q(s) = \prod_i q(s_i | s_{j < i})$ 。首先，变量被事先制定了一个顺序， $j < i$ 代表在变量 i 之前的变量，在这里称为历史变量， $s_{j < i}$ 代表它们的构型。 $q(s_i | s_{j < i})$ 的意义是给定 i 之前变量的构型，变量 i 取 s_i 的条件概率。图 6 给出的例子中 s_i 为二值变量，因此在给定历史变量构型之后的条件概率可以用 Sigmoid 函数来参数化。以 4 个变量的自回归模型为例，严格的概率分布为

$q(s_1, s_2, s_3, s_4)$ 。使用贝叶斯公式，可以先将它逐步写成条件概率的乘积：

$$q(s_1, s_2, s_3, s_4) =$$

$$q(s_1)q(s_2|s_1)q(s_3|s_1, s_2)q(s_4|s_1, s_2, s_3) \cdot$$

每个条件概率通过神经网络来参数化，如图 6 所示。

注意，对于不同历史构型的条件概率的参数是共享的。在实际应用中，不同变量条件概率的参数也可以共享。在图 6 所展示的参数化下，可以看到，给定任何一个网络构型 s_1, s_2, s_3, s_4 作为输入，网络输出的连乘 $\hat{s}_1 \hat{s}_2 \hat{s}_3 \hat{s}_4$ 即为这个构型的联合分布概率，而且其归一化因子为 1。另外，既然已经存储了所有的条件概率，对联合概率的无偏样本可以通过对每个条件概率单独采样来得到。在图 6 中我们可以取 $q(s_1 = 1)$ 采样出 s_1 ，然后带入到 $q(s_2 = 1 | s_1)$ 来采样出 s_2 ，然后将 s_1, s_2 代入到 $q(s_3 = 1 | s_2, s_1)$ 来采样 s_3 ，以此类推。而且多个样本可以被并行地采样，从而高效地利用 GPU 的计算能力。因此自回归模型用条件概率乘积的形式解决了玻尔兹曼机的配分函数计算以及无偏采样问题。

人们利用 ChatGPT 和类似的模型根据提示词生成新的文本，这个生成过程的本质是利用了自回归模型的采样，因为描述一句话中字符(token)的联合分布概率的 Transformer 加上了因果掩码(Causal mask)之后，就可以从学习到的联合分布中一个词一个词的采样，就像图 6 中所展示的，一个自旋一个自旋地生成 4 个变量的自旋构型一样。这种方式也造就了近年来大语言模型的辉煌。

在连续变量的联合分布刻画问题中，更有效的模型是流模型(flow model)、扩散模型(diffusion model)，以及流匹配方法(flow matching)，它们利用不同的方式来解决采样问题，进而提供了图片、视频以及音乐生成的精彩应用。具体的介绍可以参考作者的科普文章《写给物理学家的生成模型》。

4 未来

让我们回到文章开始提到的诺贝尔物理学奖，以及物理与机器学习的关系。在前面内容中，我

们介绍了统计物理与机器学习在20世纪末的互动，也介绍了生成学习早期的推动与启发。在2012年之后机器学习飞速发展的阶段，统计物理虽然在机器学习模型(例如扩散模型等)上提供了启发和帮助，也取得了一些对神经网络的理论研究进展，例如双下降、过参数化、玻璃性质、随机梯度下降的泛化性等。但总体上来说并不是不可或缺的，甚至可以说在“Science for AI”这个方向上，统计物理总体缺席了深度学习的这场盛宴。究其原因，作者个人认为是由于之前用于处理自旋玻璃、感知机以及Hopfield模型的那些方法，例如复本方法、空腔方法、复本对称破缺、消息传递算法等，更善于处理浅层的、线性的网络，以及刻意合成的数据，而对目前复杂的深度神经网络、生成模型以及复杂的数据缺乏有效的分析和理解能力。同时，我们也见证了在“AI for Science”方向上，基于机器学习的新方法在物理学各个方向中的应用。应用的范围从 10^{-10} m、 10^{-15} s尺度的量子力学，到微米和皮秒尺度的分子动力学，再到米和秒尺度的流体力学、经典力学，一直延伸到大尺度的天文、宇宙学中。

在统计物理领域中，机器学习方法在2016年左右就被用于识别不同的物质相，也被用于统计力学中优化变分自由能以及刻画非平衡系统的时间演化，这个方向上的应用和强化学习有很多相似之处。在非线性动力学的预测中也有很长时间的的应用历史。在量子多体领域中，神经网络模型被用于表述波函数的拟设，通过学习神经网络的参数优化变分能量，寻找系统基态，也被称为神经网络量子态或者神经网络变分蒙特卡罗。除了利用到神经网络强大的表述能力之外，好的进展包括让神经网络量子态具备例如费米子交换反对称性等内禀的对称性。另外，量子多体中的经典方法(例如张量网络等)也和机器学习有过有趣的交互，其中包括利用自动微分优化张量网络，利用张量网络构造易于解释且有特点的监督与非监督机器学习模型。在第一性原理和材料计算领域，机器学习的应用非常广泛，也非常有效。著名的工作包括如何从密度泛函的数据中学习能量和力场，甚至是哈密顿量，学习新的泛函；探索复

杂系统的自由能面和集体变量；物质生成方面，在保证对称性的情况下从数据生成走向原子、分子的生成。

在粒子物理和宇宙学方向上，例如LHC、LSST、LIGO等大科学装置需要处理大量的数据，本身就离不开机器学习的方法。而在处理数据的过程中需要做大量正向的量子场论、微扰方法、广义相对论等模拟计算，已经广泛地采用神经网络和GPU进行计算的大幅加速，机器学习方法也广泛应用于Jet物理中的flavor标记、jet聚类、谱密度分析等等。在中微子物理中，神经网络被用于信号处理从而寻找中微子相互作用位置；在引力波物理中，神经网络分类器被用于快速判断是否存在引力波信号，流模型也被用于引力波波源的参数学习。在格点量子色动力学中，即使不考虑符号问题计算也非常困难，传统方法利用Hamilton MCMC需要花费巨大的计算代价才能获取有限的组态，因此如何利用生成模型提高组态获取的效率是非常重要的问题。

近年来，机器学习与量子计算的交互变得更加显著。在量子机器学习领域中，大家关心的问题包括如何结合量子与机器学习在数据表示、量子算法设计、量子优势以及混合经典—量子计算中取得突破。机器学习也被用来设计更好的量子线路，减少线路深度，优化量子门参数。在最近的实验中，谷歌的量子计算团队利用机器学习帮助硬件进行量子纠错，充分证明利用神经网络可以高效地对具有测量噪音的表面码进行纠错，而且可以比传统的纠错方法获得更低的逻辑错误率。

如果展望物理与机器学习可能产生下一个重大突破的方向，在作者看来，最有可能的方向是机器学习与先进经典计算能力的结合，可以用于解决物理学领域的重要问题。

最后回到大自然的采样与生成学习。我们讨论过利用自然界的分布——玻尔兹曼分布进行采样启发了玻尔兹曼机等基础性的生成模型，也获得了诺贝尔奖。但也应该指出，玻尔兹曼分布具有计算归一化的概率值以及无偏采样的难题。这两个难题即使是大自然本身也是很难解决的，因为自然界的采样也会陷入亚稳态，从而产生过冷

液体、玻璃态等没有平衡的物质，因此利用了大自然计算能力的经典计算机很难彻底解决玻尔兹曼的采样问题。除了玻尔兹曼分布以外，物理学还有另外一个重要的概率分布，即玻恩法则(Born's rule)。玻恩法则中波函数的模平方所表达的概率给出了微观世界在量子力学控制范围内的观测概率。从这个概率分布中产生样本是自然界量子优势的一种体现，如果人们可以利用这种优势，也许可以在不久的将来极大地推进生成模型的发展。

参考文献

- [1] The Nobel Prize in Physics 2024. <https://www.nobelprize.org/prizes/physics/2024/summary/>
- [2] They cracked the code for proteins' amazing structures. <https://www.nobelprize.org/prizes/chemistry/2024/press-release/>
- [3] Midjourney 主页. <https://www.midjourney.com/>
- [4] WaveNet: A generative model for raw Audio. <https://deepmind.google/discover/blog/wavenet-a-generative-model-for-raw-audio/>
- [5] ChatGPT 主页. <https://www.chatgpt.com/>
- [6] 伊辛模型. https://en.wikipedia.org/wiki/Ising_model
- [7] Hopfield J J. Proceedings of the National Academy of Sciences, 1982, 79: 2554
- [8] Hebb 学习规则. https://en.wikipedia.org/wiki/Hebbian_theory
- [9] Amit D J, Gutfreund H, Sompolinsky H. Physical Review A, 1985, 32(2): 1007
- [10] Gutfreund H, Toulouse G. The Physics of Neural Networks. In: Spin Glasses and Biology. 1992. pp.7—59
- [11] Rosenblatt F. Proceedings of the IRE, 1960, 48(3): 301
- [12] Minsky M, Papert S A. Perceptrons: An Introduction to Computational Geometry. The MIT Press, 2017
- [13] Sherrington D, Kirkpatrick S. Phys. Rev. Lett., 1975, 35(26): 1792
- [14] Parisi G. Phys. Rev. Lett., 1979, 43(23): 1754
- [15] Gardner E, Derrida B. Journal of Physics A: Mathematical and General, 1988, 21(1): 271
- [16] Krauth W, Mézard M. Journal de Physique, 1989, 50(20): 3057
- [17] Cortes C. Machine Learning, 1995, 20: 273
- [18] Nguyen H C, Zecchina R, Berg J. Advances in Physics, 2017, 66(3): 197
- [19] Ackley D H, Hinton G E, Sejnowski T J. Cognitive Science, 1985, 9(1): 147
- [20] Hinton G E. Momentum, 2010, 9(1): 926
- [21] Hinton G E, Salakhutdinov R R. Science, 2006, 313(5786): 504
- [22] Neal R M. Statistics and Computing, 2001, 11: 125
- [23] Deng J, Dong W, Socher R *et al.* ImageNet: A large-scale Hierarchical Image Database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA. pp.248—255
- [24] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25, 2012
- [25] Goodfellow I, Pouget-Abadie J, Mirza M *et al.* Generative adversarial nets. In: Advances in Neural Information Processing Systems 27, 2014
- [26] Kingma D P. 2013, arXiv: 1312.6114
- [27] Vaswani A *et al.* Attention is all you need. In: Advances in Neural Information Processing Systems 30, 2017

新书推荐

读者和编者

《物理夜航船：直觉与猜算》一书以徐一鸿先生无与伦比、引人入胜的风格，向学习物理的学生介绍如何利用物理推理和明智的猜测来找到问题的关键。在典型的物理课堂上，学生们试图掌握一套庞大的数学工具，来进行物理中的精确计算。因此，学生往往会产生一种不幸的印象，认为物理由定义明确的问题组成，可以通过严密的推理和合乎逻辑的步骤来解决。理想化的课本练习和家庭作业问题强化了

这种错误印象。因此，即使是最优秀的学生也会发现自己完全没有准备好应对实际研究中的挑战。实际上，物理学充满了各种近似、简要估计、对数量级的猜测和跳跃的逻辑。这本不可或缺的书包括与物理学前沿主题相关的令人兴奋的问题(从霍金辐射到引力波)，通过精妙的“夜行法”分析，帮助学生深入理解他们所学的方程，并期待能对研究者们产生物理直觉有所帮助。



作者：徐一鸿 译者：姬扬
出版时间：2024年10月
定价：63元