

神经网络理论研究的物理学思想

黄海平

(中山大学物理学院 510275)

神经网络在当今人工智能研究和应用中发挥着不可替代的作用。它是人类在理解自我(大脑)的过程中产生的副产品,以此副产品,人类希望建造一个机器智能来实现机器文明。这个目标在当下如火如荼的人工智能研究中被无限倍凸显,甚至被认为是一场新的工业革命到来的标志。

在人类社会前几次工业革命浪潮中,物理学扮演了十分重要的角色,或者说,这些革命的理论基石在于物理学原理的突破,如热学、量子力学和相对论。但当今的人工智能革命似乎是经验科学(启发式的诀窍,如 Transformer)所驱动的,在过去 20 年间,尤其是谷歌等互联网巨头加入这场浪潮之后,人工神经网络的架构出现了快速迭代。物理学对神经网络的研究历史悠久,最早^①可追溯到 20 世纪 80 年代初霍菲尔德(与辛顿一起获得 2024 年诺贝尔物理学奖)联想记忆网络的提出;物理学思想在这之后对人工神经网络和神经动力学的研究都产生了深远的影响。著名物理学家戴森有一个说法:“严谨理论赋予一个课题以智力的深度和精确。在你能证明一个严格理论之前,你不可能全面了解你所关注的概念的意义。”^②获得玻尔兹曼奖的物理学家霍菲尔德也曾在一次访谈中提到,“如果你不能用数学的语言去描述大脑,那将永远不知道大脑是怎么工作的。”而鉴于他自身的习惯,“如果一个问题和我熟知的物理毫无联系,那我将无法取得任何的进展”。所以,在人工智能正在重塑人类社会方方面面的同时,我们有必要去了解物理学的思想如何影响人们对神经网络乃至自我的认知。

一、从伊辛模型谈起

伊辛模型是统计物理的标准模型^③。它虽然被

用来描述格点上(比如二维表面)磁矩的集体行为,但是却包含了非常丰富的物理图像(比如相变、自发对称性破缺、普适性等),更让人震惊的是,这个模型的物理图像可以向外扩展到多个似乎毫不相关的学科,如经济学、神经科学、机器学习等。我们先从物理学专业本科生所熟知的态方程讲起:

$$m = \tanh(\mathbf{J}m + \mathbf{h})$$

这显然是个迭代方程,因为变量 m 出现在方程式等号的两边,其中 \mathbf{J} 描述了自旋之间的相互作用, m 表示磁化强度矢量, \mathbf{h} 则表示外加磁场。注意到,该态方程在没有外加磁场并且相互作用较弱情况下,有且只有一个平庸解,即所有磁化为零,用物理学语言讲叫顺磁态。然而,当增大相互作用到一定程度时,顺磁态将失去稳定,该方程出现两个非平庸解(物理上叫铁磁解,即 $m = \pm M$)。这个过程叫自发对称性破缺或连续相变。

这个迭代蕴含了神经网络的形式。神经网络的基本属性可以总结为 DNA,即数据(data)、网络(network)和算法(algorithm),如图 1 所示。你把初始化 m_0 看成输入数据,每迭代一次将生成一个新的 m ,这个就是神经网络的中间隐层表示。然而,奇妙的是,神经网络把 \mathbf{J} 也变成可以变化的量,这就意味着这个模型是可以变聪明的(即能处理每一个输入)。这在传统物理学里很不可思议,因为模型通常需要大物理学家猜出来。而外场可以等价于神经网络的偏置(见图 1)。那么如何更新 \mathbf{J} 呢?你只需要写下一个目标函数,即这个神经网络,或者学习中的模型要达到什么样的目标。比如,实现数据的二分类,你可以轻松地写下 $E = \frac{1}{2} \sum_a [f_j(\mathbf{x}^a) - y^a]^2$,这里的 a 表示数据输入-输出对 (\mathbf{x}, y) (y 在机器学习

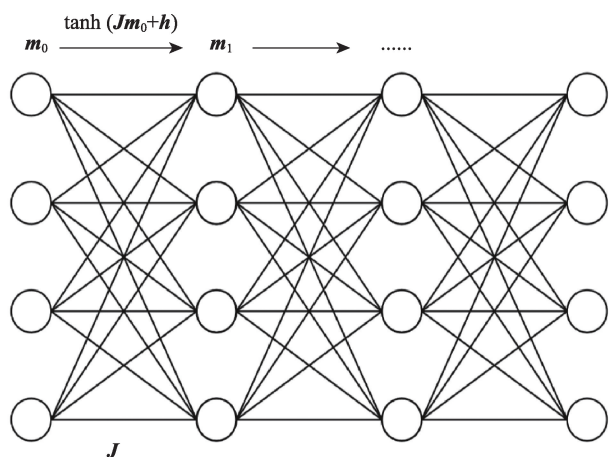


图1 神经网络的迭代示意图

叫标签),而 f 就是这个被 J 参数化的神经网络(其本质显然是一个非常复杂的嵌套函数,类似于上面态方程的多次迭代,只不过每次迭代的 J 都不一样)。

接下来你需要一个算法来驱动这个网络自我更新。这个算法其实就是梯度下降： $\frac{dJ}{dt} = -\nabla J E$ 。聪明的读者一眼就认出这是个过阻尼的朗之万动力学,因为人们在训练神经网络时通常在上方的方程右边加入微弱的白噪声。所以,神经网络的学习过程是在你为它定义的势能函数下的随机游走(或者布朗运动,见图2),如果你稍微学过一点随机动力学的话,你立马知道这个神经网络的学习过程存在平衡态,其分布正好是玻尔兹曼分布 $P(J) = \frac{1}{Z} e^{-E/T}$,其中 Z 就是统计物理的地标——配分函数,而温度 T 则控制学习过程随机涨落的程度,类似一个粒子在相同温度的溶液里运动。此刻,相信你已经获得

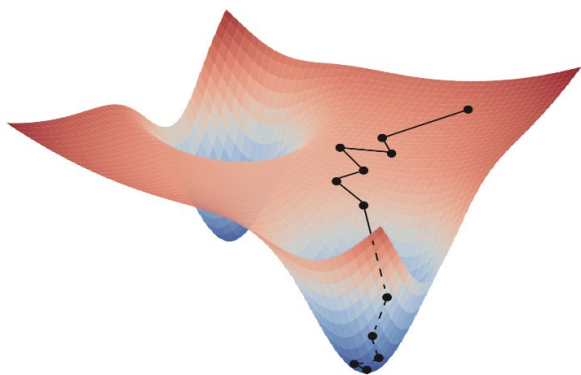


图2 神经网络学习过程

足够深刻的理解:神经网络的本质是一个从简单函数(如上述的 $\tanh()$,这个函数的形式源自物理上经典自旋有两个取值并且服从玻尔兹曼正则分布)反复迭代出来的超级复杂并且表达能力超强的函数。这个函数需要不断更新它的参数,即 J 和 h ,这些参数则构成一个聪明的物理学模型(能自我更新,无需靠大物理学家来定义);而这个模型的更新又是一个布朗运动的过程,服从朗之万动力学。所以神经网络的DNA本质在于物理学。

二、感知机学习的几何景观

接下来首先介绍感知机模型。这个模型当之无愧可称为人工智能的伊辛模型。它研究的是一群神经元如何实现对输入数据的分类,这从数学上可以表达为一个不等式 $w^T x \geq \kappa$,这里向量 w 是神经连接, x 为神经输入(例如,机器学习常用的MNIST数据集中每张手写体数字为784维实向量),而 κ 通常称为学习的稳定性指标(越大越稳定)。当 $\kappa=0, w_i = \pm 1$ 时,我们可以定义这样的玻尔兹曼统计系综:

$$P(w) = \frac{1}{Z} \prod_{\mu=1}^p \Theta\left(\frac{1}{\sqrt{N}} w^T x^\mu\right),$$

其中, P 代表分类图片总数, N 代表神经连接数目,而 Z 则为统计物理学中的配分函数。如果上面的不等式针对每个输入模式都能满足的话,则该 Z 显然具有构型数(解的数目)的特征,从而可定义自由熵: $S = \ln Z$ 。这个统计系综的设计归功于20世纪80年代一位杰出的年轻物理学家伊丽莎白·加德纳^[2],她考虑权重的分布而不是构型从而超越了霍费尔德模型的框架。因为数据的随机性,求解该熵并非易事,我们这里省去细节(感兴趣者可参阅教科书^[3])。1989年,法国物理学家马克·梅扎尔和他的博士生沃纳·克劳斯利用复本方法进行了计算,得出当 $\alpha = P/N \sim 0.833$ 时,自由熵消失(意味着该学习问题无解)。这是凝聚态物理理论(自旋玻璃)在计算机和统计学交叉学科的早期典型应用。非常奇妙的是,该结果于今年初被数学家完全严格证明^[4],而当今

高维随机统计预测在数学里是相当有生命力的一个分支。

该模型自从被提出以来伴随着不可协调的矛盾,因为长期以来在 $\alpha < 0.833$ 区间,解是存在的,但很多算法找不到它们,或者随维度升高,算法所能求解的最大 α 变小。这显示这个统计推断问题虽然定义上简单但从算法复杂度看高度非平庸!这个问题的解释要等到2013~2014年间两篇论文的问世^[5]。论文作者的出发点是解空间的几何结构,类似物理上构型空间的形态或者熵景观。解决一个难问题通常需要新思路!为了描绘熵景观,论文作者先从构型空间选取一个典型构型(物理上服从上述玻尔兹曼分布),然后在构型周围计数与选定参考构型存在一定汉明距离的构型(或者学习问题的解)。这在物理上等价于自旋玻璃理论的弗兰之-帕里西势能^[2,3]。通过复杂推导,作者惊奇地发现,在汉明距离很小的区间,自由熵为负数,哪怕是 α 非常靠近零。这从物理上意味着,该熵景观存在大量孤岛形态(犹如高尔夫球洞),这也解释了以往局域算法(如蒙特卡洛)求解的困难性。在松弛不等式的单向性的情况下,数学家近期已经给出了严格证明^[6,7]。他们在摘要中把这个物理结论称为Huang-Wong-Kabashima猜想。

一个重要问题的解决通常伴随新的重要问题的出现,这是科学研究最为迷人的地方。论文^[5]在展望中指出了有些特别设计的算法依然可在孤岛间找到解,这是跟孤岛熵景观格格不入的。这个新的重要问题看似非常难,但很快就被意大利物理学家理查德·泽奇纳及其合作者解决了^[8]。这个解决思路也十分巧妙,当然需要很深厚的数学和物理功力。既然自由熵为负,那么可以认为这可能是传统玻尔兹曼测度的结果,因此把自由熵当成随机变量,考虑其统计分布并且服从大偏差原理(即 $P(S) \sim e^{-Nr(S)}$,其中 $r(S)$ 称为率函数)。这么定义之后,理查德·泽奇纳等人发现,这个感知器的学习空间居然存在稀有的稠密解团簇!而且,那些高效的经验算法就是被这些解所吸引的,而完全避开了高尔夫球洞(实

际上它们是无法被找到的,掩藏于自由能深谷中)。而这一绝美的物理图像,同样于近期被数学家严格证明^[9]。至此,我们可以总结,虽然感知学习从数学形式上看非常简洁,但是从物理上可以获得直观且非常深刻的见解,并大部分结论能从数学上严格证明。从科学上去完全理解一个非平庸的命题应该也必须成为科学文化的一部分,而非一味盲从避开了模型只依赖于数据的现代机器学习方法。

这些研究始于一群喜欢跨学科的物理学家的好奇心,最后却激起数学家严格证明的欲望,让人们看到高维空间统计推断的优美。虽然大多物理学家考虑的问题带有随机性的成分(比如上述高斯随机输入数据),但是,在统计物理学的世界里,存在普适性这个重要的概念,或者说,在某些情况下可以被放心舍弃的细节依然不影响事物的本质。这或许是物理学思想的魅力,也是其他学科的科学家的或多或少难以理解之处。这些研究目前已经发展成一个更大的猜想,是否在深度学习乃至大语言模型的解空间里存在大偏差的稀有团簇?这些团簇或许能够实现举一反三的逻辑推理能力。

三、无师自通与对称性破缺

上一个例子讲述的是统计物理在理解监督学习的重要作用。接下来我们研究一下无监督学习,即无师自通。无监督学习是让机器从原始数据中自发地发现隐藏规律,类似人类在婴儿时期的观察和学习过程,所以是一种更为重要的认知方式。这个自然界最不可思议的是它的可理解性(爱因斯坦语录),所以人类可通过模型(几条合理性的假设)依靠逻辑演绎导出简洁的物理方程(如牛顿力学、广义相对论等),从而达到对成千上万种经验观察的高度压缩。这个与当前大语言模型所做的压缩即智能有很大的不同^③。那么,对于无监督学习,我们如何建模以直击其本质?

如图3所示, σ 代表输入原始数据(没有标签), $\xi^{1,2}$ 代表两层神经网络对数据规律的表示, x, y 分别为输出神经元。无监督学习从数学上可以表达为已知

数据推断连接 $\xi^{1,2}$ 的过程。为了建立理论模型,我们首先假定存在一个老师网络,它的连接是完全可知的,因此我们可以通过该老师网络来生成训练数据,这个规则叫受限玻尔兹曼机^④,图3中的*i,j,k,l*标示显层神经元,*x,y*是隐层神经元,因此如图的连接是个伊辛模型,显层与隐层神经元互为条件独立,因此便于蒙特卡洛模拟来生成数据。这样一来,那么具有相同结构的学生网络能否单从数据悟得老师的连接矩阵呢?这就是一个统计物理可研究的课题。

接下来,我们容易通过贝叶斯定理写出如下的学生网络的概率分布:

$$P(\xi^1, \xi^2 | \{\sigma^a\}_{a=1}^M) = \frac{\prod_a P(\sigma^a | \xi^1, \xi^2)}{\sum_{\xi^1, \xi^2} \prod_a P(\sigma^a | \xi^1, \xi^2)} P_0(\xi^1, \xi^2)$$

该分布在图3的具体表示为^[10]:

$$\frac{1}{\Omega} \prod_a \frac{1}{Z(\xi^1, \xi^2)} \cosh\left(\frac{\beta}{\sqrt{N}} \xi^1 \cdot \sigma^a\right) \cosh\left(\frac{\beta}{\sqrt{N}} \xi^2 \cdot \sigma^a\right) \prod_i [P_0(\xi^1, \xi^2)_i]$$

其中,*Z*为网络结构的配合函数,*N*为每个隐层神经元的神经连接数, β 为温度倒数, P_0 为先验分布, Ω 则为无监督学习的配分函数。在这里,我们做了两个重要假设:每个数据是独立生成的,并且先验分布对神经元标号是独立的。我们稍微观察以上的系综分布就可以发现, $\xi^{1,2} \rightarrow -\xi^{1,2}$ 和 $1 \leftrightarrow 2$,该分布是不变的,显示了 \mathbb{Z}_2 和 \mathbb{S}_2 对称性,因为我们的连接权重取为Ising自旋值。那么,一个有趣的物理问题就产生了:学习的过程是对称性破缺的过程吗?

经过复杂的计算(细节参看文献[3]),我们发现:随着数据量的增长达到第一个阈值,与 \mathbb{Z}_2 对称性相关的第一个连续性相变发生,学生开始推断老师连接权重相同的那部分(即 $\xi^1 = \xi^2$),这种类型的转变被称为自发对称破缺,就像在标准伊辛模型中遇到的铁磁相变那样。随着数据量进一步增加,学生开始推断老师连接权重不同的那部分(即 $\xi^1 = -\xi^2$),这被称为第一种置换(\mathbb{S}_2)对称破缺,即学生开始意识到它的两个感受野(ξ^1, ξ^2)也是不同的。不妨总结为“先求同,后存异”。随着数据量进

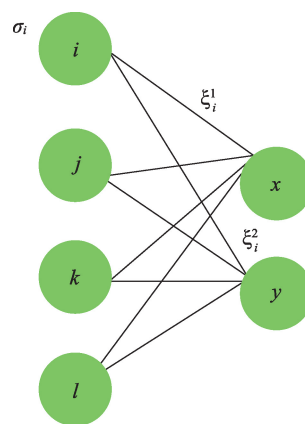


图3 受限玻尔兹曼机的学习过程示意图

一步增加,学生开始能够区分老师(或基本规律)体系结构中两个隐藏节点的内在顺序。我们将这个转变称为 \mathbb{S}_2 对称性破缺的第二个亚型。仅在此转变之后,自由能才有两个同等重要的谷底。但学生只推断其中一种可能性,并取决于初始条件。这两个谷底对应于基本规律的两种可能顺序(*x, y*)或(*y, x*),这也是原始无监督学习概率分布的内在置换对称性。因此,通过统计物理分析,我们得出的数据可以自发驱动层级式的连续相变直至数据中的客观规律被机器所捕获,并且也揭示了先验的作用^[10]:极大减少自发对称破缺的最小数据量,并且融合了两个亚型,即在先验的帮助下,学生认识自我和客观是同时发生的;然而在没有先验情况下,认识自我则先于客观。

从一个简单模型出发,我们可以揭示无监督学习丰富的物理图像,即对称性破缺是支配学习过程的重要力量。这种概念在今年又在非平衡的生成扩散过程中被完整诠释^[11],让人不得不感叹物理思维的巧妙与精确,再次印证了著名物理学家戴森那句名言。

四、非平衡稳态动力学的伪势表示法

前面两个例子并未涉及动力学,然而动力学是理解大脑认知的关键过程。我们注意到,在神经网络训练过程中,前面提到其本质为梯度力作用下的朗之万方程。事实上,在认知动力学层面上,几乎

所有的动力学并不存在梯度力,即下面方程

$$\frac{dx}{dt} = f(x) + \zeta,$$

其中 f 为不显含时间,但并不能写为某个标量势的梯度,即不存在李雅普诺夫函数。 ζ 表示神经回路的背景噪声(在这里暂且将其忽略)。在高维空间里,如上的动力学方程可以涌现出混沌行为,成为众多理论神经科学家(实际上多数为理论物理出身)展示数学物理功力的首选研究对象。在过去三十多年来,经典工作不断涌现,每次都加深了人们对于高维混沌动力学的理解。

图4给出了一个3维的例子:只有三个神经元的系统,在它们连接矩阵属性改变时,系统的相空间由一个全局稳定点破缺为对称的两个焦点。虽然原系统无法通过梯度力来研究(不存在李雅普诺夫函数),但是如果变换研究的兴趣为非平衡稳态(即零速率极限,即 $f=0$),那么我们就可以非常直观地写下一个动能函数 $E(x) = \frac{1}{2}f^2$ (单位质量)来作为非平衡稳态的伪势^[12],这个伪势将让我们能够定义正则系综来研究非平衡神经动力学稳态问题;这在此前的所有研究中是无法想象的。原则上,人们应该通过复杂的动力学平均场或路径积分来推导稳态方程,这对于更复杂的神经动力学(比如 f 的形式较复杂)来说甚至是十分艰巨的一个计算任务。

有了这个新思路,当三个神经元系统被推广至无穷($N \rightarrow \infty$)神经元系统(比如大脑具有大概860亿

级的神经细胞),并且假设相互作用矩阵是非厄米的随机矩阵[矩阵元 $J_{ij} \sim \mathcal{N}(0, g^2/N)$],我们可以发现当 g 增加到1时将触发一个连续的动力学相变(从有序走向混沌),其序参量为神经网络活动水平的涨落。注意这些序参量并不是人为设定的,而是来源于上面正则系综计算的逻辑演绎。让人惊讶的是,该计算还会导出另一个序参量,恰是统计力学中的响应函数,它刻画了动力系统在面对微弱扰动时的响应能力。我们发现在相变点附近,该响应函数出现峰值,从物理上证实了混沌边缘的优越性。无独有偶,2022年的一项实验研究表明人类大脑的脑电动力学在清醒时在混沌边缘具有最大的信息丰度^[13],从而暗示了统计力学推导的响应函数峰值可能从数学上讲是意识的必要条件。

这个例子告诉我们,即便是十分复杂的非梯度动力学,我们依然可以另辟蹊径从统计力学角度提出模型,并且通过严谨计算获得深刻认识。因此,年轻学生应该掌握必要的数学工具,并且勇于挑战既有思想框架,通过提供新的见解来发展古老的学科。

五、大语言模型示例泛化的奥妙

大语言模型是2023年初火遍全球的Chat GPT的原动力,它凭借海量数据文本和计算力通过预测下一个单词赢得了世人的赞叹和兴趣^[14]。经过预训练的聊天工具尤其展示了一种示例泛化的能力,

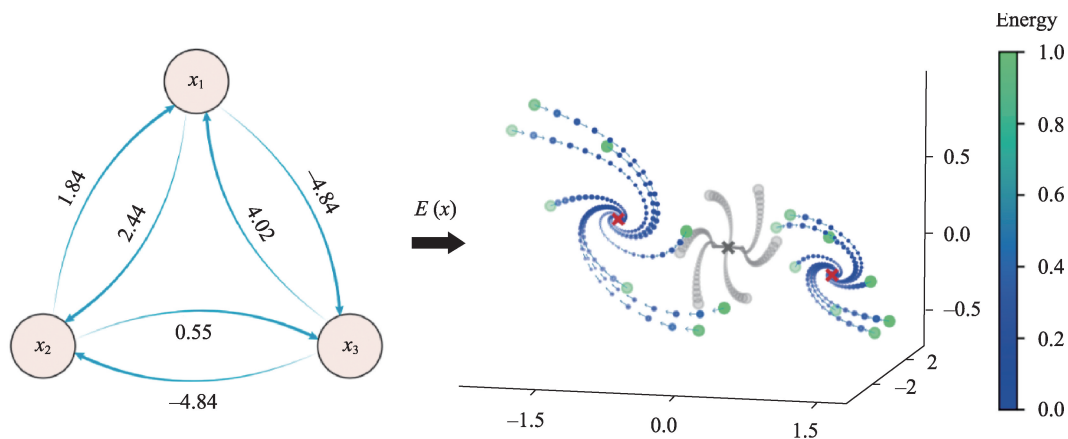


图4 非线性动力学的伪势法

这在以往所有机器模型中均未出现过。简单来说,就是给少数几个例子(不管是数学的,还是语言的),然后基础模型也不用再训练,它能够对新问题给出准确答案了。这困惑了大家好一阵子,直到下面的事情发生。

我们说过,任何一种复杂现象都需要模型驱动的研究,才能找到潜藏的简单规律(如果存在)。为了找到答案,我们先考虑一个线性回归函数类,如 $y = \mathbf{w}^T \mathbf{x}$ 。我们首先固定一个随机的任务向量 \mathbf{w} ,然后生成多个随机的 \mathbf{x} 计算其标签 y ,就有了针对示例泛化的预训练数据:

$$\mathbf{X}^\mu = \begin{bmatrix} \mathbf{x}_1^\mu & \mathbf{x}_2^\mu & \cdots & \mathbf{x}_n^\mu & \mathbf{x}^\mu \\ y_1^\mu & y_2^\mu & \cdots & y_n^\mu & 0 \end{bmatrix},$$

注意最后一列为让基础模型推断的数据,故遮掉了真实的标签。这个矩阵相当于给了有答案的 n 道数学题,然后问一道(最后一列),看机器能否准确推断。这显然是一个难的问题!但是,神奇的是,聪明的机器做到了,不禁让人看到通用人工智能微弱的曙光。

其实通过简单变换,比如假设单层线性自注意力机制(细节见文献[15]),我们喜出望外地发现预训练的机器参数服从如下的哈密顿量:

$$\mathcal{H}(\boldsymbol{\sigma}) = -\sum_{i \neq j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i + \frac{1}{2} \sum_i \lambda_i \sigma_i^2,$$

最后一项为机器参数的高斯先验。这显然是一个两体相互作用的实自旋模型,它的基态就是基础模型示例泛化能力的根源。我们可以通过高斯分布假设来求解这个模型的基态,最后发现哪怕在有限尺寸的网络,依然可以得到如下最优解:

$$\hat{y} = \mathbf{w}^T (\mathbf{W}_{11} + \mathbf{w} \mathbf{W}_{21}) \tilde{\mathbf{x}}$$

其中 \hat{y} 为神经网络给出的答案,上述物理模型的基态意味着 $\mathbf{W}_{11} = \mathbb{I}_D$ 和 $\mathbf{W}_{21} = 0$, D 是数据的维度。分块矩阵 \mathbf{W} 与自旋耦合 \mathbf{J} 一一对应。因此我们就明白了,只要找到该基态,示例泛化即可达成,并无需再微调参数!这个模型还揭示了任务向量的多样性对预训练效果起到至关重要的作用。因此,也就不难理解大语言模型需要海量多模态的文本库

了。我们可以大胆地想象,只要穷尽承载人类文明的所有知识,也许有一天我们真能制造出无所不能的智慧机器,至少在人类已掌握技能的疆域内是没有问题的。

六、总结和展望

本文从物理学的概念出发介绍了神经网络的DNA,数据相当于一种初始化,可以驱动网络连接权重的连续更新以获得一个聪明的自适应的物理模型,而这个更新过程是端对端地优化一个目标函数,优化的过程即执行在高维空间的朗之万动力学。神经网络的奥秘正是在于高维的权重空间,它本质上服从正则系综分布。半严格的物理分析给出了权重空间的分布和数据驱动的权重的对称性破缺。从物理直观出发,人们可以获取非平衡神经动力学的稳态全貌以及隐藏的动力学相变;甚至,人们可以将大语言模型的示例泛化归结为两体自旋模型,依此可以洞察智能的本质。数学的具象化为物理,而物理的尽头则为数学,数学与物理相辅相成,成为理解神经网络乃至智能本质不可或缺的手段。本文借助少数几个例子,希望启发青年学生欣赏数学的魅力,习得物理的洞察力,为揭开大脑智能神秘的面纱贡献自己的智慧。

参考文献

- ① 神经网络的源头可追溯至麦可洛和皮茨在1943年发表的关于逻辑演算的研究。
- ② 戴森应邀为美国数学学会的爱因斯坦讲座所准备,题目为鸟与青蛙。
- ③ 大模型的压缩并不意味着“理解”。
- ④ 对受限玻尔兹曼机训练算法的研究是2024年诺贝尔物理学奖得主辛顿的重要贡献之一。

参考文献

- [1] Ising, E. Beitrag zur Theorie des Ferromagnetismus. *Z. Physik*, 1925, 31: 253-258.
- [2] 黄海平, 统计物理、无序系统与神经网络, 科学, 2022, 74: 40-44
- [3] 黄海平, 神经网络的统计力学(英文版), 高等教育出版社, 2021.
- [4] Brice Huang, Capacity threshold for the Ising perceptron, arXiv:

- 2404.18902.
- [5] H. Huang, K. M. Wong and Y. Kabashima, Entropy landscape of solutions in the binary perceptron problem, *Journal of Physics A: Mathematical and Theoretical*, 2013, 46(37):375002; H.Huang and Y. Kabashima, Origin of the computational hardness for learning with binary synapses, *Physical Review E*, 2014, 90(5):052813.
- [6] W. Perkins and C. Xu, Frozen 1-rsb structure of the symmetric ising perceptron, *Random Structures & Algorithms*, 2024, 64(4):856.
- [7] E. Abbe, S. Li and A. Sly, Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron, In 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), pp. 327-338.
- [8] C. Baldassi, A. Ingrosso, C. Lucibello, L. Saglietti and R. Zecchina, Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses, *Physical review letters*, 2015, 115(12):128101.
- [9] E. Abbe, S. Li and A. Sly, Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster, In 2022 Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, pp. 860-873.
- [10] Tianqi Hou and Haiping Huang, Statistical physics of unsupervised learning with prior knowledge in neural networks. *Phys. Rev. Lett.*, 2020, 124:248302.
- [11] Gabriel Raya and Lca Ambrogioni. Spontaneous symmetry breaking in generative diffusion models. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 66377-66389. Curran Associates, Inc., 2023; Z. Y and H. Huang, Nonequilibrium physics of generative diffusion models, arXiv: 2405.11932.
- [12] Junbin Qiu and Haiping Huang. An optimization-based equilibrium measure describes non-equilibrium steady state dynamics: application to edge of chaos. arXiv:2401.10009.
- [13] Daniel Toker, Ioannis Pappas, Janna D Lendner, Joel Frohlich, Diego M Mateos, Suresh Muthukumaraswamy, Robin Carhart-Harris, Michelle Paff, Paul M Vespa, Martin M Monti, et al. Consciousness is supported by near-critical slow cortical electrodynamics. *Proceedings of the National Academy of Sciences*, 2022, 119(7):e2024455119.
- [14] Sebastien Bubeck, Varun Chandrasekaran, Ronen Eldan, John A. Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanfang Li, Scott M. Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv:2303.12712.
- [15] Yuhao Li, Ruoran Bai, and Haiping Huang, Spin glass model of in-context learning, arXiv: 2408.02288.



科学家发现亨廷顿氏症导致的分子破坏

荷兰乌得勒支大学(Utrecht University)的科学家发现,由突变的亨廷顿蛋白(huntingtin protein)聚集而成的有毒蛋白会破坏核膜,导致神经元中的DNA损伤和基因失调。这一突破可能将类似的机制与其他神经退行性疾病联系起来,揭示了神经元损伤的共同途径,并为治疗干预提供了潜在的新靶点。

亨廷顿氏病(Huntington's disease)是一种毁灭性的神经组织生长异常病变,由HTT基因突变引起,会使细胞产生异常大量的亨廷顿蛋白,它们聚集在细胞内并破坏细胞。但是,其导致神经细胞死亡的机制尚不清楚。

核膜(nuclear envelope)是保护和调节细胞核内染色体的屏障,允许其根据需要打开、关闭基因。研究人员发现,亨廷顿蛋白破坏了核膜底层的蛋白质网,使核膜更容易破裂。研究人员通过专业技术看到了

微小的原纤维从聚集体(图中绿色部分)中伸出,穿过核膜下面的网状结构(图中洋红色部分)。这可能损害细胞在核膜破裂后重新密封的能力,严重损害了核膜的屏障功能,随着时间推移,还可能导致细胞DNA的损伤和神经元基因的失调。

包括某些类型的肌萎缩性侧索硬化症(amyotrophic lateral sclerosis)、额颞叶痴呆(frontotemporal dementia)在内的其他几种神经变性疾病,都与细胞核内蛋白聚集体的形成有关。研究人员推测,核聚集体诱导的核膜破裂,是神经退行性病变的一个共同因素,它引发了一系列失控过程,最终导致神经元死亡和神经炎症。

(高凌云编译自2024年8月16日SciTechDaily网站)