

机器学习的量子动力学

王鹏[†] 麦麦提尼亚孜·麦麦提阿卜杜拉

(西南民族大学计算机科学与工程学院, 成都 610225)

(2024 年 7 月 17 日收到; 2024 年 11 月 27 日收到修改稿)

基于第一性原理思想, 采用量子动力学方法对机器学习的迭代运动过程进行建模. 在机器学习的参数空间定义广义目标函数, 利用 Schrödinger 方程和势能等效得到机器学习过程的量子动力学方程, 通过 Wick 转动进一步建立了量子动力学与热动力学的关系, 这为利用物理理论和数学理论对机器学习的迭代过程进行研究提供了可能. 本文工作将机器学习的迭代过程转化为含时偏微分方程来进行精确数学表述, 该方程表明机器学习过程可能存在多尺度的退火过程和同一尺度下的时间演化过程. 利用量子动力学方程证明了机器学习在时间演化时的收敛性, 解释了机器学习中的扩散模型是量子动力学方程在经典近似和低阶泰勒近似下的映射模型, 导出了人工智能中常用的 Softmax 和 Sigmoid 函数. 这些结果表明量子动力学方法在研究机器学习理论中是有效的.

关键词: 量子动力学, 机器学习, 扩散模型, 薛定谔方程

PACS: 07.05.Mh, 03.65.-w, 05.40.-a, 02.30.Jr

DOI: 10.7498/aps.74.20240999

CSTR: 32037.14.aps.74.20240999

1 引言

机器学习是一个典型的优化问题, 其学习过程就是在参数空间中的迭代寻优过程, 将这种算法的迭代运动过程视为动力学过程是一种很自然的思路. 经过长期的发展, 动力学理论体系已十分完备, 有量子动力学、牛顿动力学、热动力学、电动力学、分子动力学, 这些动力学通过建立一组动力学方程对运动规律进行理论描述. 建立机器学习的动力学理论有望解决这一领域缺乏理论模型的问题, 从而推进机器学习理论和应用的发展.

针对优化问题采用动力学的方法进行理论建模最早始于 Metropolis 等^[1]在 1953 年基于热力学提出的解的接收准则, 后来被 Kirkpatrick 等^[2]应用于求解优化问题模拟退火算法, 这是将热动力学应用于优化问题的早期尝试. 1994 年 Finnila 等^[3]

提出将目标函数视为 Schrödinger 方程中的势能, 从而把优化问题转化为约束态量子基态波函数问题, 这是量子动力学理论在优化问题中的首次应用. 2013 年我们开始尝试提出优化问题的量子动力学框架, 研究结果表明 Schrödinger 方程可以有效地对优化算法的基本迭代过程进行描述^[4-7]. 量子动力学在人工智能中能成功应用的证据还包括最早实现商业应用的量子计算机 D-wave, 它利用量子退火实现了对智能优化问题的成功求解^[8].

2015 年 Sohl-Dickstein 等^[9]利用非平衡热力学思想提出了扩散概率模型 (diffusion probabilistic model). 随后, 扩散模型 (diffusion model) 得到了快速的发展, 并被广泛应用在人工智能领域. 扩散模型是动力学理论在机器学习领域成功应用的范例. 近年来动力学理论在机器学习中展现出越来越大的理论和应用价值, 并发展出基于随机微分方程的新型动力学方法^[10], 其有效性不断被实验结果证实.

[†] 通信作者. E-mail: wp002005@163.com

机器学习的动力学理论最重要的是要建立机器学习过程的动力学方程. 由于机器学习过程是一个概率运动过程, 而量子动力学和热动力学分别从微观和宏观两个角度描述了物质世界的概率运动过程, 特别是量子动力学反映了物质世界最基本的运动规律. 量子动力学的核心动力学方程是 Schrödinger 方程, 其利用一个确定性的含时偏微分方程描述了物质世界广泛存在的概率运动规律. 这表明量子动力学可以作为描述机器学习运动过程的第一性原理. 本文通过建立机器学习的量子动力学方程, 从量子动力学角度出发展开对机器学习迭代过程的研究.

2 优化问题的量子动力学框架

优化问题的量子动力学框架 (quantum dynamics framework, QDF) 最早源于 2013 年提出的多尺度量子谐振子算法 (multi-scale quantum harmonic oscillator algorithm, MQHOA)^[6]. 随后

我们对其进行了持续的研究逐步形成了优化问题的量子动力学框架, 该框架已在函数优化理论和实践中都得到了成功的应用^[11,12].

QDF 的基本结构如图 1 所示. 量子力学的基本动力学方程为 Schrödinger 方程, 在研究优化问题时可以不考虑物理学中 Schrödinger 方程中的物理量和常数, 则 Schrödinger 方程可以简化为

$$i\frac{\partial\psi(x,t)}{\partial t} = \left[-D\frac{\partial^2}{\partial x^2} + V(x)\right]\psi(x,t), \quad (1)$$

其中 D 为常数, 该常数影响系统的动能大小; $V(x)$ 为约束势能; $\psi(x,t)$ 为粒子的波函数.

QDF 的基本思想是将优化问题的迭代过程视为量子动力学过程, 将优化问题中的目标函数 $f(x)$ 视为 Schrödinger 方程中的势能项 $V(x)$, 即 $V(x) = f(x)$, 从而将优化问题转化为求解量子约束态的基态波函数问题, 相应的优化问题的量子动力学方程如下:

$$i\frac{\partial\psi(x,t)}{\partial t} = \left[-D\frac{\partial^2}{\partial x^2} + f(x)\right]\psi(x,t). \quad (2)$$

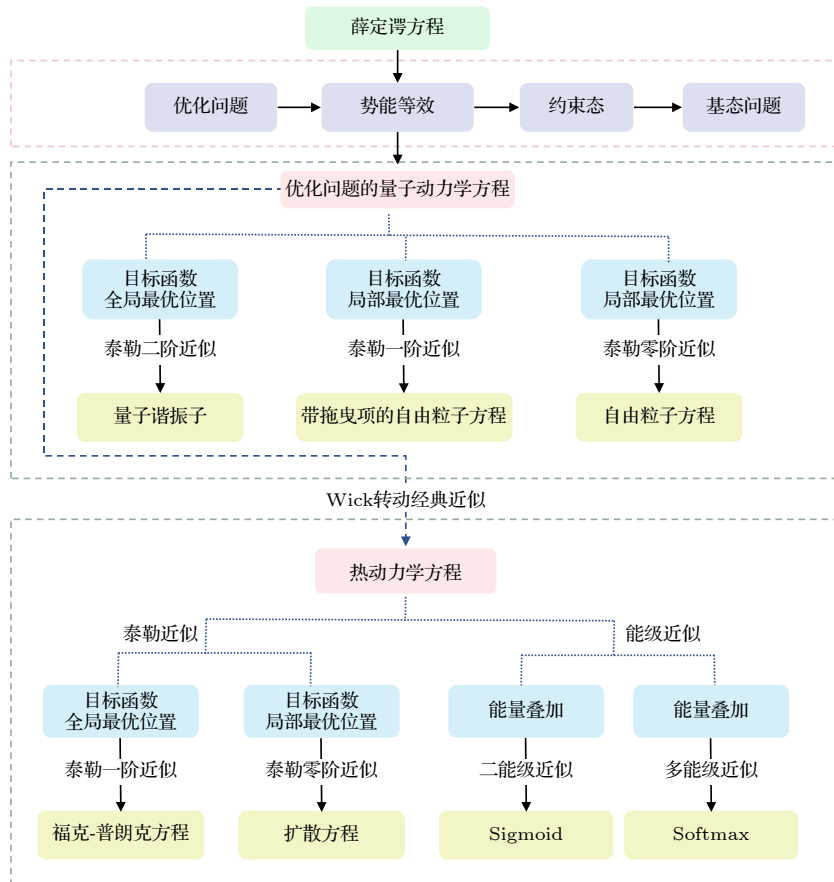


图 1 优化问题的量子动力学框架

Fig. 1. Quantum dynamical framework for optimization problems.

通过量子动力学方程, 优化问题的迭代过程转化为含时偏微分动力学方程进行描述. 其中波函数 $\psi(x, t)$ 的时间演化过程就是优化问题解的概率分布的演化过程.

3 机器学习的量子动力学

3.1 机器学习量子动力学方程

机器学习过程也是对由大量连接权重构成的参数空间进行参数优化的过程, 其优化目标是为了寻找一个最优的连接权重组合. 通常不能直接获得参数空间优化时的目标函数, 为了和函数优化问题中的目标函数 $f(x)$ 区别, 这里用广义目标函数 $\zeta(x)$ 来形式化地表示任意问题在神经网络参数空间中目标函数的数学关系. 如果将参数空间中的广义目标函数视为 Schrödinger 方程中的势能项, 即进行势能等效:

$$V(x) = \zeta(x), \quad (3)$$

则薛定谔方程转化为如下方程:

$$i \frac{\partial \psi(x, t)}{\partial t} = \left[-D \frac{\partial^2}{\partial x^2} + \zeta(x) \right] \psi(x, t). \quad (4)$$

这个方程就是机器学习的量子动力学方程, 该方程的物理含义为被势能 $\zeta(x)$ 约束的量子束缚态系统. 根据玻恩对波函数的概率解释, 波函数 $\psi(x, t)$ 的模方 $|\psi(x, t)|^2$ 的意义为 t 时刻解的概率分布. 机器学习的过程就是波函数 $\psi(x, t)$ 的时间演化过程.

系统的哈密顿算符为

$$\hat{H} = -D \frac{\partial^2}{\partial x^2} + \zeta(x), \quad (5)$$

其中参数 D 决定系统动能的大小, 通过缓慢降低参数 D 的值可以实现系统的退火. 因此, 根据机器学习的量子动力学方程可以预测, 机器学习的迭代过程可能包含两个主要的迭代循环: 第一个迭代循环是减小 D 值的量子退火迭代过程, 对应于在参数空间针对广义目标函数的 $\zeta(x)$ 采样步长的逐步减小. D 值大时, 系统动能大, 系统在基态时的零点能 E_0 也大, 为了获得精确的全局最优解, 需要逐步退火降低 D 值, 最后获得较低的零点能. 第二个迭代循环是同一动能下系统向基态的时间演化过程.

建立机器学习量子动力学方程的重要意义在于以下两点: 第一, 从物理意义上将机器学习对参

数的优化学习过程转化为了求解以广义目标函数 $\zeta(x)$ 为约束势能的量子约束态基态波函数问题; 第二, 将机器学习的迭代过程从数学上转化为严格的含时偏微分方程来表述. 由于量子物理经过一百多年的发展已具有完备且丰富的物理和数学理论体系, 可以为研究机器学习提供支持.

3.2 广义目标函数的 Taylor 近似

由于量子动力学方程的解析求解非常困难, 同时无法获得机器学习参数空间广义目标函数 $\zeta(x)$ 解析式, 因此采用针对广义目标函数 Taylor 展开的方式来对量子动力学方程进行近似分析.

3.2.1. 广义目标函数的 0 阶 Taylor 近似

参数空间的广义目标函数 $\zeta(x)$ 在当前位置 x_0 的 Taylor 零阶近似为 $\zeta(x) = 0$, 这种情况表明对广义目标函数 $\zeta(x)$ 的当前位置 x_0 的情况一无所知. 因此, 机器学习的量子动力学方程在广义目标函数 $\zeta(x)$ 的 0 阶 Taylor 近似就是自由电子方程:

$$i \frac{\partial \psi(x, t)}{\partial t} = -D \frac{\partial^2 \psi(x, t)}{\partial x^2}. \quad (6)$$

自由电子的运动未受到任何势能的约束, 所以自由电子方程中没有势能项 $\zeta(x)$. 自由电子方程相比经典的扩散方程只是多了一个表达波动性的虚数符号 i , 它描述了机器学习过程中对广义目标函数 $\zeta(x)$ 的信息完全未知时算法的量子动力学规律.

根据量子力学中自由电子的运动理论, 0 阶 Taylor 近似下机器学习的量子动力学过程表现为波包的色散过程. 色散过程是由于自由粒子的波函数由许多不同频率成分的态叠加组成的波包, 这些不同频率的成分随时间的演化会逐步弥散开来 (如图 2). 这一过程与扩散过程相似, 但却是更复杂的波包色散过程, 这正是量子动力学系统存在波粒二象性所造成的.

3.2.2. 广义目标函数的 1 阶 Taylor 近似

参数空间的广义目标函数 $\zeta(x)$ 在当前位置 x_0 的 1 阶 Taylor 近似为 $\zeta(x) \approx \frac{\partial \zeta(x_0)}{\partial x}$, 此时机器学习的量子动力学方程为

$$i \frac{\partial \psi(x, t)}{\partial t} = \left(-D \frac{\partial^2}{\partial x^2} + \frac{\partial \zeta(x_0)}{\partial x} \right) \psi(x, t). \quad (7)$$

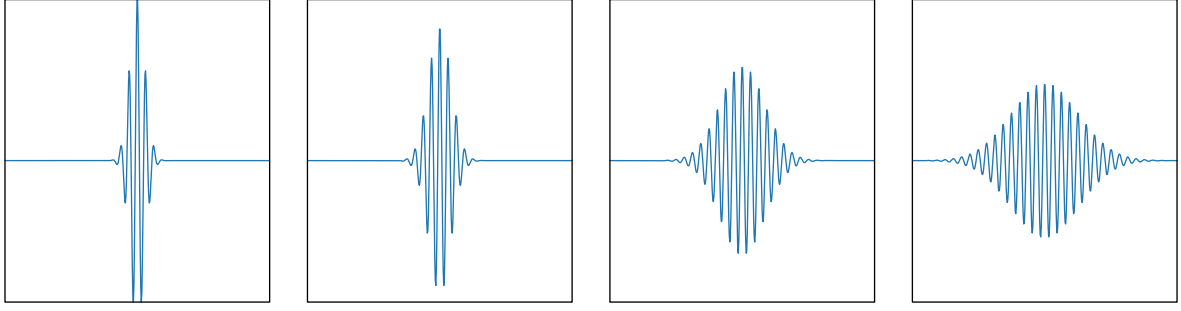


图 2 波包色散过程

Fig. 2. Process of wave packet dispersion.

这一方程表示电子的波包移动到 x_0 位置时, 该位置约束势能 $\zeta(x)$ 的斜率为 $\partial\zeta(x_0)/\partial x$, 这一动力学过程对应于算法在参数空间通过两次采样来估计广义目标函数在 x_0 位置的斜率, 从而进行梯度下降的优化操作.

3.3 机器学习量子动力学方程的经典近似

量子动力学中的波包动力学过程在经典计算机中非常难以实现, 为了在经典计算机上近似实现量子动力学过程, 这里将机器学习的量子动力学方程进行 Wick 转动^[13]. 令虚时间 $\tau = it$, 则量子动力学方程将失去波动特性被变换为虚时间下经典的扩散反应方程:

$$\frac{\partial\psi(x, \tau)}{\partial\tau} = \left(D \frac{\partial^2}{\partial x^2} - \zeta(x)\right)\psi(x, \tau). \quad (8)$$

Wick 转动后广义目标函数 $\zeta(x)$ 的 0 阶 Taylor 近似方程就由量子化的自由电子方程变换为标准的经典扩散方程:

$$\frac{\partial\psi(x, \tau)}{\partial\tau} = D \frac{\partial^2\psi(x, \tau)}{\partial x^2}. \quad (9)$$

算法的动力学过程从波包色散退化为经典扩散过程 (如图 3). 机器学习算法在参数空间中的采样可以采用点源的扩散过程来描述, 即扩散方程的初始条件 $\tau = 0$ 时 $\psi(x, 0) = \delta(x)$, $\delta(x)$ 为狄拉克函数. 这一初始条件对应的解为格林函数, 一维扩散方程的格林函数为高斯函数:

$$\psi(x, \tau) = \frac{1}{(4\pi D\tau)^{3/2}} e^{-\frac{x^2}{4D\tau}}. \quad (10)$$

格林函数物理意义是描述一个点源的扩散过程, 即一个位于原点 0 的点在 τ 时刻扩散到 x 位置的概率. 优化问题中解空间内的采样行为可以采用点源的扩散来描述. 经典的扩散过程在函数优化问题中可采用群体高斯随机行走采样的方法来实现.

这里参数 D 为扩散系数, D 值越大系统动能越大, 对应的高斯采样的标准差越大.

从扩散方程可以得出: 如果对广义目标函数 $\zeta(x)$ 的信息完全未知, 机器学习的基本动力学过程就是在参数空间中进行的高斯随机扩散过程.

Wick 转动后广义目标函数 $\zeta(x)$ 的 1 阶 Taylor 近似为 Fokker-Planck 方程:

$$\frac{\partial\psi(x, \tau)}{\partial\tau} = \left(D \frac{\partial^2}{\partial x^2} - \frac{\partial\zeta(x)}{\partial x}\right)\psi(x, \tau), \quad (11)$$

其中拖拽项 $\partial\zeta(x)/\partial x$ 表达的是在机器学习的迭代过程中获得广义目标函数 $\zeta(x)$ 当前采样位置的导数. 与扩散方程描述的对广义目标函数 $\zeta(x)$ 完全未知的情况相比, Fokker-Planck 方程描述了迭代过程中可以获得当前目标函数采样位置导数并进行梯度下降时的动力学规律. 机器学习中的 Fokker-Planck 方程描述了学习过程中基于高斯采样扩散的梯度下降过程. 在机器学习迭代过程中针对一个样本的学习迭代可以被映射为参数空间中针对广义目标函数 $\zeta(x)$ 的一次梯度下降操作.

3.4 Fokker-Planck 方程的离散化

Fokker-Planck 方程中的时间变量 τ 和空间变量 x 均为连续变量, 但机器学习的迭代过程是一个离散的动力学过程, 因此采用 Fokker-Planck 方程的离散化形式更能准确地描述机器学习的迭代过程.

采用 ψ_i^n 表示在时刻 $\tau = \tau_n$ 和位置 $x = x_i$ 处方程的解, 机器学习两次迭代过程的时刻分别为 τ_n 和 τ_{n+1} , 则时间间隔 $\Delta\tau = \tau_{n+1} - \tau_n$; 两次迭代在参数空间中的位置分别为 x_i 和 x_{i+1} , 空间位置间隔 $\Delta x = x_{i+1} - x_i$, 对应于参数空间广义目标函数值分别为 $\zeta(x_i)$ 和 $\zeta(x_{i+1})$, 广义目标函数值的差 $\Delta\zeta = \zeta(x_{i+1}) - \zeta(x_i)$.

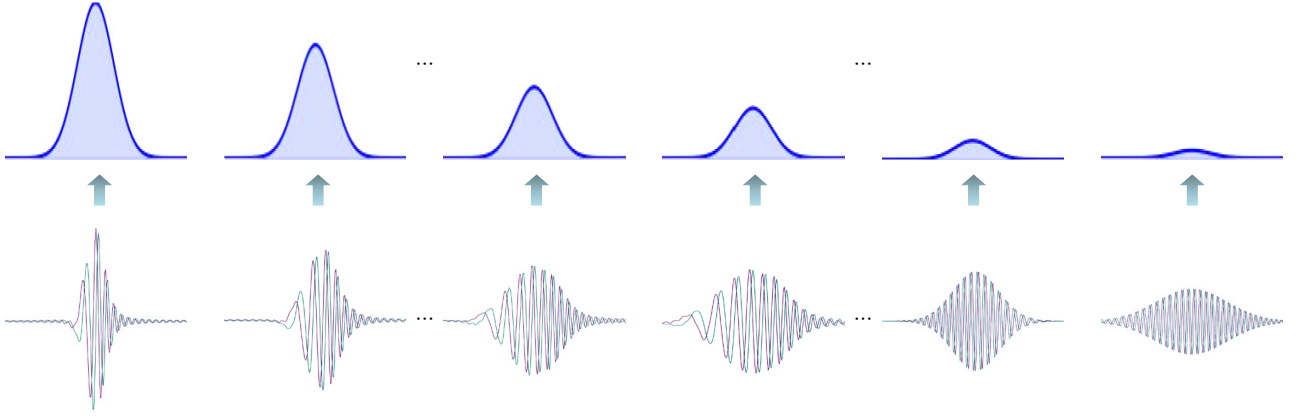


图 3 波包色散到经典扩散的转化

Fig. 3. Transition from wave packet dispersion to classical diffusion.

根据 Fokker-Planck 方程 (11), 对时间的微分可以离散化为

$$\frac{\partial \psi(x, \tau)}{\partial \tau} \approx \frac{\psi_i^{n+1} - \psi_i^n}{\Delta \tau}. \quad (12)$$

方程 (11) 右边的动能项对位置的二阶微分可以离散化为

$$\frac{\partial^2 \psi(x, \tau)}{\partial x^2} \approx \frac{\psi_{i+1}^n - 2\psi_i^n + \psi_{i-1}^n}{(\Delta x)^2}. \quad (13)$$

方程 (11) 右边的广义目标函数 $\zeta(x)$ 的 Taylor 二阶近似离散化为

$$\frac{\partial \zeta(x)}{\partial x} \psi(x, \tau) \approx \frac{\zeta(x_{i+1}) - \zeta(x_i)}{x_{i+1} - x_i} \psi_i^n = \frac{\Delta \zeta}{\Delta x} \psi_i^n. \quad (14)$$

由于在参数空间中 $\zeta(x)$ 的导数无法直接定义, $\partial \zeta(x)/\partial x$ 的离散形式表明可以通过在参数空间的两次采样来近似获得广义目标函数 $\zeta(x)$ 的梯度信息. 根据 Fokker-Planck 方程两边相等的关系和上面的离散化表达式可得 Fokker-Planck 方程的离散化形式:

$$\frac{\psi_i^{n+1} - \psi_i^n}{\Delta \tau} = D \frac{\psi_{i+1}^n - 2\psi_i^n + \psi_{i-1}^n}{(\Delta x)^2} - \frac{\Delta \zeta}{\Delta x} \psi_i^n. \quad (15)$$

变形后可得到从 τ_n 时刻到 τ_{n+1} 时刻的离散时间演化迭代公式:

$$\psi_i^{n+1} = \psi_i^n + D \frac{(\psi_{i+1}^n - 2\psi_i^n + \psi_{i-1}^n) \Delta \tau}{(\Delta x)^2} - \frac{\Delta \zeta \Delta \tau}{\Delta x} \psi_i^n. \quad (16)$$

4 机器学习量子动力学理论的应用

4.1 机器学习的收敛性分析

机器学习的量子动力学方程的建立为研究机

器学习的迭代过程提供了可靠的理论支持, 利用这一理论可以分析和解释机器学习的一些核心操作和算法. 首先采用量子动力学方程对机器学习迭代过程的收敛性进行分析.

机器学习的迭代过程是一种时间演化过程, 机器学习的量子动力学方程的含时通解可以写成如下形式:

$$\psi(x, t) = \sum_n c_n \varphi_n(x) \exp(-iE_n t). \quad (17)$$

通解公式表明量子动力学方程的解是由一系列不同能级的态按概率叠加形成的叠加态. 对这一叠加态进行测量时, 叠加态将按概率坍缩到其中的一个态. 通解中的能量 E_n 对应于系统的动能 E_d 加上势能 E_p :

$$E_n = E_d + E_p. \quad (18)$$

量子动力学方程进行 Wick 转动后经典扩散反应方程的通解为

$$\psi(x, \tau) = \sum_n c_n \varphi_n(x) \exp(-E_n \tau). \quad (19)$$

求最小值的优化问题, 就可获得基态能量 E_0 .

Wick 转动后通解的指数部分会导致各能量对应状态随 τ 的演化指数衰减, 能量值 E_n 越大衰减越快, 其中基态能量 E_0 对应的项会以较大的概率被保留下来, 也就是说优解会被大概率保留下来. 当 $\tau \rightarrow \infty$ 时, 可以得到

$$\lim_{\tau \rightarrow \infty} \psi(x, \tau) = \lim_{\tau \rightarrow \infty} \sum_n c_n \varphi_n(x) \exp(-E_n \tau) \approx c_0 \varphi_0(x) \exp(-E_0 \tau), \quad (20)$$

其中 $c_0 \varphi_0(x) \exp(-E_0 \tau)$ 为能量最低的状态.

这一结果表明机器学习的迭代过程在 $\tau \rightarrow \infty$ 时是可以收敛到基态的, 即全局最优解. 这一结果也表明对机器学习的量子动力学方程进行 Wick 转动后, 虽然丢失了波动性, 但仍然能有效地表达算法向基态的演化收敛特性. 热动力学的时间演化过程是一个针对系统能量的低通滤波器, 低能量的状态会以较大概率被保留下来.

根据这一时间演化的收敛性和动力学方程中与动能相关的系数 D 的物理意义可以得出机器学习的动力学过程应包含以下两个基本迭代过程: 第一, 不断减小系数 D 的值以逐步降低系统的动能的退火过程, 这一过程在实际中对应于采样步长的逐步减小, 从而获得更低的基态能量 E_0 , 这意味着机器学习算法应具有某种多尺度的学习过程; 第二, 在同一动能下向系统基态的时间演化过程, 这一迭代演化过程在采用 Softmax 或 Sigmoid 准则的基础上进行梯度下降操作.

4.2 Softmax 和 Sigmoid 的导出

同样, 利用经典扩散反应方程的通解公式可以得到机器学习中被广泛使用的 Softmax 和 Sigmoid 函数.

假设量子动力学方程进行 Wick 转动后经典扩散反应方程的通解是由 n 个能级的状态构成:

$$\psi(x, \tau) = \sum_{i=1}^n c_i \varphi_i(x) \exp(-E_i \tau), \quad (21)$$

其中第 k 个能级 τ 时刻能量对应的状态为 $c_k \varphi_k(x) \times \exp(-E_k \tau)$.

算法演化的初始时刻 $\tau = 0$, 此时系统状态为

$$\psi(x, 0) = \sum_{i=1}^n c_i \varphi_i(x). \quad (22)$$

由于时间 τ 在指数位置, 随时间的演化, 每一项的大小主要由指数项决定, 同时能量越大的状态随时间演化会衰减得越快. 当算法进行了长时间演化后, 指数项将占主导地位, 因此可以忽略非指数项的影响, 能量为 E_k 的态出现的概率 P_{E_k} 可以近似表示为

$$P_{E_k} \approx \frac{\exp(-E_k \tau)}{\sum_{i=1}^n \exp(-E_i \tau)}. \quad (23)$$

(23) 式即为含时的 Softmax 函数, Softmax 会随时间 τ 进行动力学演化, 当 $\tau = 0$ 时各种状态的概率

相同; 逐渐增大时间 τ , 则低能量的状态出现的概率将增大; 当 $\tau \rightarrow \infty$ 时, 系统将趋近于能量最低的 E_0 状态. 在这里 Softmax 实现了对系统能量的低通滤波, 以较大的概率保留了能量较小状态.

如果将系统能量进一步减少为两个, 即对系统进行二能级近似, 分别用能量 E_a 和 E_b 来表示. 假设 $E_a < E_b$, E_a 为低能态对应优解, E_b 为高能态对应差解, 则低能态 E_a 出现的概率 P_{E_a} 为

$$\begin{aligned} P_{E_a} &\approx \frac{\exp(-E_a \tau)}{\exp(-E_a \tau) + \exp(-E_b \tau)} \\ &= 1/[1 + \exp(-(E_b - E_a)\tau)] \\ &= 1/[1 + \exp(-\Delta E \tau)]. \end{aligned} \quad (24)$$

从 (24) 式的形式看, 在二能级近似后的概率函数就是含时的 Sigmoid 函数, Sigmoid 函数的形状会随时间的演化发生改变. ΔE 相同时, 随着时间 τ 的演化, 低能量状态 E_a 的选择概率会增大; 当 $\tau \rightarrow \infty$ 时, 低能量状态 E_a 的选择概率 P_{E_a} 会趋近于 1.

所以根据二能级近似, 算法在迭代演化的过程中会逐步降低对差解的接受概率. 当 $\tau > 0$ 且 $\Delta E > 0$ 时表明算法采样到了优解, 对差解的接收概率是按相同的规律演化.

令 $\Delta E = x$, 则上述概率式就是机器学习中常见的带时间演化过程的 Sigmoid 激活函数:

$$S(x) = 1/(1 + e^{-x\tau}). \quad (25)$$

图 4 给出了不同时间 τ 下 Sigmoid 函数 $S(x)$ 的演化图像.

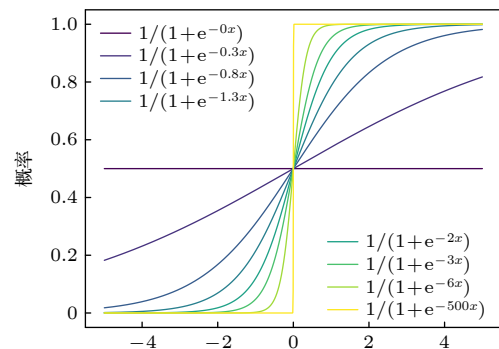


图 4 Sigmoid 函数随时间的演化

Fig. 4. Evolution of the Sigmoid function over time.

4.3 扩散模型的量子动力学解释

扩散模型是 2015 年提出的一种基于热力学理论的机器学习算法, 它在生成式人工智能等多个

应用领域展现出了良好的性能和应用前景^[14-17]. 扩散模型采用对图像进行逐级添加高斯噪声后逐级对噪声图像进行去噪的高斯扩散方法来实现网络参数的学习.

针对量子动力学和扩散过程的关系早在 1975 年 Anderson^[18] 就首次根据 Schrödinger 方程和扩散方程的同构性提出了采用随机行走 (random walk) 求解分子基态波函数的扩散蒙特卡罗 (diffusion Monte Carlo, DMC) 方法. DMC 被认为是一种获得量子系统基态能量和波函数的标准方法. 如果把一个优化问题转化为量子问题就可以采用 DMC 方法计算基态波函数^[19]. DMC 利用了 Schrödinger 方程和扩散方程之间的同构性 (isomorph), 模拟扩

散过程推动波函数逐步向基态演化, 从而实现对目标函数全局最优解的搜索^[20].

根据对机器学习量子动力学方程的分析, 经典近似后广义目标函数的 0 阶 Taylor 近似要求机器学习算法在参数空间对广义目标函数 $\zeta(x)$ 进行高斯采样, 1 阶 Taylor 近似要求机器学习算法在参数空间对广义目标函数 $\zeta(x)$ 进行两次采样来获得对梯度的估计. 但广义目标函数 $\zeta(x)$ 只是对参数空间目标函数的形式化描述, 因此无法像函数优化问题一样精确地定义其采样邻域, 进而无法实现在参数空间的直接高斯采样. 所以扩散模型在学习空间中的高斯噪声扩散和去噪过程可以视为参数空间的高斯采样过程在学习空间中的近似. 如图 5 所示,

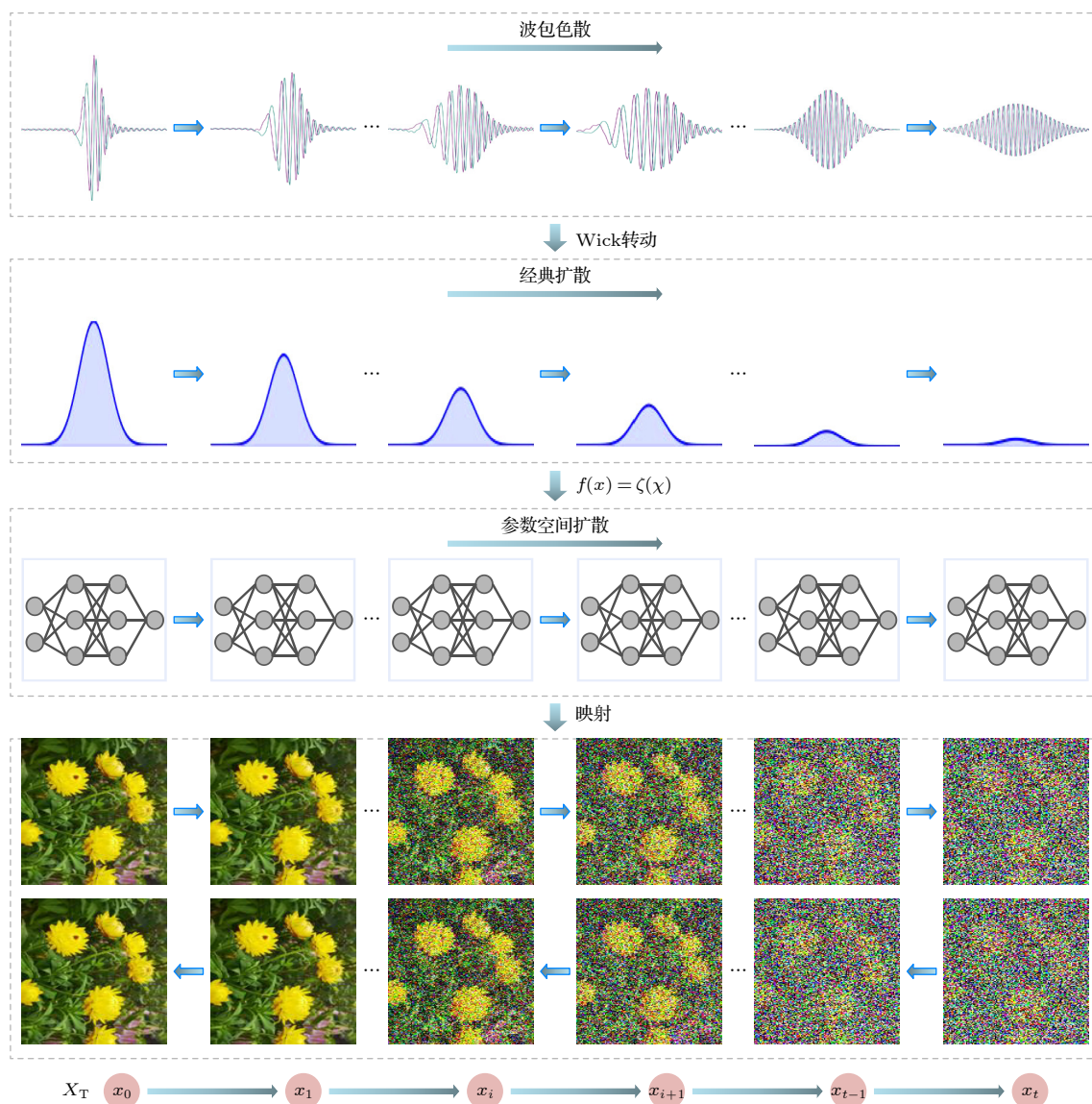


图 5 扩散模型的量子动力学诠释

Fig. 5. Quantum dynamical interpretation of diffusion models.

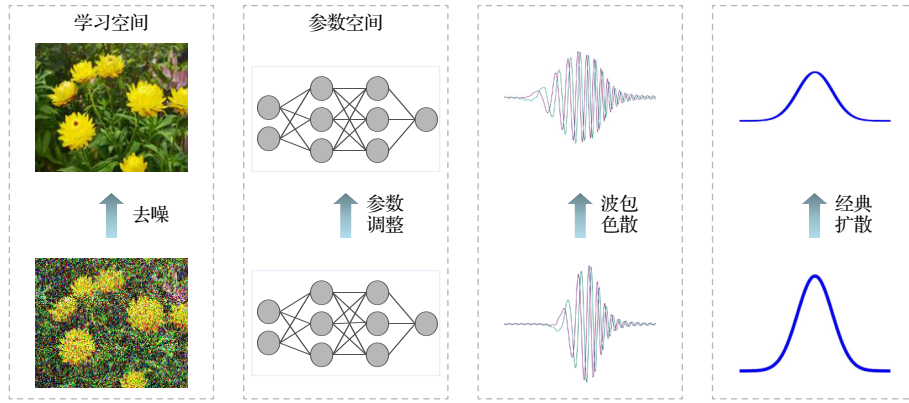


图 6 参数空间的采样映射

Fig. 6. Sampling mapping of parameter space.

扩散模型通过在学习空间中的高斯噪声扩散过程间接实现了机器学习算法在参数空间中的高斯采样扩散过程,从而解决了机器学习算法在参数空间中无法直接进行高斯采样的问题.所以,扩散模型的迭代过程是符合机器学习的量子动力学理论的.根据量子动力学的理论,扩散模型可以被认为是机器学习量子动力学经典近似下广义目标函数的0阶和1阶迭代操作.扩散模型中的每一次去噪迭代都对应于在广义参数空间中的一次高斯采样,一个样本的一次加噪和去噪迭代序列就对应于参数空间中一个采样点的高斯扩散采样序列.

从量子动力学的角度来看,机器学习中的扩散模型原本是基于在参数空间中进行高斯采样的过程.但由于在参数空间中无法定义目标函数,直接进行高斯采样变得不可行.因此,扩散模型将学习空间映射到参数空间,从而实现了在参数空间中进行高斯采样的目标.这一方法有效地克服了参数空间中目标函数不可定义的限制,为模型的学习和优化提供了新的理论框架.

机器学习就是利用学习空间中的训练集所形成的目标函数,通过训练迭代实现对参数空间广义目标函数 $\zeta(x)$ 的优化.机器学习的目标是对参数空间的广义目标函数进行优化,从而获得最优的模型参数,但通常算法无法实现在参数空间的直接采样,所以通过学习空间向参数空间映射来间接实现针对参数空间的优化.如图6所示,扩散模型通过去除高斯噪声的学习过程,间接实现了在参数空间中的高斯采样扩散过程,完成了学习空间向参数空间中高斯采样和梯度获取的映射.而参数空间中的高斯采样则是量子动力学理论中波包色散过程的

经典近似.

同时,根据机器学习的量子动力学,机器学习迭代过程也应该有逐步降低动能的退火过程,这表明在扩散模型中可能需要添加从大到小不同标准差的噪声,从而实现系统的退火.添加多尺度噪声的方法在扩散模型应用中的有效性目前也已被实验证明^[21].所以,完整的扩散模型也应该包含两个迭代循环过程:第一个迭代循环是高斯噪声标准差不断减小的退火过程;第二个迭代循环是扩散演化过程.

从量子动力学的角度和参数映射关系角度看扩散模型是基于目标函数0阶和1阶Taylor近似的生成式模型,输入图像通过扩散模型后所生成的图像为经过0阶和1阶优化后的低阶推理结果,因此,可以将扩散模型作为其他模型数据预处理的通用模型(如图7).原始图像首先经过扩散模型的处理,再将扩散模型生成的输出结果作为其他模型的输入.扩散模型在这里作为低阶推理过程完成图像的预处理,其他模型作为高阶推理过程完成进一步推理过程.这种新的推理结构将有可能提高整个模型的推理效果.

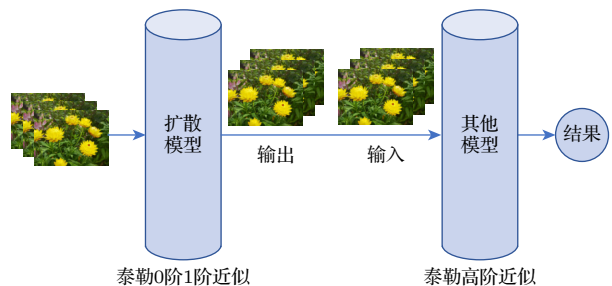


图 7 基于扩散模型的推理结构

Fig. 7. Inference structure based on diffusion models.

5 结 语

本文在 Schrödinger 方程的基础上建立了机器学习迭代过程的量子动力学方程, 并通过 Wick 转动将量子动力学方程与热动力学方程联系起来. 通过机器学习的量子动力学方程, 首先利用 Wick 转动和 Taylor 近似得到了机器学习的基本迭代过程; 其次, 利用方程的通解获得了机器学中常用的 Softmax 和 Sigmoid 函数, 从而在物理和数学上解释了 Softmax 和 Sigmoid 的动力学意义; 最后, 还采用量子动力学方程对扩散模型的基本迭代过程给出了动力学解释.

机器学习的量子动力学方程的建立, 将机器学习的迭代过程转化为含时偏微分方程, 实现了对机器学习过程的精确数学化描述, 从而可以充分利用量子力学和数学中成熟理论体系对机器学习展开研究. 本文的工作为建立机器学习的精确理论基础提供了一个新的研究视角和思路, 也有望为未来在量子计算机上实现人工智能算法提供理论依据^[22].

参考文献

- [1] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H, Teller E 1953 *J. Chem. Phys.* **21** 1087
- [2] Kirkpatrick S, Gelatt C D, Vecchi M P 1983 *Science* **220** 671
- [3] Finnila A B, Gomez M A, Sebenik C, Stenson C, Doll J D 1994 *Chem. Phys. Lett.* **219** 343
- [4] Wang F, Wang P 2024 *Quantum Inf. Process.* **23** 66
- [5] Wang P, Xin G 2023 *Acta Autom. Sin.* **49** 2396 (in Chinese) [王鹏, 辛罡 2023 *自动化学报* **49** 2396]
- [6] Wang P, Huang Y, Ren C, Guo Y 2013 *Acta Electron. Sin.* **41** 2468 (in Chinese) [王鹏, 黄焱, 任超, 郭又铭 2013 *电子学报* **41** 2468]
- [7] Wang P, Wang F 2022 *J. Univ. Electron. Sci. Technol. (Nat. Sci. Ed.)* **51** 2 (in Chinese) [王鹏, 王方 2022 *电子科技大学学报 (自然科学版)* **51** 2]
- [8] Johnson M W, Amin M H S, Gildert S 2011 *Nature* **473** 194
- [9] Sohl-Dickstein J, Weiss E, Maheswaranathan N, Ganguli S 2015 *Proceedings of the 32nd International Conference on Machine Learning* Lille, France, July 7–9, 2015 p2256
- [10] Song Y, Sohl-Dickstein J, Kingma D P, Kumar A, Ermon S, Poole B 2020 *arXiv: 2011.13456* [cs.LG]
- [11] Xin G, Wang P, Jiao Y 2021 *Expert. Syst. Appl.* **185** 115615
- [12] Jin J, Wang P 2021 *Swarm Evol. Comput.* **65** 100916
- [13] Wick G C 1954 *Phys. Rev.* **96** 1124
- [14] Dhariwal P, Nichol A 2021 *Advances in Neural Information Processing Systems (NeurIPS 2021)* December 7–10, 2021 (Virtual-only Conference) p8780
- [15] Ho J, Jain A, Abbeel P 2020 *Advances in Neural Information Processing Systems (NeurIPS 2020)* December 6–12, 2020 (Virtual-only Conference) p6840
- [16] Nichol A Q, Dhariwal P 2021 *Proceedings of the 38th International Conference on Machine Learning* July 18–24, 2021 (Virtual-only Conference) p8162
- [17] Lim S, Yoon E, Byun T, Kang T, Kim S, Lee K, Choi S 2023 *Advances in Neural Information Processing Systems (NeurIPS 2023)* New Orleans, USA, December 10–16, 2023 p37799
- [18] Anderson J B 1975 *J. Chem. Phys.* **63** 1499
- [19] Kosztin I, Faber B, Schulten K 1996 *Am. J. Phys.* **64** 633
- [20] Haghighi M K, Lüchow A 2017 *J. Phys. Chem. A* **121** 6165
- [21] Jeong J, Shin J 2023 *Advances in Neural Information Processing Systems (NeurIPS 2023)* New Orleans, USA, December 10–16, 2023 p67374
- [22] Morawietz T, Artrith N 2021 *J. Comput. Aid. Mol. Des.* **35** 557

Quantum dynamics of machine learning

WANG Peng[†] MAIMAITINIYAZI Maimaitiabudula

(*School of Computer Science and Engineering, Southwest Minzu University, Chengdu 610225, China*)

(Received 17 July 2024; revised manuscript received 27 November 2024)

Abstract

In order to solve the current lack of rigorous theoretical models in the machine learning process, in this paper the iterative motion process of machine learning is modeled by using quantum dynamic method based on the principles of first-principles thinking. This approach treats the iterative evolution of algorithms as a physical motion process, defines a generalized objective function in the parameter space of machine learning algorithms, and regards the iterative process of machine learning as the process of seeking the optimal value of this generalized objective function. In physical terms, this process corresponds to the system reaching its ground energy state. Since the dynamic equation of a quantum system is the Schrödinger equation, we can obtain the quantum dynamic equation that describes the iterative process of machine learning by treating the generalized objective function as the potential energy term in the Schrödinger equation. Therefore, machine learning is the process of seeking the ground energy state of the quantum system constrained by a generalized objective function. The quantum dynamic equation for machine learning transforms the iterative process into a time-dependent partial differential equation for precise mathematical representation, enabling the use of physical and mathematical theories to study the iterative process of machine learning. This provides theoretical support for implementing the iterative process of machine learning by using quantum computers. In order to further explain the iterative process of machine learning on classical computers by using quantum dynamic equation, the Wick rotation is used to transform the quantum dynamic equation into a thermodynamic equation, demonstrating the convergence of the time evolution process in machine learning. The system will be transformed into the ground energy state as time approaches infinity. Taylor expansion is used to approximate the generalized objective function, which has no analytical expression in the parameter space. Under the zero-order Taylor approximation of the generalized objective function, the quantum dynamic equation and thermodynamic equation for machine learning degrade into the free-particle equation and diffusion equation, respectively. This result indicates that the most basic dynamic processes during the iteration of machine learning on quantum computers and classical computers are wave packet dispersion and wave packet diffusion, respectively, thereby explaining, from a dynamic perspective, the basic principles of diffusion models that have been successfully utilized in the generative neural networks in recent years. Diffusion models indirectly realize the thermal diffusion process in the parameter space by adding Gaussian noise to and removing Gaussian noise from the image, thereby optimizing the generalized objective function in the parameter space. The diffusion process is the dynamic process in the zero-order approximation of the generalized objective function. Meanwhile, we also use the thermodynamic equation of machine learning to derive the Softmax function and Sigmoid function, which are commonly used in artificial intelligence. These results show that the quantum dynamic method is an effective theoretical approach to studying the iterative process of machine learning, which provides a rigorous mathematical and physical model for studying the iterative process of machine learning on both quantum computers and classical computers.

Keywords: quantum dynamics, machine learning, diffusion model, Schrödinger equation

PACS: 07.05.Mh, 03.65.-w, 05.40.-a, 02.30.Jr

DOI: [10.7498/aps.74.20240999](https://doi.org/10.7498/aps.74.20240999)

CSTR: [32037.14.aps.74.20240999](https://cstr.cn/32037.14.aps.74.20240999)

[†] Corresponding author. E-mail: wp002005@163.com



机器学习的量子动力学

王鹏 麦麦提尼亚孜·麦麦提阿卜杜拉

Quantum dynamics of machine learning

WANG Peng MAIMAITINIYAZI Maimaitiabudula

引用信息 Citation: *Acta Physica Sinica*, 74, 060701 (2025) DOI: 10.7498/aps.74.20240999

CSTR: 32037.14.aps.74.20240999

在线阅读 View online: <https://doi.org/10.7498/aps.74.20240999>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于波动与扩散物理系统的机器学习

Machine learning based on wave and diffusion physical systems

物理学报. 2021, 70(14): 144204 <https://doi.org/10.7498/aps.70.20210879>

机器学习辅助绝热量子算法设计

Machine learning assisted quantum adiabatic algorithm design

物理学报. 2021, 70(14): 140306 <https://doi.org/10.7498/aps.70.20210831>

蛋白质计算中的机器学习

Machine learning for *in silico* protein research

物理学报. 2024, 73(6): 069301 <https://doi.org/10.7498/aps.73.20231618>

生物分子模拟中的机器学习方法

Machine learning in molecular simulations of biomolecules

物理学报. 2023, 72(24): 248708 <https://doi.org/10.7498/aps.72.20231624>

机器学习结合固溶强化模型预测高熵合金硬度

Machine learning combined with solid solution strengthening model for predicting hardness of high entropy alloys

物理学报. 2023, 72(18): 180701 <https://doi.org/10.7498/aps.72.20230646>

机器学习在光电子能谱中的应用及展望

Application and prospect of machine learning in photoelectron spectroscopy

物理学报. 2024, 73(21): 210701 <https://doi.org/10.7498/aps.73.20240957>