

量子态的数学与神经网络表示

丛爽

中国科学技术大学自动化系, 合肥 230027

摘要 随着深度学习的广泛应用,神经网络模型的数据自生成以及概率模拟等功能在量子态重构与估计方面的应用得到人们关注。通过对量子态的各种不同的数学表示,引导到神经网络量子态的不同表示;从量子态不同物理变量之间的关系,推导出相应神经网络结构上的输入/输出之间的非线性映射关系,为采用不同类型的神经网络模型根据自身数据自生成以及概率模拟功能,实现量子态估计应用,提供网络函数关系及其数据生成的理论设计基础。

关键词 量子态;神经网络表示;概率模拟;密度矩阵重构

量子系统的状态可以在希尔伯特空间采用波函数表示,也可以采用密度矩阵 ρ 描述。密度矩阵可以通过对波函数的外积叠加获得。对一个量子系统中的密度矩阵的估计,需要用到量子系统输出的测量值。量子状态(简称量子态)估计可以分为完备测量和不完备测量。完备测量是指在选定的正交测量基矢下测量次数与待估计参数值数目相等的测量。一个 n 比特量子系统状态的密度矩阵 $\rho \in \mathbb{C}^{d \times d}$ ($d=2^n$)所具有的维数是 $d=2^n$,密度矩阵同时还需要满足半正定($\rho \geq 0$),迹为1($\text{tr}(\rho)=1$)和共轭对称($\rho^\dagger=\rho$)的性质。由于待估计 ρ 的未知元素数目为 $d \times d=2^n \times 2^n=4^n$ 个,所以对量子系统中一个量子态的完备测量次数为 4^n 量级,测量次数是随着量子位数的增加呈指数增长^[1]。量子状态层析是最早用来

进行量子状态估计的方法^[2]。Smith等基于最小二乘法从带有噪声的数据中实现了4比特量子状态估计^[3]。Li等^[4]提出了一种低测量比率下的高量子位的快速重构。Zhang等^[5]提出了带有噪声和干扰的量子态滤波器。采用神经网络进行量子态估计,有可能通过对神经网络输入量子位数 n ,即 n 输入节点,得到期望的 2^n 个重构参数,这样就有可能将 2^n 的计算复杂度减少为 n 多项式的计算复杂度,从而减少训练网络时间,加快收敛速度,提高逼近精度。这是近年来人们采用神经网络进行量子态重构的动因。Xin等^[6]使用全连接神经网络从局部测量信息实现了量子态层析,为大规模量子系统的可扩展状态层析提供了有效方法。Lohani等^[7]实现了利用卷积神经网络进行量子态层析,并获得较高的

收稿日期:2022-03-03;修回日期:2022-05-27

基金项目:国家自然科学基金项目(61973290)

作者简介:丛爽,教授,研究方向为先进控制策略、人工智能、智能控制及量子系统控制,电子信箱:scong@ustc.edu.cn

引用格式:丛爽. 量子态的数学与神经网络表示[J]. 科技导报, 2023, 41(19): 48-58; doi: 10.3981/j.issn.1000-7857.2023.19.005

平均保真度。当采用神经网络进行应用时,除了需要掌握神经网络本身的设计,更重要的是要将需要解决的问题和实现的目标转换成能够以神经网络实现的函数映射关系,对于采用神经网络进行量子态估计,重要的一步就是实现神经网络的量子态表示。

本文根据量子态的波函数和密度矩阵的数学表示,分别对量子本征态、纯态以及混合态和纠缠态在选择不同测量基下的不同表示,以及它们之间的不同转换和关系,通过具体数值例子对其进行特性分析,给出2种典型的神经网络量子态的表示;根据不同物理变量之间的关系,以及神经网络自身所具有的特性,推导出相应神经网络结构上的输入/输出之间的非线性映射关系,在理论上提供神经网络函数关系及其数据生成的设计原则。

1 量子态的数学表示

1.1 波函数的数学表示

量子态的波函数 $\psi(t)$ 又称为波矢,是一个矢量。狄拉克(Dirak)提出采用右矢 $|\cdot\rangle$ 符号表示量子波函数 $|\psi\rangle$,采用左矢 $\langle\cdot|$ 表示量子波函数的共轭转置 $\langle\psi| = (|\psi\rangle)^\dagger$ 。最简单的单量子位的波矢是一个 2×1 的列向量,可以表示为 $|\psi\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = [1 \ 0]^\top$,或者 $|\psi\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = [0 \ 1]^\top$ 。这2个矢量在二维平面内正好组成一个正交基,波函数的所有状态可以由这2个正交的基本矢量(简称基矢)线性叠加而成。最常用定义单量子位的2个正交基矢分别为 $|0\rangle = [1 \ 0]^\top$ 和 $|1\rangle = [0 \ 1]^\top$ 。这里所定义的2个基矢又称为本征态,选用本征态为位基矢是最直接的,当然还可以定义其他基矢,将在本文的1.3节叙述。任意一个波函数的表达式可以为这2个本征态基矢的线性叠加:

$$|\psi\rangle = a|0\rangle + b|1\rangle \quad (1)$$

式中, a 和 b 为满足 $a^2+b^2=1$ 的复数。

因为 $b=0$ 时,有 $a^2=1$ 成立; $a=0$ 时,有 $b^2=1$ 成立,此时的波函数是(以概率1)分别处于 $|0\rangle$ 和 $|1\rangle$,所以 a^2 和 b^2 分别表示波函数本征态等于 $|0\rangle$ 和 $|1\rangle$ 的概

率。 a 和 b 被称为(概率)幅(开)方。由于 a 和 b 为任意复数,原则上通过适当地确定 a 和 b 的值,就可以对一个量子位编码出无穷多个量子态波函数。所谓的量子态的制备、估计、重构以及调控,都是由本征态组成波函数的系数 a 和 b 的值的设计与操纵。例如,可以采用 $|0\rangle$ 和 $|1\rangle$ 基矢分别表示半自旋粒子系统的自旋向上和自旋向下的状态: $|\uparrow_x\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ 和 $|\downarrow_x\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$;以及 $|\uparrow_y\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$ 和 $|\downarrow_y\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle)$ 。

当量子位为 $n=2$ 时,波函数的本征态个数有 $2^n=2^2=4$ 个,当采用本征态作为测量基时,波函数的表达式为

$$|\psi\rangle = a_1|00\rangle + a_2|01\rangle + a_3|10\rangle + a_4|11\rangle \quad (2)$$

式中, $a_1^2 + a_2^2 + a_3^2 + a_4^2 = 1$ 。

以此类推可得, n 量子位波函数可以由所定义的 2^n 个基矢的叠加组成:

$$|\psi\rangle = \sum_{i=1}^{2^n} a_i |x_1 x_2 \cdots x_n\rangle, \quad x_i \in \{0,1\} \quad (3)$$

从式(3)中可以看出,对于 n 量子位组成的波函数,有 n 个变量, 2^n 个基矢,以及与其对应的 2^n 个幅方系数 $a_i, i=1, 2, \dots, 2^n$ 。

1.2 密度矩阵的数学表示

由于波函数平方后得到的系数具有概率的物理意义,人们通过计算波函数的外积获得矩阵,得到其元素具有明确物理意义的密度矩阵

$$\rho = |\psi\rangle\langle\psi| = \sum_i p_i |\psi_i\rangle\langle\psi_i| \quad (4)$$

式中, p_i 为状态对应 $|\psi_i\rangle\langle\psi_i|$ 的概率,且满足 $\sum_i p_i = 1$ 。

密度矩阵中的每一个元素都对应着所定义的基矢以及不同基矢间的相干出现的概率,这个概率可以通过对实际量子系统的实验,反复对同一个量子态的测量并进行平均计算出来。通过这样的反复测量实验,获得含有量子状态信息的实验数据,对数据进行处理后,可重构(或估计)出量子状态的密度矩阵。

需要特别强调和提醒的是,对于任意一个量子系统的状态制备、测量、估计以及调控,都是对幅方 a_i , $i=1, 2, \dots, 2^n$ (波函数系数), 或 a^2 (密度矩阵系数) 的操控, 也就是人们只能对各个基矢及其相干态在组成量子态中出现的概率进行测量和操控, 而不是对本征态本身的操控。所以量子态或系统的控制是一种对概率的控制, 这与宏观系统中直接对系统状态的控制概念完全不同。

举个简单的例子(例1)。向空中抛一枚硬币, 当此硬币落入手中时, 是观测到正面还是反面? 采用量子态的概念可以很精确求解如下:

1) 当此硬币抛向空中时, 此硬币在空中不断旋转, 所处的状态是|正面〉和|反面〉的叠加态: $|\psi\rangle = a|\text{正面}\rangle + b|\text{反面}\rangle$ 。因为 a 和 b 可以取任意复数, 所以 $|\psi\rangle$ 是一个连续不断变化的状态。

2) 当硬币落入手中观测其结果时, $|\psi\rangle$ 不再是一个叠加态, 只能是|正面〉或者|反面〉中的一种结果。换句话说, 硬币塌缩到组成硬币正或反面(本征态)中的一个, 这就是量子态观测的塌缩性。至于是塌缩到哪一个本征态(|正面〉或者|反面〉), 这是随机的, 这就是量子态观测的随机性, 它取决于观测的量子态本身在各个本征态上的幅方 a_i , $i=1, 2, \dots, 2^n$ 的大小。所以通过一次观测就准确说出落入手中的硬币的状态是做不到的, 因为要确定的是组成硬币状态的幅方 a_i , $i=1, 2, \dots, 2^n$ 的值, 不过 a_i , $i=1, 2, \dots, 2^n$ 的值可以通过多次实验结果求出。

3) 反复抛出硬币, 观测并记下其落入手中的结果, 看到|正面〉时, 记录一个“+”号; 看到|反面〉时, 记录一个“-”号。做100次观测实验后, 分别统计|正面〉“+”号和|反面〉“-”号的总数, 可能的结果是49:51, 那么, 一次观测到落入手中硬币为|正面〉的概率为 $a^2=49/100$, 为|反面〉的概率为 $b^2=51/100$ 。随着观测实验次数的增加, 这2个概率数值将越来越接近50%:50%, 所以出现|正面〉和|反面〉的概率各占50%。这是一个精确答案。硬币抛向空中落入手中的精确量子态的表达式为

$$|\text{硬币观测状态}\rangle = \frac{1}{\sqrt{2}}(|\text{正面}\rangle + |\text{反面}\rangle) \quad (5)$$

1.3 量子态密度矩阵的不同表示方法

例1通过获取多次实验的、硬币塌缩到某一个本征态的测量值, 以及对测量值的统计计算, 求出观测硬币正、反两面的精确幅方系数 a 和 b 各等于 $1/\sqrt{2}$, 这个过程就是基于测量值的量子态重构(估计)过程。过程中最关键的信息是, 根据实验获得的观测数据, 此数据不是 a 和 b 本身, 但其中是包含着所要重构的 a 和 b 的信息。根据这些数据, 就可以对 a 和 b 进行估计, 因为在观测数据与参数 a 和 b 之间, 存在某种关系式。例如, 例1中对落下的硬币状态的估计就是通过求观测频率的平均值 y 来求得, 它们之间的关系为

$$y = \text{tr}(\mathbf{M}\rho) \quad (6)$$

式中, ρ 为量子系统状态, y 是通过观测结果计算出各个状态出现频率的平均值, \mathbf{M} 是由所定义的所有正交基组成的测量算符。

实际在数学计算上, 每一次测量值的获得就是通过将所定义的测量算符 \mathbf{M} 作用在待估计的量子态 ρ 上, 计算出的随机塌缩到的某个本征态(正交投影)上的值, 即 $\mathbf{M}_i\rho$, 然后通过求迹运算 $\text{tr}(\mathbf{M}\rho)$, 就可得到所有观测值的平均值。在实际实验中, 测量算符 \mathbf{M} 是做实验时定义的, 所以是已知的; 而 y 值是直接从实验装置上测量到的, 也是已知的, 所以理论上根据公式(6)求逆, 就可以求出密度矩阵的估计值 $\hat{\rho}$ 。在实际应用中, 人们所选择或定义的正交基是不同的, 这样就得到不同的测量算符, 不同情况下所获得的观测值不同, 进而进行量子态的估计过程也可能不同。可以说对量子态观测值获取的关键在于选取或构造不同的测量算符 \mathbf{M} 。

\mathbf{M} 由不同的正交基组成, 正交基矢的定义或选取不同, 与之对应的测量算符 \mathbf{M} 的表达式也不同。例1中选取的基矢分别为 $|0\rangle = [1 \ 0]^T$ 和 $|1\rangle = [0 \ 1]^T$, 所对应的测量算符 \mathbf{M} 为: $\mathbf{M}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ 和 $\mathbf{M}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 。除了采用 $|0\rangle$ 和 $|1\rangle$ 表示单量子位 $|\psi(a, b)\rangle = a|0\rangle + b|1\rangle$ 外, 还可以采用其他方式来表示。比如, 采用空间坐标形式来表示一个量子位的波函数的表达式为

$$|\psi(\theta, \phi)\rangle = \cos\theta|0\rangle + e^{i\phi}\sin\theta|1\rangle \quad (7)$$

此时,需要确定的是2个实参数 θ 和 ϕ ,通过这2个参数所确定的量子态是空间单位球面上的一个点,并且任意一个量子纯态都可在此球面上表示出来。单量子位可以被表示为空间三维坐标 (x, y, z) 中的一个点,其中,本征态 $|0\rangle$ 的坐标为 $|0\rangle = (\sin 0 \cos \phi \quad \sin 0 \sin \phi \quad \cos 0) = (0 \quad 0 \quad 1)$,本征态 $|1\rangle$ 的坐标为 $|1\rangle = (\sin \pi \cos \phi \quad \sin \pi \sin \phi \quad \cos \pi) = (0 \quad 0 \quad -1)$,它们分别处于 z 轴1和-1处,也就是单位球(Bloch球)的北极和南极。

另外还可以采用水平状态 $|H\rangle$ 和垂直状态 $|V\rangle$ 作为一组正交基矢来表示单量子态,或采用Stocks,或泡利(Paoli)矩阵等来表示单量子态等。换句话说,人们可以根据具体情况来选择不同的正交基矢来表示量子态。由于这些基矢在希尔伯特空间是可以相互转换的,例如,泡利矩阵与 $|0\rangle$ 和 $|1\rangle$ 之间的关系为 $\sigma_x = |1\rangle\langle 0| + |0\rangle\langle 1| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_y = i|1\rangle\langle 0| - i|0\rangle\langle 1| = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ 。这些正交基矢都是由 $|0\rangle$ 和 $|1\rangle$ 组合而成,最常用的采用泡利矩阵作为正交基矢的密度矩阵的表示为

$$\begin{aligned} \rho &= \frac{1}{2}(I + n_x\sigma_x + n_y\sigma_y + n_z\sigma_z) \\ &= \frac{I + \mathbf{n} \cdot \boldsymbol{\sigma}}{2} = \frac{1}{2} \begin{pmatrix} 1 + n_z & n_x - in_y \\ n_x + in_y & 1 - n_z \end{pmatrix} \end{aligned} \quad (8)$$

式中, n_x, n_y 和 n_z 为 ρ 在三维空间中的坐标矢量,确定了 $\mathbf{n}=(n_x, n_y, n_z)$,就确定了密度矩阵 ρ 。

根据式(8)可以求出密度矩阵与Block矢量之间的对应关系:对于采用 $|0\rangle$ 和 $|1\rangle$ 表示的一个混合态密度矩阵 $\rho_1 = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|$,有

$$\begin{aligned} \rho_1 &= \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1| \\ &= \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 + \frac{1}{2} & 0 \\ 0 & 1 - \frac{1}{2} \end{pmatrix} \end{aligned} \quad (9)$$

令式(8)与式(9)系数相等,可得 $n_x=0, n_y=0$ 和 $n_z=1/2$,即 $\mathbf{n}=(0, 0, 1/2)$ 。

在实际应用中很有必要选择合适的正交基矢

表示量子态,其作用之一即简化对系统状态操控过程中的复杂性,可看作是一种数学处理技巧。比如,对于一个量子态进行跟踪控制,要求被控量子态在每个时刻与期望状态一致,而量子态本身带有自旋。为简化跟踪的复杂度,可通过对量子态进行一个与其自身自旋 H_0 一致的旋转操作,即进行相互作用变换,将量子态变换到自旋的轨道上,这样被控量子态与其自旋同轨,就可以在跟踪的控制过程中,不必考虑自旋的影响,如同人登上行驶的汽车后与汽车的相对速度为0,与汽车的位置相对静止。

2 量子态波函数与密度矩阵的特性

量子态一般可被分为本征态、叠加态和混合态,还可为纠缠态、压缩态等。本征态即所选测量空间的正交基组成的状态,并且所有的正交基组成一个测量算符。单量子的本征态有2个,采用波函数表示分别为 $|0\rangle = [1 \quad 0]^T$ 和 $|1\rangle = [0 \quad 1]^T$;采用密度矩阵分别表示为 $\rho_3 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ 和 $\rho_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$,其测量算符为 $M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ 。纯态是本征态的叠加态,例如 $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$,该叠加态的密度矩阵表示为 $\rho_5 = |\psi\rangle\langle\psi| = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ 。混合态是纯态的几率混合,例如 $\rho_6 = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1| = \begin{pmatrix} 3/4 & 0 \\ 0 & 1/4 \end{pmatrix}$ 。

波函数只能表示纯态,密度矩阵既可以表示纯态,也可以表示混合态。纠缠态是不可分离的叠加态,例如,2量子位叠加的波函数

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \quad (10a)$$

就是纠缠态;而2量子位叠加的波函数 $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |01\rangle)$ 就是可分离的叠加态,因为它可以分离写为 $|\psi_2\rangle = \frac{1}{\sqrt{2}}|0\rangle(|0\rangle + |1\rangle)$;2量子位是由不可分离叠加态组成的最大纠缠态。除了 $|\psi_1\rangle$ 外,还有一对:

$$|\psi_3\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) \quad (10b)$$

另一方面,由于密度矩阵 $\rho = |\psi\rangle\langle\psi|$,它同时是一个正定的厄密矩阵,满足 $\rho \geq 0, \rho^\dagger = \rho$,并且 $\rho^2 = \rho$ 。一个纯态密度矩阵满足: $\text{tr}(\rho) = \sum_i p_i \langle\psi_i|\psi\rangle = 1, \rho^2 = \rho$,合在一起的条件满足

$$\rho^\dagger = \rho, \rho \geq 0 \text{ 和 } \text{tr}(\rho) = \text{tr}(\rho^2) = 1 \quad (11)$$

一个混合态密度矩阵满足

$$\text{tr}(\rho) = 1 \text{ 和 } \text{tr}(\rho^2) < 1 \quad (12)$$

只有满足上述条件的密度矩阵才是量子态密度矩阵,单量子位的全部状态可以采用 Bloch 球显示(图1)。图1(b)中,单位球面上的所有点为纯态,球内的所有点为混合态,球心为完全(或最大)混合态。

对于密度矩阵表示特别强调的是:

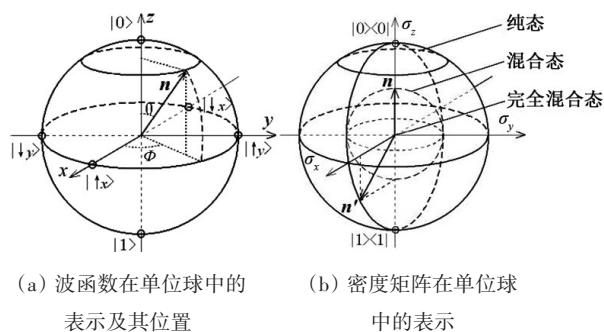


图1 单量子位状态的 Bloch 球表示

1) 一个本征态的密度矩阵形式一定是整个矩阵中只有对角线上仅有一个1,其他元素全部为0。

2) 一个叠加态的密度矩阵形式有2种:(1) 对角线上多于一个元素有值,其余元素全部为0。(2) 除了对角线上有值外,非对角线上也有值。

3) 混合态的密度矩阵形式与叠加态相似。一个密度矩阵 ρ 是叠加态还是混合态,必须通过计算其性质才能确定: $\text{tr}(\rho) = 1$ 为叠加态, $\text{tr}(\rho) < 1$ 为混合态。

在实际应用中,为了保证能够求出同时满足所需条件的密度矩阵,人们往往把条件同时加入到所求的密度矩阵中,也就是求出满足归一化条件的密度矩阵。

1) 对于由波函数表示的满足纯态条件的密度矩阵的表示为

$$\rho = \frac{\psi^\dagger \psi}{\text{tr}(\psi^\dagger \psi)} \quad (13)$$

2) 由密度矩阵表示的满足纯态条件的密度矩阵的估计值为

$$\hat{\rho}_i = \frac{\hat{\rho}^\dagger \hat{\rho}}{\text{tr}(\hat{\rho}^\dagger \hat{\rho})} \quad (14)$$

3 量子态的神经网络表示

不论是具有1到2个隐含层的浅层神经网络,还是深度神经网络,用来进行量子态重构的原理都是相似的:根据对量子态在正交基上的投影测量数值来获得测量数据 $\{y_i\}$,以及该数据与波函数及其幅方之间的关系(3)式或与密度矩阵之间的关系(6)式选择一种神经网络,并设计其结构,包括层数、输入层/隐含层/输出层的节点数,选择合适的网络权值训练算法,通过有监督或无监督的对网络参数的训练,达到逼近波函数幅方或密度矩阵中各参数的目标。

实际上各种不同类型的神经网络几乎都可以用来实现对量子态参数逼近的应用,由于各种神经网络结构上的特殊性,以及训练网络的各种快速算法近年来不断被开发,因为对密度矩阵的重构的数值迭代的优化算法都是对维数为 $d \times d$ 密度矩阵中的 d^2 个元素进行重构,由于 $d=2^n$,所以待估计的参数数量随着量子位数 n 的增加而指数增长,极大地制约了高量子位量子态的估计。

将神经网络应用于量子态重构的第一步就是要弄清楚如何采用神经网络对量子态表示,也就是神经网络输入/输出节点数的设计,以及训练样本数据的产生。根据第1节中所阐述的各种关系,神经网络的量子态表示可以分为2种: 2^n 和 n 输入节点的神经网络表示。

3.1 2^n 输入节点的神经网络表示

3.1.1 神经网络输入的纯态表示

从公式(6)中测量值与量子态密度矩阵之间的关系 $y = \text{tr}(M\rho)$ 可以看出, M 是所选定的所有正交基

组成的测量算符, y 是在 M 对 ρ 的作用下的测量值, 其物理意义为: ρ 随机塌缩投影到某个正交基上的值。现在的情况是, M 是在希尔伯特空间上选取构造的已知测量算符, y 是实验获取也是已知的, 根据公式 $y = \text{tr}(M\rho)$, 可以得到关系式:

$$\hat{\rho} = F(y_i(M_i)), i = 1, 2, \dots, 2^n \quad (15)$$

由式(15)可以看出, 神经网络的输入数据是由每一个 M_i 对应的 y_i , 网络输出是组成待估计密度矩阵 $\hat{\rho}$ 的全部元素。神经网络要达到的目标是从 $y_i(M_i)$ 到 $\hat{\rho}$ 的非线性映射, 也就是训练神经网络实现非线性函数 $\hat{\rho} = F(y_i(M_i))$ 。对于 n 量子位, 与重构元素一致的完备测量次数是 $d \times d = 2^n \times 2^n = 4^n$ 。所以神经网络的输入节点有 4^n 个, 也就是所有的测量基的数目。按排列顺序应当为 4^n 个正交基。如果将每一次获得 y_i 与 M_i 相对应, 那么所获得的测量数据中, 每 4^n 不同的 y_i 组成一组网络输入, 称之为测量向量, 用 b_k 表示: $b_k = [y_1, y_2, \dots, y_{d \times d}]$ 。每一组都是一组完备的测量。例如, 对于 2 比特量子系统, 总测量次数为 $d \times d = 4 \times 4 = 16$ 次, 16 次测量所对应的完备测量算符为 $\{M_i\} = \{II, IX, IY, IZ, XI, XX, XY, XZ, YI, YX, YY, YZ, ZI, ZX, ZY, ZZ\}$, 其中, I, X, Y, Z 分别表示单位算符和泡利算符的 $\sigma_x, \sigma_y, \sigma_z$ 分量, “YZ” 表示由 Y, Z 的直积构成的测量算符。

考虑到实际测量值中含有噪声, 对每一个待测量子密度矩阵 ρ_k 的测量向量 b_k 加上高斯噪声序列可得

$$b_k = \mathcal{A}_k \text{vec}(\rho_k) + e_k \quad (16)$$

式中, $\mathcal{A}_k = (\text{vec}(M_1), \text{vec}(M_2), \dots, \text{vec}(M_{d \times d}))^\dagger$, $\text{vec}()$ 表示将矩阵按列组合成 1 个向量。

式(16)就是利用测量算符作用在待估计的密度矩阵上构造出的神经网络的输入样本。这样对每一个网络输出的密度矩阵, 就得到了相应的输入样本(16), 也就构成了一个神经网络训练的输入/输出对。给输入样本 b_k 的每个测量值中加入高斯噪声进行训练, 也可以使网络更加鲁棒, 增加网络的抗噪能力。如果没有噪声, 理论上对于具有完备测量次数作为输入节点的神经网络, 经过一组输入

的网络, 甚至经过 1 次训练就可能得到高性能的网络量子态输出。

3.1.2 神经网络输出的纯态表示与生成

神经网络输出为密度矩阵 ρ , 其中包含的元素为 $d^2 = 4^n$ 。由于神经网络的输出是对真实密度矩阵的逼近, 但不一定能满足(11)式中密度矩阵要求的厄米矩阵、半正定性和单位迹这 3 个性质, 所以需要通过对神经网络的输出 $\hat{\rho}$ 进行归一化变换, 采用(14)式来使量子纯态满足这 3 个性质。

1) 本征态密度矩阵为仅有对角线某一位置有元素 1, 其余位置均为 0 的矩阵, 所以对于一个 2 量子位本征态, 先生成一个 $d \times d = 4 \times 4$ 的零矩阵, 然后随机将其对角线中的一个元素置为 1, 这样即得到一个本征态密度矩阵, 再结合式(16)生成 b_k , 即可得到一个本征态输入/输出样本对。

2) 叠加态的生成则根据式(13), 其中 ψ 为 $d \times r$ 的 Wishart 矩阵, 且矩阵内各元素满足均值为 0、标准差为 1 的正态分布^[8]。秩 $r=1$ 时, 为叠加态密度矩阵。当采用泡利矩阵构造的测量算符来对密度矩阵(13)进行测量, 将所有的测量结果组成向量作为输入样本。泡利矩阵是一组 3 个 2×2 的厄米复矩阵, 一般用希腊字母 $\sigma_x, \sigma_y, \sigma_z$ 表示, 将单位矩阵 I 视为第零号泡利矩阵:

$$\begin{aligned} \sigma_x &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ \sigma_z &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \quad (17)$$

对于 n 比特量子系统, 由泡利矩阵构成的测量算符 W_i 是从集合 $\{\sigma_x, \sigma_y, \sigma_z, I\}$ 中随机挑选 n 个矩阵, 然后计算这些矩阵的归一化张量积组成:

$$W_i = \frac{\sigma_{i_1} \otimes \dots \otimes \sigma_{i_n}}{\sqrt{d}} \quad (18)$$

由(18)式可以得到所有完备的测量算符 M_i 。给定一个测量算符 M_i , 对密度矩阵 ρ_k 的测量值 y_i 为

$$y_i = \text{tr}(M_i^\dagger \rho_k) = \text{vec}(M_i)^\dagger \text{vec}(\rho_k) \quad (19)$$

3.1.3 混合态输入/输出样本对的生成

混合态密度矩阵的秩 $r \neq 1$, 而是 $r > 1$, 所以对于混合态, 在对神经网络的输出进行归一化变换时无法采用式(14)。此时需要先按照公式生成(13), 其

中的 ψ 为 $d \times r$ 的 Wishart 矩阵, 令 $r=d$, ψ 变为 $d \times d$ 的矩阵, 此时生成的 ρ 为混合态密度矩阵。由于密度矩阵是厄米矩阵, 为了使最终的网络输出结果满足密度矩阵的性质, 需要对密度矩阵 ρ 进行矩阵分解: $\rho = \frac{\tau^\dagger \tau}{\text{tr}(\tau^\dagger \tau)}$, 其中 τ 矩阵为下三角矩阵。以 2 量子位为例, 三角矩阵 τ 与 ρ 中各元素关系为

$$\tau = \begin{bmatrix} \sqrt{\frac{\text{Det}(\rho)}{m_{11}^{(1)}}} & 0 & 0 & 0 \\ \frac{m_{12}^{(1)}}{\sqrt{m_{11}^{(1)} m_{11,12}^{(2)}}} & \sqrt{\frac{m_{11}^{(1)}}{m_{11,12}^{(2)}}} & 0 & 0 \\ \frac{m_{12,23}^{(2)}}{\sqrt{\rho_{44}} \sqrt{m_{11,12}^{(2)}}} & \frac{m_{11,23}^{(2)}}{\sqrt{\rho_{44}} \sqrt{m_{11,12}^{(2)}}} & \sqrt{\frac{m_{11,12}^{(2)}}{\rho_{44}}} & 0 \\ \frac{\rho_{41}}{\sqrt{\rho_{44}}} & \frac{\rho_{42}}{\sqrt{\rho_{44}}} & \frac{\rho_{43}}{\sqrt{\rho_{44}}} & \sqrt{\rho_{44}} \end{bmatrix} \quad (20)$$

式中, $m_{ij}^{(1)}$ 为 ρ 的一阶余子式, $m_{ijkl}^{(2)}$ 为 ρ 的二阶余子式。 τ 矩阵可以用参数 $t_1, t_2, \dots, t_{15}, t_{16}$ 表示为

$$\tau = \begin{bmatrix} t_1 & 0 & 0 & 0 \\ t_5 + it_6 & t_2 & 0 & 0 \\ t_{11} + it_{12} & t_7 + it_8 & t_3 & 0 \\ t_{15} + it_{16} & t_{13} + it_{14} & t_9 + it_{10} & t_4 \end{bmatrix} \quad (21)$$

神经网络的输出样本即可用 $t_1, t_2, \dots, t_{15}, t_{16}$ 表示, 神经网络的输入依旧按照式(16)进行生成, 此时网络仅需要输出 16 个元素, 为估计的 $\hat{\tau}$ 矩阵, 再通过公式:

$$\hat{\rho}_i = \frac{\hat{\tau}^\dagger \hat{\tau}}{\text{tr}(\hat{\tau}^\dagger \hat{\tau})} \quad (22)$$

就可得到满足性质的密度矩阵的估计 $\hat{\rho}_i$ 。

以 $n=2$ 比特量子状态为例, 完备的测量值 y 共有 $4 \times 4 = 16$ 个, 将其组成一组向量作为神经网络的输入, 网络的最后输出为估计出的密度矩阵 ρ 表示, 网络输出节点数为密度矩阵的维数: $d \times d = 2^n \times 2^n$, 即 4×4 , 因为密度矩阵可以是复数, 所以将密度矩阵的实数与虚数部分分开, 可以组成 2 个矩阵, 然后将这 2 个矩阵按列拉成向量, 并组合在一起得到 $2 \times 4 \times 4$ 个输出值, 也就是输出层节点数为 32 个, 其中实部和虚部各为 16 个数据。图 2 为采用前向神经网络进行 2 比特量子态估计的网络结构^[9], 其

中, 神经网络由 2 个隐含层构成, 隐含层激活函数选择对数型 Sigmoid 函数, 输入数据采用泡利矩阵作为测量算符的测量结果(6)组成的输入向量 $[y_1, y_2, \dots, y_{16}]^T$, 输出层为量子态密度矩阵实部和虚部分开后, 再按列拉成的输出向量 $[\rho_{r11}, \rho_{r21}, \dots, \rho_{r34}, \rho_{i44}, \rho_{i11}, \rho_{i21}, \dots, \rho_{i34}, \rho_{i44}]^T$, 其中下标 r 代表实数, i 代表虚数, 数字代表所处的行和列。

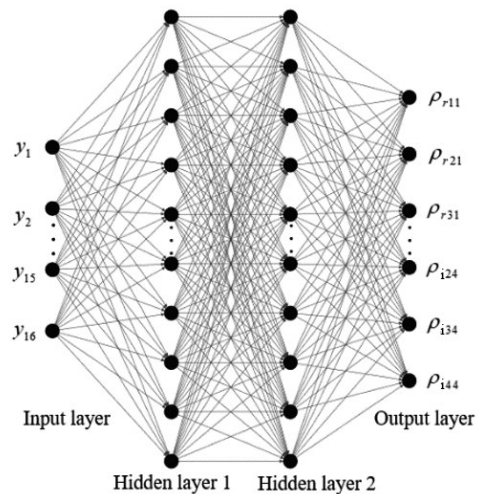


图2 2比特前向神经网络重构结构

3.2 n 输入节点的神经网络表示

对于 n 量子位, 波函数(3)中的每一个系数的平方为状态 $|i\rangle$ 的概率, 人们只要求出所有的正交基 $|i\rangle$ 的系数 $a_i, i=1, 2, \dots, 2^n$, 就确定了一个量子态波函数, 其中正交基的位数 $|x_1 x_2 \dots x_n\rangle$ 是 n 个, 由于每一位中的值可以分别取 0 或 1, 所以一共可以组成 2^n 个本征态。Xu 等^[8]通过对受限玻尔兹曼机 (restricted Boltzmann machine, RBM) 网络权值的训练, 逼近组成波函数基态 $|i\rangle$ 的概率分布, 实现网络权值参数对输入模式出现概率的记忆。此时神经网络的输入节点数是 n 个正交基, 网络的输入数据是不同的 2^n 个状态, 对应得到是每一次落在相应某个状态上的数值, 就像抛硬币落下来看是正面或反面的结果一样。多次观测结果就是对落入每一个状态上数值的叠加, 此结果做出来的曲线就是波函数在每一个状态上的概率分布。RBM 具有能够对原始输入网络数据集的概率分布进行逼近的功能, 并且它还是通过一种无监督的方式, 通过一个能量函数, 修正网络参数 $\lambda=(b, c, W)$ 重建数据集的概率分

布。RBM这种网络函数的逼近功能,使得网络的结构从 2^n 个随量子位 n 指数增加输入节点,减少为 n 个多项式增加的输入节点,极大地减少了网络构造的复杂度。

3.2.1 基于RBM的量子波函数纯态表示

一个由 n 量子比特构成的量子波函数,是由 2^n 个正交基的叠加构成,它所有正交基系数的平方之和等于1。采用一个具有 n 个二值神经元(对应 n 个量子比特)可视层以及隐含层具有 m 个二值神经元的RBM网络表示量子波函数。一个 n 量子比特的波函数 $|\Psi\rangle$ 的一般表达式为

$$|\Psi\rangle = \sum_{\mathbf{x}} \Psi(\mathbf{x})|\mathbf{x}\rangle \quad (23)$$

其中系数为

$$\Psi(\mathbf{x}) = \langle \mathbf{x} | \Psi \rangle \quad (24)$$

式中, $|\mathbf{x}\rangle = |\sigma_1^z, \dots, \sigma_n^z\rangle$ 为 n 量子位系统所选定的一个参考正交基底; $\sigma_j^z \in \{0, 1\}$, $\sigma_j^z = 0$ 表示量子处于自旋向上状态 $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\sigma_j^z = 1$ 表示量子处于自旋向下状态 $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $|1\rangle$ 和 $|0\rangle$ 构成一个量子位的正交完备基矢,分别对应于 Bloch 矢量的沿 z 轴向上和沿 z 轴向下的2个状态,称为计算基矢。

单比特量子系统参考基底 $|\sigma_i^z\rangle$ 仅包含有 $|1\rangle$ 和 $|0\rangle$ 这2个不同的计算基矢, n 比特量子系统的参考正交基底 $|\mathbf{x}\rangle$ 包含有 2^n 个不同的计算基矢,所以波函数的系数个数,就是待估计的参数也为 2^n 个。

利用RBM可以将一个待估计的 n 量子位波函数 $|\Psi\rangle$ 的各系数 $\Psi(\mathbf{x})$ 表示为 $\psi_{\lambda,\mu}(\mathbf{x})$ ^[10]:

$$\psi_{\lambda,\mu}(\mathbf{x}) = \sqrt{p_{\lambda}(\mathbf{x})} e^{i\Phi_{\mu}(\mathbf{x})/2} \quad (25)$$

式中, $\mathbf{x} = (\sigma_1^z, \dots, \sigma_n^z)$, $\sigma_j^z = 0$ 或 1 , $p_{\lambda}(\mathbf{x})$ 表示一个带有RBM完备参数 λ 的波函数的振幅; $\Phi_{\mu}(\mathbf{x})$ 表示一个带有RBM完备参数 μ 的波函数的相位, $\Phi_{\mu} = \log p_{\mu}(\mathbf{x})$ 。

RBM表示的波函数为 $|\psi_{\lambda,\mu}\rangle = \sum_{\mathbf{x}} \psi_{\lambda,\mu}(\mathbf{x})|\mathbf{x}\rangle$ 。

从(25)式中可以看出,波函数的振幅 $p_{\lambda}(\mathbf{x})$ 及其相位 $\Phi_{\mu}(\mathbf{x})$ 是重构量子态的关键,它们与RBM的关系如下:RBM网络结构如图3所示,它是一个具有双向连接的随机处理单元的2层网络^[11],分别称为

可见层(visible layer)和隐含层(hidden layer),每一层的输出是另一层的输入,其中,可见层的输入节点数为 n ,其神经元用向量 $\mathbf{v}=(v_1, \dots, v_j, \dots, v_n)$ 表示,且 $v_j \in \{0, 1\}$;隐含层输入节点数为 m ,其神经元用 $\mathbf{h}=(h_1, \dots, h_i, \dots, h_m)$ 表示, $h_i \in \{0, 1\}$ 。连接2层之间的权重由矩阵 \mathbf{W} 来描述,其中元素 W_{ij} 是连接 h_i 到 v_j 之间的权重; $\mathbf{b}=(b_1, \dots, b_n)$ 和 $\mathbf{c}=(c_1, \dots, c_m)$ 分别表示可见层 \mathbf{v} 和隐藏层 \mathbf{h} 的偏置向量。RBM的完备参数由 $\lambda = (\mathbf{b}, \mathbf{c}, \mathbf{W})$ 表示,所有的参数都是实数。

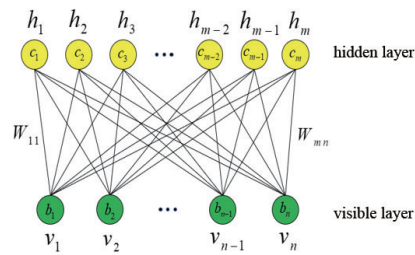


图3 RBM网络结构

对于一个由可见层和隐含层状态 (\mathbf{v}, \mathbf{h}) 组成的RBM网络,在信息双向动态流动过程中,网络中的内部能量将逐渐减少并平衡在最小点。RBM网络的能量函数 $E_{\lambda}(\mathbf{v}, \mathbf{h})$ 可以被定义为^[12]

$$E_{\lambda}(\mathbf{v}, \mathbf{h}) = -\sum_{ij} W_{ij} h_i v_j - \sum_j b_j v_j - \sum_i c_i h_i \quad (26)$$

式中, $i \in [1, m]$, $j \in [1, n]$ 。

根据 $E_{\lambda}(\mathbf{v}, \mathbf{h})$ 可以得到网络状态 (\mathbf{v}, \mathbf{h}) 的联合概率分布(玻尔兹曼分布) $p_{\lambda}(\mathbf{v}, \mathbf{h})$ 为

$$p_{\lambda}(\mathbf{v}, \mathbf{h}) = \frac{1}{Z_{\lambda}} e^{-E_{\lambda}(\mathbf{v}, \mathbf{h})} \quad (27)$$

式中, Z_{λ} 为归一化因子(也称为配分函数): $Z_{\lambda} = \sum_{\mathbf{v}, \mathbf{h}} e^{-E_{\lambda}(\mathbf{v}, \mathbf{h})}$ 表示对所有可能的 \mathbf{v}, \mathbf{h} 进行累计,一共包含 2^{n+m} 项。

由(27)式可得,可见层状态向量 \mathbf{v} 的概率分布 $p_{\lambda}(\mathbf{v})$ 可以通过求解在所有可能的隐藏层向量 \mathbf{h} 上边缘化联合概率分布 $p_{\lambda}(\mathbf{v}, \mathbf{h})$ 获得:

$$p_{\lambda}(\mathbf{v}) = \sum_{\mathbf{h}} p_{\lambda}(\mathbf{v}, \mathbf{h}) = \frac{1}{Z_{\lambda}} e^{-\varepsilon_{\lambda}(\mathbf{v})} \quad (28)$$

式中, $\varepsilon_{\lambda}(\mathbf{v})$ 为有效能量:

$$\varepsilon_{\lambda}(\mathbf{v}) = -\sum_j b_j v_j - \sum_i \ln \left(1 + e^{c_i + \sum_j W_{ij} v_j} \right) \quad (29)$$

因为归一化因子 Z_λ 是一个常数, 所以

$$Z_\lambda = \sum_{v,h} e^{-E_\lambda(v,h)} = \sum_{v,h} e^{-E_\lambda(v,h)} \quad (30)$$

对于网络特定的参数所组成的 $\lambda=(b, c, W)$, 因为同一层中各神经元不相连, 所以各层神经元向量的条件概率分布等于每个神经元的激活概率的乘积, 即

$$\begin{aligned} p_\lambda(\mathbf{h}|\mathbf{v}) &= \prod_{i=1}^m p_\lambda(h_i|\mathbf{v}) \\ p_\lambda(\mathbf{v}|\mathbf{h}) &= \prod_{j=1}^n p_\lambda(v_j|\mathbf{h}) \end{aligned} \quad (31)$$

各层神经元激活的概率为

$$\begin{aligned} p_\lambda(h_i = 1|\mathbf{v}) &= \mathcal{S}\left(c_i + \sum_j W_{ij}v_j\right) \\ p_\lambda(v_i = 1|\mathbf{h}) &= \mathcal{S}\left(b_j + \sum_j h_j W_{ij}\right) \end{aligned} \quad (32)$$

式中, 函数 $\mathcal{S}(x) = \frac{1}{1 + e^{-x}}$ 被称为 sigmoid 函数。

从(32)式可以看出: RBM 的可见层和隐含层分别含有一个 S 型激活函数, 对于各层的输入状态 (\mathbf{v}, \mathbf{h}) , RBM 各层被激活的概率都是 $[0, 1]$ 之间的一个数, 所以, RBM 是一个具有生成 $[0, 1]$ 概率功能的 2 层人工神经网络。RBM 能够对原始输入网络数据集的概率分布进行逼近, 通过一种无监督的方式, 修正参数 $\lambda=(b, c, W)$ 来重建数据集的概率分布。

3.2.2 基于 RBM 的量子密度矩阵混合态表示

一个混合态的密度矩阵总可以在更大的复合希尔伯特空间中找到纯态的密度矩阵, 这个过程叫做纯化 (purification)。如果想利用 RBM 表示混合态密度矩阵, 可以通过对纯化后纯态的密度矩阵求偏迹来得到想要的混合态的密度矩阵, 被称为约化密度矩阵, 来得到 RBM 表示的混合态。

对于 n 量子位的神经网络密度矩阵表示关系定义为一个从输入投影测量的样本数据到待重构的密度矩阵的映射 ρ_θ , 其中, θ 为其关系中的所有参数。对于由神经网络的输出描述量子态密度矩阵, 必须同时具有单位迹 $\text{tr}_\sigma[\rho_\theta]=1, \forall |\sigma\rangle$, 厄米共轭 $\rho_\theta = \rho_\theta^\dagger$, 并且是半正定 $\langle \sigma | \rho_\theta | \sigma \rangle \geq 0, \forall |\sigma\rangle$ 。这些约束条件可以通过使用一个 n_a 辅助自由度的系统 $\mathbf{a} = (a_1, \dots, a_i, \dots, a_{n_a}), a_i \in \{0, 1\}$, 纯化其希尔伯特空间构

造神经网络来满足, n_a 辅助自由度代表辅助系统的量子比特数为 n_a , 纯化得到的复合系统的波函数为 $|\psi_\theta\rangle = \sum_{\sigma a} \psi_\theta(\sigma) \otimes |a\rangle$, 其密度矩阵为 $\rho_\theta^{\sigma \oplus a} = |\psi_\theta\rangle \langle \psi_\theta|$ 。一个标准的采用波函数表示的量子纯态 RBM 网络, 包含 2 层随机二进制单元, 一个由计算基矢组成输入的可见层 $\sigma, \sigma=(\sigma_1, \dots, \sigma_j, \dots, \sigma_n), \sigma_j \in \{0, 1\}$; 一个隐藏层 $\mathbf{h}, \mathbf{h}=(h_1, \dots, h_i, \dots, h_m), h_i \in \{0, 1\}$ 。将一个 n_a 辅助自由度的辅助单元 \mathbf{a} 嵌入到神经网络的隐藏层 \mathbf{h} 中, 也就是将隐藏层的节点数由 m 个扩大到 $(m+n_a)$ 个, 如图 4^[13] 所示, 其中可见层表示输入状态 σ , 而其他 2 层分别是隐藏层 \mathbf{h} 和辅助单元层 \mathbf{a} ^[13]。采用图 4 中的权值参数表示的玻尔兹曼概率分布为 $p_\theta(\sigma, \mathbf{h}, \mathbf{a})$, 其中, θ 为网络参数: $\theta = \{W_\theta, U_\theta, b_\theta, c_\theta, d_\theta\}$; W_θ, U_θ 为网络权重; $b_\theta, c_\theta, d_\theta$ 为网络的偏置。为了表示纯化后系统的纯态波函数, 需要 2 组 RBM 网络分别表示振幅和相位, 对此, 使用 2 组参数 $\theta=\lambda$ 或 μ 定义复合系统的量子态; λ, μ 分别描述振幅和相位, 即有 $\lambda=\{W_\lambda, U_\lambda, b_\lambda, c_\lambda, d_\lambda\}$ 和 $\mu=\{W_\mu, U_\mu, b_\mu, c_\mu, d_\mu\}$ 。

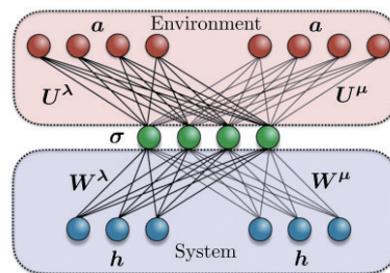


图 4 RBM 的网络结构及各参数变量表示

由神经网络表示的玻尔兹曼概率分布为

$$p_\theta(\sigma, \mathbf{h}, \mathbf{a}) = e^{h^T W_\theta \sigma + a^T U_\theta \sigma + b_\theta^T \sigma + c_\theta^T h + d_\theta^T a} \quad (33)$$

通过在隐藏变量 \mathbf{h} 对概率分布 $p_\theta(\sigma, \mathbf{h}, \mathbf{a})$ 求和, 可以获得描述复合 (纯) 系统的分布 $p_\theta(\sigma, \mathbf{a})$

$$p_\theta(\sigma, \mathbf{a}) = e^{\sum_{i=1}^m \log(1 + e^{W_\theta^{(i)} \sigma + c_\theta^{(i)}}) + a^T U_\theta \sigma + b_\theta^T \sigma + d_\theta^T a} \quad (34)$$

式中, $W_\theta^{(i)}$ 和 $c_\theta^{(i)}$ 表示权重矩阵和偏置向量的第 i 行, $\theta=\lambda$ 或 μ 。

由此可以得到纯化后的量子态波函数 $\psi_{\lambda\mu}(\sigma, \mathbf{a})$ 为

$$\psi_{\lambda\mu}(\sigma, \mathbf{a}) = Z_\lambda^{-\frac{1}{2}} \sqrt{p_\lambda(\sigma, \mathbf{a})} e^{i\phi_\mu(\sigma, \mathbf{a})} \quad (35)$$

其中, $\phi_{\lambda}(\boldsymbol{\sigma}, \boldsymbol{a}) = \log p_{\lambda}(\boldsymbol{\sigma}, \boldsymbol{a})/2$; $Z_{\lambda} = \sum_{\boldsymbol{\sigma}, \boldsymbol{a}} p_{\lambda}(\boldsymbol{\sigma}, \boldsymbol{a})$, 是一个常数。

通过求辅助系统的偏迹 $\rho_{\lambda\mu} = \text{tr}_{\boldsymbol{a}} \{ |\psi_{\theta}\rangle \langle \psi_{\theta}| \}$ 就能直接获得神经网络表示的待求密度矩阵元素 $\rho_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$:

$$\rho_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = \sum_{\boldsymbol{a}} \psi_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{a}) \psi_{\lambda\mu}^*(\boldsymbol{\sigma}', \boldsymbol{a}) \quad (36)$$

式中, $\psi_{\lambda\mu}^*(\boldsymbol{\sigma}', \boldsymbol{a})$ 为 $\psi_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{a})$ 的共轭复数。

由于辅助单元嵌入在网络的潜在空间中, 可以精确地执行(36)中等式中的求和, 获得

$$\rho_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = Z_{\lambda}^{-1} \tilde{\rho}_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \quad (37)$$

式中, $\tilde{\rho}_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ 为归一化之前矩阵, $\rho_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ 为归一化之后的密度矩阵, $\tilde{\rho}_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ 的矩阵元素为

$$\begin{aligned} \tilde{\rho}_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') &= e^{\Gamma_{\lambda}^{+1}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') + i\Gamma_{\mu}^{-1}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') + \Pi_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')} \\ \Gamma_{\theta}^{\pm 1}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') &= \frac{1}{2} \left(\sum_i \log(1 + e^{w_{\theta}^{i1}\sigma + c_{\theta}^{i1}}) \pm \sum_i \log(1 + e^{w_{\theta}^{i1}\sigma' + c_{\theta}^{i1}}) + \mathbf{b}_{\theta}^{\top}(\boldsymbol{\sigma} \pm \boldsymbol{\sigma}') \right) \\ \Pi_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') &= \sum_k \log \left(1 + \exp \left[\frac{1}{2} U_{\lambda}^{[k]}(\boldsymbol{\sigma} + \boldsymbol{\sigma}') + \frac{i}{2} U_{\mu}^{[k]}(\boldsymbol{\sigma} - \boldsymbol{\sigma}') + \mathbf{d}_{\lambda}^{[k]} \right] \right) \end{aligned}$$

以及 $\theta = \lambda$ 或 μ 。

(37)式就是最终所要求的密度矩阵 $\tilde{\rho}_{\lambda\mu}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ 表达式。

考虑一个由正实系数组成的 2 量子比特波函数, 其波函数是由 $2^n = 2^2 = 4$, 也就是 4 个基矢的叠加组成, 采用 RBM 表示波函数的网络输入节点为 $n=2$; 波函数为 $|\psi\rangle = \sum_{\boldsymbol{x}} \psi_{\lambda}(\boldsymbol{x}) |\boldsymbol{x}\rangle$, 其中 $\boldsymbol{x} = 00, 01, 10, 11$ 。对于给定一个由 Ising 的基态波函数组成的波函数的估计, 就是对一个 $|\Psi\rangle = a_1|00\rangle + a_2|01\rangle + a_3|10\rangle + a_4|11\rangle$ 中 4 个系数 $a_i, i=1, 2, 3, 4$ 的估计。网络输入数据由 $\{0, 1\}$ 构成, 为选择基底 $|\boldsymbol{x}\rangle$ 下的投影测量得到的测量结果, 理论上无穷多次的测量数据组成 n 量子位波函数所有本征态系数的概率分布。通过训练 RBM 网络中权值与偏置参数, 可以在给定的有限次的输入数据下, RBM 通过无监督训练, 达到对网络输入数据集概率分布的逼近。

通过网络设计与训练, 前向神经网络、深度自编码器、生成对抗网络以及递归网络等都具有数据自生成以及概率模拟等功能, 可以借助于 RBM 进行量子态的表示, 之后就可以结合各深度学习网络所具有的功能与特点, 采用合适的算法对网络的权值进行训练来达到对量子态重构的目标。以 RBM 为例, 为了衡量输入数据的预测概率分布和真实分布之间的距离, RBM 使用 KL 散度 (Kullback-Leibler divergence) 度量 2 个分布的相似性。KL 散度测量的是 2 条曲线的非重叠区域或是发散区域, RBM 的优化算法尝试最小化这些区域, 所以当共享权重与第 1 个隐藏层的激活值相乘时就可以得出原始输入的近似。RBM 的快速学习算法, 就是对比散度 (contrastive divergence, CD) 算法。对比散度是用来计算梯度 (该斜率表示网络权重与其误差之间的关系) 的方法, 该算法中通过预训练学习数据, 获得 v_0 初始值后, 只需要进行吉布斯采样 (Gibbs sampling) 1 到 2 次, 就能够完成最后的概率逼近。

4 结论

采用神经网络对量子态估计, 有可能将 2^n 的计算复杂度减少为 n 多项式的计算复杂度, 从而减少训练网络时间, 加快收敛速度, 提高逼近精度。本文根据波函数和密度矩阵不同的数学和神经网络的表示, 以及各自所具有的特性, 推导出相应神经网络输入/输出之间的非线性映射关系, 为采用神经网络进行量子态估计提供了数学上的网络函数关系以及网络训练用数据生成的设计理论。

参考文献 (References)

- [1] D'Alessandro D. On quantum state observability and measurement[J]. Journal of Physics A: Mathematical and General, 2003, 36(37): 9721-9735
- [2] Fano U. Description of states in quantum mechanics by density matrix and operator techniques[J]. Reviews of Modern Physics, 1957, 29(1): 74-93.
- [3] Smith A, Riofrío C A, Anderson B E, et al. Quantum

- state tomography by continuous measurement and compressed sensing[J]. *Physical Review A*, 2013, 87(3): 030102.
- [4] Li K, Zhang J, Cong S. Fast reconstruction of high-qubit quantum states via low rate measurements[J]. *Physical Review A*, 2017, 96(1): 012334.
- [5] Zhang J J, Cong S, Ling Q, et al. Quantum state filter with disturbance and noise[J]. *IEEE Transactions on Automatic Control*, 2020, 65(7): 2856–2866.
- [6] Xin T, Lu S R, Cao N, et al. Local-measurement-based quantum state tomography via neural networks[J]. *NPJ Quantum Information*, 2019, 5(1): 109.
- [7] Lohani S, Kirby B T, Brodsky M, et al. Machine learning assisted quantum state estimation[J]. *Machine Learning: Science and Technology*, 2020, 1(3): 035007.
- [8] Xu Q, Xu S Q. Neural network state estimation for full quantum state tomography[J]. *arXiv e-prints*, 2018: 1811.06654.
- [9] 潘成威, 丛爽. 基于宽度神经网络的2比特量子态的估计[C]//第22届中国系统仿真年会. 合肥: 中国科学技术大学, 2021: 80–84.
- [10] Zyczkowski K, Penson K A, Nechita I, et al. Generating random density matrices[J]. *Journal of Mathematical Physics*, 2011, 52(6): 062201.
- [11] 蒋文杰, 邓东灵. 神经网络量子态及其应用[J]. *物理*, 2021, 50(2): 76–83.
- [12] Hinton G E. Training products of experts by minimizing contrastive divergence[J]. *Neural Computation*, 2002, 14(8): 1771–1800.
- [13] Torlai G, Melko R G. Latent space purification via neural density operators[J]. *Physical Review Letters*, 2018, 120(24): 240503.

Mathematical and neural networks representations of quantum states

CONG Shuang

Department of Automation, University of Science and Technology of China, Hefei 230027, China

Abstract With wide application of deep learning, applications of data self-generation and probability simulation of neural network models in quantum state reconstruction and estimation have attracted people's attention. In this paper, from various mathematical representations of quantum states, we derive different representations of quantum states in neural networks. Nonlinear mapping relationships between input/output on corresponding neural network structures are deduced from the relationship between different physical variables of quantum state. This work provides a theoretical design basis of network function relationship and data generation for using different types of neural network models to realize quantum state estimation by means of their own data self-generation and probability simulation function.

Keywords quantum state; neural network representation; probabilistic simulation; density matrix reconstruction ●



(责任编辑 王志敏)