

## 量子机器学习

陆思聪<sup>1</sup>, 郑 昱<sup>2</sup>, 王晓霆<sup>3</sup>, 吴热冰<sup>1,4†</sup>

(1. 清华大学 自动化系, 北京 100084; 2. 清华大学 微纳电子系, 北京 100084;

3. 成都电子科技大学 基础与前沿研究院, 四川 成都 611731;

4. 清华信息科学与技术国家实验室 量子信息科学与技术研究中心, 北京 100084)

**摘要:** 人工智能和量子物理是上世纪发展起来的两个截然不同但又影响深远的学科. 近年来, 它们在数据科学方面的结合引起了学术界的高度关注, 形成了量子机器学习这个新兴领域. 利用量子态的叠加性, 量子机器学习有望通过量子并行解决目前机器学习中数据量大, 训练过程慢的困难, 并有望从量子物理的角度提出新的学习模型. 目前该领域的研究还处于探索阶段, 涵盖内容虽然广泛, 但还缺乏系统的梳理. 本文将从数据和算法角度总结量子机器学习与经典机器学习的不同, 以及其中涉及的关键加速技巧, 针对数据结构(数字型、模拟型), 计算技巧(相位估计、Grover搜索、内积计算), 基础算法(求解线性系统、主成分分析、梯度算法), 学习模型(支持向量机、近邻法、感知器、玻尔兹曼机)等4个方面对现有研究成果进行综述, 并建议一些可能的研究方向, 供本领域的研究人员参考.

**关键词:** 量子力学; 量子计算; 机器学习; 深度学习; 神经网络

中图分类号: O413.1 文献标识码: A

## Quantum machine learning

LU Si-cong<sup>1</sup>, ZHENG Yu<sup>2</sup>, WANG Xiao-ting<sup>3</sup>, WU Re-bing<sup>1,4†</sup>

(1. Department of Automation, Tsinghua University, Beijing 100084;

2. Department of Micro-Nano Electronics, Tsinghua University, Beijing 100084;

3. Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu Sichuan 611731;

4. Center for Quantum Information Science and Technology, Tsinghua National Laboratory for information science and technology, Beijing 100084)

**Abstract:** Artificial intelligence and quantum physics are two most influential disciplines developed in the last century. Recent years, their marriage in data science attracted much attention, forming the frontier field of quantum machine learning. Exploiting quantum superposition, quantum machine learning provides the hope for resolving difficulties in big data and training process, and new learning models illuminated by quantum physics. So far, the field is still in its infancy. Although many problems had been explored, a systematic theory is still lacking. This review will summarize the major differences in data structure and algorithms between classical and quantum machine learning, as well as the key acceleration techniques. Several topics will be covered, including the data structure (digital and analog), computation skills (phase estimation, Grover search and inner product), fundamental algorithms (linear equation, principle component analysis and gradient algorithms) and several learning models (supporting vector machine, nearest-neighbor method, perceptron networks and Boltzmann machine). Finally, we propose several promising directions in this field.

**Key words:** quantum mechanics; quantum computation; machine learning; deep learning; neural network

### 1 引言(Introduction)

机器学习, 尤其是基于人工神经网络的深度学习近年来得到了迅猛发展<sup>[1]</sup>. 而在量子信息科学领域, 它与量子计算技术的结合也正在成为一个飞速发展的研究方向. 事实上, 机器学习的思想很早就应用于量子力学系统的优化控制中, 并在量子化学和物理实验的广泛应用中获得巨大成功<sup>[2]</sup>. 近年来, 随着深度

学习的流行, 各种机器学习算法被用来探索量子多体物理中许多难于解析分析的问题, 并得到了许多有趣的结果.

机器学习与量子物理结合的另一个自然思路是利用量子状态的叠加和量子算法的加速, 来解决当前数据科学中数据量巨大, 训练过程缓慢的困难. 这方面的研究在量子计算的发展初期实际已经有人开始研

收稿日期: 2017-09-08; 录用日期: 2017-11-22.

†通信作者. E-mail: rbwu@tsinghua.edu.cn.

本文责任编辑: 崔巍.

国家自然科学基金项目(61374091, 61773232)资助.

Supported by National Natural Science Foundation of China (61374091, 61773232).

究,但是限于实验条件和当时学术界对量子计算发展前景的困惑,并没有得到长足的发展.近年来,随着量子计算机在计算规模和稳定性的突破,基于量子算法的机器学习重新得到关注,并成为一个迅速发展的研究方向<sup>[3]</sup>.

事实上,从经典-量子的二元概念出发可以将机器学习问题按照数据和算法类型的不同分为4类(如表1),本文将对C-Q(利用经典算法解决量子物理问题)进行简单介绍,然后重点对Q-C(利用量子算法加速机器学习)进行综合讨论,因为本文认为后者更体现量子机器学习的本质所在,对未来机器学习的发展推动作用更为巨大.Q-Q是一个开放的领域,也在此不做评论.

表1 机器学习的分类

Table 1 Classification of machine learning problems

简称	算法类型	数据类型	应用举例
C-C	经典	经典	传统机器学习
C-Q	经典	量子	用机器学习解决量子多体物理问题;量子优化控制
Q-C	量子	经典	量子支持向量机等
Q-Q	量子	量子	量子反馈控制

量子系统控制中的一个基本问题是对控制对象的建模,即对系统的哈密顿量以及确定扰动,噪声等参数特征进行辨识,这些问题都属于C-Q的范畴.例如,研究人员提出使用贝叶斯推断中的似然函数对量子系统的哈密顿量进行学习,并进行了实验验证,结果表明学习算法具有一定的鲁棒性,即使假设模型结构中有缺项,学习结果仍是实际模型的最佳近似<sup>[4-5]</sup>.

基于一定量的量子测量数据出发判断和确定其未知量子态的性质,例如如何判别一个量子态是分离还是纠缠的,也可以看作一类机器学习问题.研究人员提出用监督式学习的方法对量子态做分类,容易理解,直接使用量子模型,量子数据的全量子方式,其分类

正确率比使用经典模型和经过量子随机存储器(quantum random access memory, QRAM)调用转换的量子数据的半经典方式更高.其后,他们提出不需要使用QRAM的量子态分类方法,且算法所需的经典存储器规模仅随训练比特数以对数速度增长<sup>[6]</sup>.此外,他们还专门研究了光子相干态的分类问题,发现对训练样本和反射信号进行联合全局测量,比先对训练集进行基于高斯测量的幅值估计,再对反射信号进行状态判别的识别率要高<sup>[7]</sup>.对于量子系统演化所对应的酉变换,如何从有限样本集中进行辨识学习,也是一个重要的问题.该问题可以归结为一个双重优化问题,同时需要最优输入态和最优测量<sup>[8]</sup>.近年来,也有学者开始着手于研究用机器学习方法获取量子系统的属性和统计参数,如相变点,期望值等.该思想也可用于量子多体系统的模拟问题.在此不做重点描述.

利用量子理论改进机器学习(Q-C)的方法大致可以分为两种:1)通过量子算法使某些在经典计算机上不可计算的问题变为可计算的,从而大幅降低机器学习算法的计算复杂度,如量子退火(quantum annealing, QA)算法、Gibbs采样等;2)量子理论的并行性等加速特点直接与某些机器学习算法深度结合,催生出一批全新的量子机器学习模型,如张量网络、概率图模型(probabilistic graphical model, PGM)等.

## 2 数据的量子化表达和读写(Quantum data: expression, read-in and read-out)

量子机器学习的训练数据必须以某种可以为量子计算机识别的格式载入,经过量子机器学习算法处理以后形成输出,而此时的输出结果是量子格式的,需要经过测量读出来得到笔者所需的最终结果(如图1所示).因此,量子机器学习的复杂度与这3个过程都有关系.大致说来,基于量子相位估计的学习算法可以实现算法的指数加速,但是对数据的载入和完整结果的读出要求很高,而基于量子搜索的算法相对较复杂,但是数据的读入读出则相对比较容易.本章介绍常用的两种方式以及为之需要匹配的读入读出方式.



图1 量子机器学习的基本流程

Fig. 1 The basic process of quantum machine learning

### 2.1 数字型(Digital type)

众所周知,数字计算机中所有数据都是有限字长的.假设某训练数据向量为 $\vec{x} = [x_1 \cdots x_n]$ ,若每个元素都用长度为 $m$ 的二进制数据表示,则整个向量需要 $n \cdot m$ 个比特,在量子计算机中它们对应于 $N =$

$n \cdot m$ 个量子位.每个数据向量都对应于输入态希尔伯特空间的一个基矢 $|\vec{x}\rangle$ .

这其实是一个量子化的过程.类似的,输出数据 $\vec{y} = [y_1 \cdots y_q]$ 也可以用适当长度的二进制数表示相应,并通过量子化对应于输出态希尔伯特空间的一个

基矢. 对比下面将要提到的模拟型数据, 这种表示通过冯诺伊曼投影测量即可进行读出, 其代价是最低的. 进一步, 如果训练数据集中有  $M$  个数据, 自然的方式是用  $N \cdot M$  个量子位存储这些数据, 但这样耗费的资源无疑十分巨大, 特别在海量数据集的情形下. 为此可以充分利用量子系统的叠加特性, 将这些数据以“叠加”的方式存储在  $N$  个比特上, 用状态  $\sum_j c_j |\vec{x}_j\rangle$  表示. 这样可以存储最多  $2^N$  个数据向量, 并且数据处理可以并行进行, 其中单个或多个数据的提取根据数据特征(模式)采用 Grover 算法搜索得到, 因此称作量子关联存储器(quantum associative memory, QAM). 另一种方式是用  $\log(M)$  个比特标记数据下标(地址), 从而数据可以并行存储在  $\log(M) + n \cdot m$  个比特中, 其中第  $j$  个数据表示为  $|j\rangle |\vec{x}_j\rangle$ . 可以通过设计存储器根据寻址请求调用相应的数据, 以叠加方式  $\sum_j c_j |j\rangle |\vec{x}_j\rangle$  并行载入数据, 这是与经典数据载入非常不同的地方. 这里的概率幅系数  $c_j$  对应于经典数据处理中的随机采样概率分布, 但是以相干叠加的方式进行, 使得后续的数据处理可以并行进行, 设计巧妙的算法利用相干性带来的相长相消得到期望的结果, 以实现量子加速(quantum speed-up).

## 2.2 模拟型(Analog type)

训练数据还可以用半模拟的方式编码在量子态上. 假设  $\vec{x} = [x_1 \cdots x_n]$  已经归一化, 则可以用  $\log(n)$  个比特, 其中数据分量对应于用叠加态  $|\vec{x}\rangle = \sum_{k=1}^n x_k |k\rangle$  中各基矢的概率幅系数. 与数字型数据相比, 这种编码方式可以极大节省所需的量子位资源, 并避免数字化带来的圆整误差. 但它的最大问题是制备和读出的困难, 精确制备这样的量子态在实验上是非常困难的事, 而读出所需的精确层析技术也需要通过大量的重复实验才能准确重构这样的量子态. 由于所需重复样本随测控精度提高而上升, 因此也需要很大的时间复杂度代价.

与数字型数据类似, 如果训练数据集中有  $M$  个数据, 可以用  $\log(M)$  个比特表示下标, 从而用  $\log(n \cdot M)$  个比特表示整个数据集, 而且数据可以以  $\sum_j c_j |j\rangle |\vec{x}_j\rangle$  的方式并行加载和处理.

## 2.3 量子数据集的转换 (Transformation of quantum data sets)

从经典数据到量子数据的转换, 需要通过存储器来实现. 对于数字型数据, 其实对应于一个量子化的数字逻辑电路, 或者一个量子子程序, 通过执行该子程序将寄存器的状态制备到训练数据对应的状态. 这个子程序是模块化的, 由于在机器学习中会经常随机抽取或者检索该数据集中的数据, 因此它常常作为一

个 oracle 供 Grover 算法调用, 而调用的次数则成为衡量机器学习算法复杂度的重要指标之一. 这类存储器具有专用名称—量子关联存储器(quantum associative memory, QAM)<sup>[9]</sup>.

显而易见, 模拟型量子数据是最节省资源的存储方式, 但是从物理实现的角度上, 需要对量子寄存器进行精确地状态制备和存储, 利用所谓的 QRAM 调用<sup>[10]</sup>.

$$\sum_j c_j |j\rangle \xrightarrow{\text{QRAM}} \sum_j c_j |j\rangle |x_j\rangle. \quad (1)$$

在训练过程中, 还应该能够高效地根据访问地址调用所需数据, 这些对量子系统的控制, 以及量子计算机体系结构和算法设计提出了重要挑战. 此外, 数据的读出是另外一个困难问题, 因为在这种情况下, 通常需要对寄存器状态的最终值进行测量读出, 而这需要大量耗时的状态层析, 如果不精心设计的话, 所需代价甚至可能超过存储节省的资源.

## 3 量子计算的加速技巧 (Speed-up skills of quantum computation)

### 3.1 相位估计 (Phase estimation)

相位估计指对于一个特征向量是  $|u\rangle$  的酉算符  $U$ , 估计它相应的特征值  $e^{2\pi i \varphi}$ , 它是 Shor 算法, 解线性方程等很多量子算法的核心部分. 量子相位估计被广泛运用的根本原因是由于它数字型和模拟型数据的桥梁.

相位估计中的关键部分是量子傅里叶变换<sup>[11]</sup>. 都知道离散的傅里叶变换是把一组长度为  $N$  的向量  $[x_1 \cdots x_n]$  变换为  $[y_1 \cdots y_n]$ , 其中

$$y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{\frac{2\pi i j k}{N}}. \quad (2)$$

类似地, 量子傅里叶变换是一个线性的酉变换, 它将一组正交基  $\{|0\rangle, \dots, |N-1\rangle\}$  变换为另一组正交基:

$$|k\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{\frac{2\pi i j k}{N}} |j\rangle. \quad (3)$$

相比于经典快速傅里叶变换, 用量子线路实现的量子傅里叶变换可以实现  $O[N \log N]$  到  $O[\log^2 N]$  的指数加速. 利用量子傅里叶变换, 相位估计算法可以将待估计的相位转化到由  $t$  个量子比特编码的数字量上, 其中  $2^{-t}$  对应于估计精度, 这样就将  $\varphi$  以模拟信息的形式编码到辅助比特上, 最后经过反傅里叶变换就得到了  $\varphi$ .

### 3.2 Grover 搜索 (Grover search)

很多数据处理算法都有遍历或搜索数据的需求, 比如要找到  $N$  个数据中的最小值. 在经典计算机上, 这需要  $O(N)$  次操作才能找到, 因为最糟糕的情况是

最后一次操作才找到最大值. 但是奇妙的是, Grover在1996年提出一个量子搜索算法<sup>[13]</sup>, 只需要 $O(\sqrt{N})$ 次就能找到答案. Grover算法的思路是将所有的数据以相干叠加在一起, 并设计量子线路(Grover oracle)使目标数据对应的概率幅反相, 然后将所有态的幅值以平均态为轴进行翻转. 如此迭代这两个酉操作, 正确解的幅值就慢慢放大, 并在 $\sqrt{N}$ 次左右达到最大, 这样如果对量子态进行测量, 就可以以接近1的概率得到目标解. 理论证明, 对于这样的问题Grover算法是最优的方案<sup>[14]</sup>.

正如其他的量子算法一样, Grover算法也是以小于1的概率给出的正确解. 如果一个问题的搜索空间是 $N$ , 正确解的个数是 $K$ , 则Grover算法需 $\pi\sqrt{N/K}/4$ 次迭代可以以最大概率给出正解. 如果不知道正确解的个数, 则算法最多需要 $\pi(2 + \sqrt{2})\sqrt{N}/4$ 次迭代. 第一个量子支持向量机<sup>[15]</sup>就是通过Grover算法来寻找优化目标的最小值.

### 3.3 内积运算(Inner product)

求取向量内积是很多机器学习算法中的基本运算, 例如支持向量机中核函数的运算. 针对模拟型数据, 量子算法可以通过引入幅值比特和测量, 将计算复杂度降到 $O(1)$ , 但其缺点是测量可能会破坏数据, 因此该算法通常在算法的最后读出时使用. 针对数字型数据, 可以用 $m$ 个量子比特编码, 通过量子傅里叶加法和量子傅里叶乘法器来得到内积的结果<sup>[12]</sup>, 那么内积运算所用的门复杂度是 $O(m^3)$ , 相比于经典运算能实现指数加速<sup>[16]</sup>. 在量子SVM算法中<sup>[17]</sup>, 这两种内积的运算都发挥作用, 其中数字型内积运算加速核函数的实现, 而模拟型内积运算用于读出最后的判别结果, 极大降低了运算复杂度.

## 4 数据预处理和训练过程(Preprocessing and training of data sets)

### 4.1 线性方程组算法与拟合(Linear systems and regression)

求解线性方程组既是一个基本的代数计算问题, 也是求解许多复杂问题的重要中间步骤. 该问题通常被表述为, 已知矩阵 $A_{n \times n}$ 和向量 $b_{n \times 1}$ , 求向量 $x_{n \times 1}$ , 使 $Ax = b$ (或 $Ax$ 在某种度量准则下最接近 $b$ ). 一般性问题在电子计算机上的时间复杂度(下文中如不特别说明, “复杂度”均指时间复杂度)介于 $O(n^2)$ 与 $O(n^3)$ 之间, 可统一记为Poly( $n$ ).

进一步, 在大量实用场景中, 通常可假设 $A$ 具有一定的稀疏性, 记其条件数为 $c$ , 并假设不需直接求出 $x$ , 而是需要求解与其相关的某些函数值, 如 $x^\dagger Bx$ ( $B_{n \times n}$ 为某个给定矩阵). 目前已发表的求解此问题最快的经典算法复杂度约为 $O(n\sqrt{c})$ , 当 $c$ 较小时, 可近似认为是 $O(n)$ .

在此基础上, 研究人员提出了一个用于求解稀疏线性方程组的量子算法<sup>[18-20]</sup>. 假设 $A$ 是厄米矩阵,  $b$ 是单位向量, 并将 $b$ 中的信息编码到量子态上, 即制备 $|b\rangle = \sum_{k=1}^n b_k |k\rangle$ . 然后用 $e^{iAt}$ 作用于 $|b\rangle$ , 这相当于将其分解到 $A$ 的特征空间, 并求出各特征值 $\lambda_k$ . 此时系统状态可近似写为 $\sum_{k=1}^n \beta_k |u_k\rangle |\lambda_k\rangle$ , 其中 $u_k$ 即为 $A$ 的特征向量, 张成 $A$ 的特征空间,  $|b\rangle$ 则可写为 $|b\rangle = \sum_{k=1}^n |u_k\rangle$ . 再通过变换 $|\lambda_k\rangle \rightarrow \frac{N}{\lambda_k} |\lambda_k\rangle$  ( $N$ 为归一化常数), 即可规避计算 $|\lambda_k\rangle$ 而直接得到量子态 $\sum_{k=1}^n \frac{\beta_k}{\lambda_k} |u_k\rangle = A^{-1}|b\rangle = |x\rangle$ . 此处需要特别注意的是, 该变换不是酉变换, 因此有一定的概率会失败.

关于计算结果, 如果需读出 $|x\rangle$ 中的每个分量, 则需运行上述过程至少 $n$ 次, 这显然并不节省计算资源. 但很多实际问题并不需要确切知道 $x$ 中每个分量的具体取值, 而是要求关于 $x$ 的某些函数值(例如分类问题中只需要知道分类的结果). 例如 $x^\dagger Bx$ 的形式, 只需将矩阵 $B$ 作为观测量, 在状态 $|x\rangle$ 下做测量平均, 即可得到二次型函数 $\langle x|B|x\rangle$ 的一个估计. 用此方法可得到 $x$ 的许多特征, 如归一化、状态空间分布权重、矩等.

该算法已经得到了实验验证: 2012年, 研究人员用基本量子门构建出一个用于求解 $4 \times 4$ 线性系统的7位量子线路图, 对其功能进行了仿真验证, 并对其计算复杂度进行了量化估计<sup>[21]</sup>; 2013年, 研究人员在该成果的基础上, 使用核磁共振方法完成了一组求解 $2 \times 2$ 线性系统的物理实验, 且所有结果量子态的保真度均在96%以上<sup>[22]</sup>; 这些实验为求解一系列与线性方程组有关的实际问题指明了方向, 可以作为实现许多量子算法的基本模块.

### 4.2 主成分分析(Principal component analysis)

主成分分析(principal component analysis, PCA)方法是对数据进行特征提取的常用方法. 假设数据已编码为一组 $n$ 维向量 $v_k, k = 1, 2, \dots, m$ , 那么它们实际上组成了一个 $n \times m$ 矩阵 $V_{n \times m} = [v_1 v_2 \dots v_m]$ , 此时数据的协方差矩阵可表示为 $C_{n \times n} = VV^\dagger$ . 这里需要注意的是, 根据协方差的定义,  $V$ 应该是去均值化后的样本集, 即各数据向量通常带有某种改变量的含义, 在此意义下,  $C$ 刻画的是数据间的相关关系, 特别是数据不同成分之间的相关关系, 主成分分析的算法命名正是由此而来.

经典主成分分析方法的本质是对协方差矩阵 $C$ 进行对角化处理. 容易看出,  $C^\dagger = (VV^\dagger)^\dagger = VV^\dagger = C$ , 因此 $C$ 是一个厄米矩阵, 一般情况下, 其特征向量构成一组完备正交基, 可用于表示各数据向量. 因此, 对 $C$ 进行特征值分解有 $C = \sum_{k=1}^n \lambda_k e_k e_k^\dagger$ , 其中不妨假

设 $|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_n|$ 为 $C$ 的 $n$ 个特征值,  $e_k$ 为与 $\lambda_k$ 对应的特征向量. 根据矩阵特征值和特征值分解的定义, 把数据投影到绝对值较大的特征值对应的特征向量方向上, 能使数据间的相关关系被更显著地表现出来; 与之相反, 舍弃数据在接近于零的特征值对应特征向量方向上的分量不会对数据的本质有太大影响. 因此, 显著非零的特征值对应的特征向量称为 $V$ 的主成分, 代表数据互相关联的主要方式. 将数据表示为主成分的线性组合既能实现降维, 又能消除测量获取过程中引入的噪声(假设数据的信噪比较高, 即在数据中真实值是“主成分”, 而噪声是次要成分). 经典PCA方法的计算复杂度为 $O(n^2)$ .

用量子算法处理经典数据的主成分分析问题, 即量子主成分分析(quantum PCA, QPCA)<sup>[23]</sup>, 其操作方法如下: 使用QRAM将随机选择的数据向量 $v_k$ 转换为量子态 $|v_k\rangle$ , 由于选择的随机性, 可以得到密度矩阵为 $P = \frac{1}{m} \sum_{k=1}^m |v_k\rangle\langle v_k|$ 的量子态, 将其与协方差矩阵 $C$ 对比发现, 量子形式的密度矩阵 $P$ 即是经典形式的 $C$ , 两者之间仅相差一个系数. 对数据集重复采样并使用密度矩阵求幂和相位估计算法, 即可得到数据向量的量子形式及其在主成分 $|e_k\rangle$ 上的分解 $\sum_{k=1}^s v_k |e_k\rangle |\lambda_k\rangle$ , 主成分向量也可通过测量得出. QPCA的计算复杂度为 $O(\log^2 n)$ (表示所需的量子位数为 $\log n$ ), 因此对经典算法具有指数加速效果.

### 4.3 量子梯度算法(Quantum gradient algorithm)

梯度算法是机器学习, 特别是训练人工神经网络中必不可少的优化工具, 因此如何在量子机器学习模型中有效地计算和利用梯度进行学习, 是一个重要的问题.

梯度算法的基本思路是, 对于损失函数 $L(w, \dots, w)$ , 其中 $x_k$ 是学习模型中需要调整的变量. 如果可以计算出梯度函数

$$\nabla L(w, \dots, w) = \left( \frac{\delta L}{\delta w_1}, \dots, \frac{\delta L}{\delta w_n} \right), \quad (4)$$

则可以以迭代的方式

$$w_k^{(l+1)} = w_k^{(l)} - \alpha_l \frac{\delta L}{\delta w_k}, \quad k = 1, \dots, n. \quad (5)$$

寻找适当的解. 由于函数沿负梯度方向总是下降的, 因此只要合理选择步长(学习率), 就可以逐步减小损失函数, 最终达到一个局部极小.

推广到量子算法, 梯度算法需要解决两个问题: 1) 如何有效的计算梯度; 2) 如何实现迭代过程. 对于梯度的计算, 2005年Jordan指出, 如果函数是一个由Oracle给出的布尔函数, 则利用量子并行性可以只需调用一次Oracle就完成梯度的计算, 而经典计算基于差分求解则需要至少 $n + 1$ 次调用<sup>[24]</sup>. 对于模拟型数

据, 最近的研究考虑了当函数具有多项式表达的时候, 根据梯度函数的解析表达式设计量子算法进行迭代计算. 由于函数的非线性, 以及为实现矩阵乘法, 算法中需要原始数据的多个拷贝以满足系统约化平均的需要, 并通过测量干预并进行conditional的操作, 而这两种操作都是不可逆的. 因此, 该梯度算法对于原始数据拷贝的数量要求较高, 不适于迭代次数较多的优化<sup>[25]</sup>.

## 5 量子机器学习算法(Quantum machine learning algorithms)

表2概述了目前文献中见到的一些典型量子机器学习算法, 及其所需资源和性能改善特征. 本章将把这些算法分别进行介绍.

表 2 主要量子机器学习算法汇总

Table 2 A summary of quantum machine learning algorithms

算法	Grover 搜索	量子加速	量子数据	泛化性能	实验
K-均值	需要	平方	不需要	无	无
K-近邻	需要	平方	不需要	数值分析	无
主成分分析	不需要	指数	需要	无	无
神经网络	需要	-	不需要	数值分析	有
支持向量机1	需要	平方	不需要	解析分析	无
支持向量机2	不需要	指数	需要	解析	有
回归	不需要	-	需要	无	无
Boosting	不需要	平方	不需要	解析	有
玻尔兹曼机	不需要	量子退火	不需要	概率生成	有

### 5.1 支持向量机(Supporting vector machine)

支持向量机(support vector machine, SVM)是最直观且易于理解的监督式机器学习模型之一. 其想法是通过一个最佳超平面将特征空间中的数据按类别标签分隔开来, 而衡量“最佳”的标准则是最大化分类间隔, 训练得出的参数即是分类超平面的法向量和截距.

通常把SVM的数学模型写为原凸优化问题的对偶问题, 即

$$\begin{aligned} \max_{\vec{\alpha}} L(\vec{\alpha}) &= \sum_{k=1}^N \alpha_k y_k - \frac{1}{2} \sum_{k,l=1}^N \alpha_k P_{k,l} \alpha_l, \\ \text{s.t.} \quad \sum_{k=1}^N \alpha_k &= 0, \quad \alpha_k y_k \geq 0, \quad k = 1, 2, \dots, N. \end{aligned} \quad (6)$$

而最佳分类超平面 $\vec{w}^\dagger \vec{x} + b = 0$ 的参数可表示为

$$\begin{cases} \vec{w}^\dagger = \sum_{k=1}^N \alpha_k \vec{x}_k, \\ b = \frac{y_k - \vec{w}^\dagger \vec{x}_k}{|\alpha_k \neq 0|}, \end{cases} \quad (7)$$

其中满足 $\alpha_k \neq 0$ 的对应的 $\vec{x}_k$ 即称为支持向量. 由于模型的核心矩阵 $P = (P_{k,l})_{N \times N}$ 中只含有样本向量两两

之间的内积, 可将其统一表示为  $P_{k,l} = \vec{x}_k^T \vec{x}_l = K(\vec{x}_k, \vec{x}_l)$ , 因此通过引入核函数的方式, 很容易将线性SVM模型推广到非线性场景. 自1995年被提出至今, SVM是应用最广泛的监督式机器学习模型之一, 可以处理图像分割、人脸、指纹识别、生物信息等各种分类决策问题.

与经典SVM类似, 量子支持向量机 (quantum SVM, QSVM) 同样是监督式量子机器学习模型的代表. 由于支持向量机模型本身的物理含义较明确, 量子算法的加速效果主要体现在寻找支持向量和计算核函数矩阵上.

2003年研究人员曾考虑借助Grover搜索算法的二次加速效果处理SVM模型的优化计算问题<sup>[15]</sup>, 即在  $m$  个数据向量组成的训练集中搜索出  $s$  个支持向量, 大约需要  $\sqrt{\frac{m}{s}}$  次迭代计算. 近年来, 最小二乘量子支持向量机 (least-square QSVM, LSQSVM) 模型由于能够有效利用量子基础代数算法加快计算速度而被广泛讨论. LSQSVM可以接受不同类型的训练数据, 如通过QRAM转换的经典数据或事先制备好的量子态, 然后通过相位估计和矩阵求逆 (HHL算法) 得出训练结果<sup>[17]</sup>. QSVM的训练和推断计算复杂度均为  $\text{Poly}(\log n)$ . 此外, 文献[26–27]等还讨论了多项式、径向基函数、高斯函数等非线性核QSVM.

QSVM已经在核磁共振平台上得到了实验验证: 2015年, 研究人员搭建了一个4位QSVM, 并对标准字符字体集中的数字6和数字9进行训练, 然后以MNIST中的“6”和“9”为对象进行手写数字图像二分类测试, 实验结果表明QSVM是物理可实现的<sup>[28]</sup>.

## 5.2 近邻法 (Nearest-neighbor method)

在机器学习理论中, 近邻法是比较特殊的一族: 其核心思想可派生出监督式和非监督式两类机器学习模型, 分别对应于字面上的“近邻法”和K-均值聚类 (由于在当前背景下, 监督式学习模型的应用更为广泛, 所以在不引起歧义的前提下, 以“近邻法”同时指代两类模型). 此外, 图像、语音、生物等多领域的应用实例已经证明, 基于距离的近邻法的分类性能普遍较好, 在某些应用背景下甚至能够超越SVM和人工神经网络 (artificial neural network, ANN). 不过, 近邻法的推断复杂度较高, 尤其当训练集规模较大时,  $O(mn)$  的计算复杂度往往无法保证实时性, 因此, 引入量子计算的加速效果具有较高的理论和应用价值.

基于距离的算法通常分为两步: 1) 计算测试样本与所有训练特征之间的距离, 在该步骤中, 使用量子算法能够起到多项式加速效果; 2) 在所有距离中选出最小的一个或数个, 并取出对应的训练样本, 在该步骤中, 使用Grover搜索的幅值估计算法能够起到平方

加速效果; 所以总地来说, 量子计算对近邻法能够起到多项式加速效果.

需要注意的是, 由于上述两个步骤各需要一次测量, 这在很大程度上局限了量子算法的加速效果, 文献[29]针对这个问题提出了一种不需要进行中间测量的相干算法. 该算法分为3步: 1) 对测试样本和每一个训练样本, 制备一个量子态, 用其幅值表示两者之间的某种距离度量; 2) 使用幅值放大算法将距离存储为一个量子位串, 而不对其进行测量; 3) 使用Grover搜索的变种DürHøyer最小化算法找出最小距离.

## 5.3 感知器与感知器神经网络 (Perceptron and neural networks)

感知器是机器学习中的一个基本分类器模型, 由此发展的SVM和以此为基本单元的神经网络都是运用十分广泛并有效的模型.

量子感知器有多个实现方式. 在1994年, Lewenstein提出了基于感知器的量子神经网络模型<sup>[31]</sup>, 用酉操作来表示权值矩阵, 通过训练数据不断调整这个酉操作中的参数, 实现学习的过程. 其中一种方案是基于无损线性光学搭建  $N$  个端口, 实现输入光算符转化为输出光算符这样一个酉操作; 第2种方案是利用  $N$  个端口的非线性光学网络来实现更为普适的转化. 除此之外, 感知机基于Grover算法进行训练<sup>[30]</sup>. 一种方式是在线训练, 每次判别训练数据是否被错误分类过程中使用量子搜索算法, 这样可实现平方加速; 另外一种方式是通过哈夫变换找到问题的对偶表达, 即将分类超平面由点  $w$  来描述, 而数据则转换为超平面表达, 这样可以证明当  $w$  服从高斯分布并具有足够多采样点时, 一定能找到可以正确分类所有点的  $w$ . 于是整个问题就变为了搜索问题, 可以用Grover来找到问题的解, 这里算法亦可实现平方加速.

对于感知器构成的神经网络, 也可以基于搜索算法来训练网络, 找到正确的权值<sup>[32]</sup>. 这时训练数据与其标签分别存于独立的寄存器中, 将所有权值的可能值叠加在一个向量  $w$  中, 向量  $p$  用来标定被正确分类的样本个数,  $p$  越大表示网络训练的越接近目标. 如此,  $p$  大于一定值时作为Grover算法中标记正确解的 oracle, 用来找到藏于  $w$  中的正确解. 由于搜索空间与权值的个数和编码数成指数关系, Grover搜索也只是平方加速, 一次全部搜索计算量太大, 因此可以采用分阶段搜索. 初始时刻所有的权值取随机值, 每次搜索一个神经元的权值向量, 当其正确度达到一定时就固定下来, 再随机选择下一个神经元训练, 直到所有的权值被确定下来. 值得一提的是, 能否基于经典机器学习中常用的梯度算法训练量子感知器神经网络的研究还很少, 因为算法迭代收敛的不可逆要求与量子酉操作的不可逆性是不相容的, 这方面的研究还是个

开放的课题。

#### 5.4 玻尔兹曼机(Boltzmann machine)

前述机器学习模型有一个共同点, 即不论SVM、近邻法还是感知器, 它们的输出结果都是一个最终的类别判断, 除此之外没有其他附加信息, 所以有时将这类模型称为判别式模型。与判别式模型相对应的是产生式模型, 这类模型通常并不直接给出分类结果, 而是给出测试样本被归到每一类的概率, 最后由用户进行判断, 玻尔兹曼机(Boltzmann machine, BM)就是一种典型的产生式模型。根据信息论的基本原理, 产生式模型提供的信息量比判别式模型显著地大, 所以其训练代价更高, 为达到与判别式模型相同的性能, 需要的训练集也更大。

以 $a_k, k = 1, 2, \dots, n$ 表示系统中的所有结点, 其中 $a_k \in \{1, -1\}$ , 则系统能量为

$$E = - \sum_{k=1}^n \sum_{l=1}^n w_{k,l} a_k a_l - \sum_{k=1}^n b_k a_k, \quad (8)$$

其中 $w_{k,l}, b_k$ 就是训练过程所需确定的参数。在BM模型中, 观测值(训练特征)的概率分布是以隐藏结点系统能量为参数的玻尔兹曼分布, 这也正是其名称的由来。通过最大似然方法可对该模型的优化问题进行梯度求解, 即正向迭代和反向迭代两个相互交错的迭代过程。

综上所述, 容易看出, SVM、近邻法和感知器等机器学习模型的核心思想较为直观, 符合人类进行分类操作的一般心理过程, 而BM的想法则更加“物理”, 它是一种基于系统能量的模型。因此, 与之相对应地, 量子玻尔兹曼机(quantum BM, QBM)通过将上述能量函数量子化, 并引入额外的非对易项<sup>[33]</sup>, 它可以基于量子退火算法(quantum annealing, QA)处理器来实现。目前的商用化QA处理器, 如D-Wave芯片等, 均未设计专门针对BM模型的计算功能, 无法直接产生玻尔兹曼分布, 故而需对其进行适当的修改和调整, 才能适用于QBM的训练和推断。此外, QBM模型还存在一些变种, 如限制量子玻尔兹曼机 (restricted QBM, RQBM)、有界量子玻尔兹曼机 (bound-based QBM, BQBM)等, 它们均是QBM深度化的重要参考。

#### 6 结论与展望(Conclusions and perspectives)

如前所述, 量子机器学习算法相对于在经典计算平台上执行的经典算法而言, 有如下明显优势: 1) 由于量子态的可叠加性, 量子算法可以在不增加硬件的基础上实现并行计算, 在此基础上利用量子相位估计、Grover搜索等算法, 可实现相对于完成同样功能的经典算法的二次甚至指数加速; 2) 将经典数据编码为量子数据, 并利用量子并行性进行存储, 可实现指数级节省存储硬件需求。

目前量子机器学习的研究绝大部分还处在理论层面, 而与之匹配的量子计算机硬件还远远落后。但是近年来基于超导电路、离子阱等方案的量子计算实验进展迅速。通过云计算共享技术, 小规模量子计算机(小于20量子位)已经进入测试阶段, 并有望在几年内实现更大规模的量子计算, 以真正超越现有的经典计算(即所谓的量子霸权, quantum supremacy)。一些专门的量子计算器, 如量子模拟器、QA处理器等也在逐步规模化、复杂化。事实上, 支持这些硬件设备正规化和更新换代的技术路线已经较为明确, 成品实现主要是时间问题。

人们普遍相信量子计算实用化已经为时不远, 因此针对性地开展量子算法应用正当其时, 量子机器学习作为这样一个交叉方向, 有很大的发展前景。作者认为, 以下将是近期内量子机器学习领域较为关注和亟待解决的问题:

- 量子计算的优势主要体现在数据处理而非读取输入方面, 某些情况下数据的获取(包括某些量子态的制备)甚至可能成为算法复杂度的主导部分, 所以关于量子数据的制备和读取有必要做专门研究。
- 在计算复杂度方面, 目前可以明确的是, 当问题规模足够大时, 量子算法相对于经典算法具有明显优势。但这些算法具体如何实现, 量子线路如何能够进一步优化, 相关的设计研究非常缺乏, 很难形成理论对实验的具体指导, 这个鸿沟尚待填平。
- 需要从基础理论角度刻画量子机器算法的优势所在。目前仍然不能说明某个量子算法的性能比所有经典算法都好, 因为没有找到同样复杂度的经典算法, 并不代表这样的经典算法不存在。此外, 对一批代表性问题, 经典机器学习理论中有一些对应的启发式方法能够巧妙而快速地予以解决, 而量子机器学习算法是否仍能匹敌, 有待进一步研究。

#### 参考文献(References):

- [1] GOODFELLOW I, BENGIO Y, COURVILLE A. *Deep Learning* [M]. Massachusetts: MIT Press, 2016.
- [2] BRIF C, CHAKRABARTI R, RABITZ H. Control of quantum phenomena: past, present and future [J]. *New Journal of Physics*, 2010, 12(7): 075008.
- [3] BIAMONTE J, WITTEK P, PANCOTTI N, et al. *Quantum machine learning* [EB/OL]. <https://arxiv.org/abs/1611.09347>, 2016.
- [4] GRABADE C, FERRIE C, WIEBE N, et al. Robust online hamiltonian learning [J]. *New Journal of Physics*, 2012, 14(10): 103013 – 103043.
- [5] WIEBE N, GRANADE C, FERRIE C, et al. Hamiltonian learning and certification using quantum resources [J]. *Physical Review Letters*, 2014, 112(19): 190501.
- [6] SENTIS G, CALSAMIGLIA J, MUNOZTAPIA R, et al. Quantum learning without quantum memory [J]. *Scientific Reports*, 2012, 2: 708.

- [7] SENTIS G, GUTA M, ADESSO G. Quantum learning of coherent states [J]. *Epl Quantum Technology*, 2015, 2(1): 1 – 22.
- [8] BISIO A, CHIRIBELL G, D'ARIANO G M, et al. Optimal quantum learning of a unitary transformation [J]. *Physical Review A*, 2010, 81(3): 76 – 76.
- [9] DAN V, MARTINEZ T. Quantum associative memory [J]. *Information Sciences*, 2000, 124(1): 273 – 296.
- [10] GIOVANNETTI V, LLOYD S, MACCONE L. Quantum random access memory [J]. *Physical Review Letters*, 2008, 100(16): 160501.
- [11] NIELSON M A, CHUANG I L. *Quantum Computation and Quantum Information* [M]. Cambridge: Cambridge University Press, 2000: 558 – 559.
- [12] RUIZ-PEREZ L, GARCIA-ESCAVIN J C. Quantum arithmetic with the quantum fourier transform [J]. *Quantum Information Processing*, 2017, 16(6): 152.
- [13] GROVER L K. A fast quantum mechanical algorithm for database search [C] // *The 28th ACM Symposium on Theory of Computing*. ACM, 1996: 212 – 219.
- [14] BENNETT C H, BERNSTEIN E, BRASSARD G, et al. Strengths and weaknesses of quantum computing [J]. *Siam Journal on Computing*, 1997, 26(5): 1510 – 1523.
- [15] ANGUIA D, RIDELLA S, RIVIECCIOI F, et al. Quantum optimization for training support vector machines [J]. *Neural Networks*, 2003, 16(5/6): 763 – 770.
- [16] LLOYD S, MOHSENI M, REBENTROST P. *Quantum Algorithms for Supervised and Unsupervised Machine Learning* [EB/OL]. <https://arxiv.org/abs/1307.0411>, 2013.
- [17] REBENTROST P, MOHSENI M, LLOYD S. Quantum support vector machine for big data classification [J]. *Physical Review Letters*, 2014, 113(13): 130503.
- [18] HARROW A W, HASSIDIM A, LLOYD S. Quantum algorithm for linear systems of equations [J]. *Physical Review Letters*, 2009, 103(15): 150502.
- [19] CLADER B D, JACOBS B C, SPROUSE C R. Preconditioned quantum linear system algorithm [J]. *Physical Review Letters*, 2013, 110(25): 250504.
- [20] WOSSNIG L, ZHAO Z, PRAKASH A. *A quantum linear system algorithm for dense matrices* [EB/OL]. <https://arxiv.org/abs/1704.06174>, 2017.
- [21] CAO Y, DASKIN A, FRANKEL S, et al. Quantum circuit design for solving linear systems of equations [J]. *Molecular Physics*, 2012, 110(15/16): 1675 – 1680.
- [22] PAN J, CAO Y, YAO X, et al. Experimental realization of quantum algorithm for solving linear systems of equations [J]. *Physical Review A*, 2013, 89(2): 1150 – 1154.
- [23] LLOYD S, MOHSENI M, REBENTROST P. Quantum principal component analysis [J]. *Nature Physics*, 2014, 10(9): 631 – 633.
- [24] JORDAN S P. Fast quantum algorithm for numerical gradient estimation [J]. *Physical Review Letters*, 2005, 95(5): 050501.
- [25] REBENTROST P, SCHULD M, PETRUCCIONE F, et al. *Quantum gradient descent and Newton's method for constrained polynomial optimization* [EB/OL]. <https://arxiv.org/abs/1612.01789>, 2016.
- [26] CHATTERJEE R, YU T. *Generalized coherent states, reproducing kernels, and quantum support vector machines* [EB/OL]. <https://arxiv.org/abs/1612.03713>, 2016.
- [27] ZHAO Z, FITZSIMONS J K, FITZSIMONS J F. *Quantum assisted Gaussian process regression* [EB/OL]. <https://arxiv.org/abs/1512.03929>, 2015.
- [28] LI Z, LIU X, XU N, et al. Experimental realization of a quantum support vector machine [J]. *Physical Review Letters*, 2015, 114(14): 140504.
- [29] WIEBE N, KAPOOR A, SVORE K. Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning [J]. *Quantum Information & Computation*, 2014, 15(3): 318 – 358.
- [30] WIEBE N, KAPOOR A, SVORE K M. Quantum perceptron models [EB/OL]. *arXiv*, 2016: 1602.04799.
- [31] LEWENSTEIN M. Quantum perceptrons [J]. *Journal of Modern Optics*, 1994, 41(12): 2491 – 2501.
- [32] RICKS B, DAN V. Training a quantum neural network [J]. *Advances in Neural Information Processing Systems*, 2004, 14(6): 621 – 622.
- [33] KULCHYTSKYI B, ANDRIYASH E, AMIN M, et al. Quantum Boltzmann Machine [C] // *APS Meeting Abstracts*. Maryland, 2016.

#### 作者简介:

**陆思聪** (1989–), 男, 博士研究生, 2008–2012年就读于清华大学自动化系, 获工学学士学位, 2013至今在清华大学自动化系攻读工学博士学位, 目前研究方向为机器学习、自然语言处理, E-mail: [lusicong@gmail.com](mailto:lusicong@gmail.com);

**郑昱** (1992–), 女, 硕士研究生, 2015年毕业于东北大学自动化专业, 现硕士就读于清华大学微纳电子系, 研究方向为量子算法与机器学习, E-mail: [zhengyu15@mails.tsinghua.edu.cn](mailto:zhengyu15@mails.tsinghua.edu.cn);

**王晓霆** (1982–), 男, 博士, 教授, 2001–2005年就读于武汉大学数学系, 获理学学士学位, 2005–2011年就读于剑桥大学应用数学与理论物理系, 依次获得剑桥大学硕士和博士学位, 2011–2017年分别在麻省大学波士顿分校, 麻省理工大学, 路易斯安那州立大学从事博士后研究工作, 2017年至今担任电子科技大学教授, 研究方向包括量子信息学、量子优化、量子控制、量子机器学习, E-mail: [xiaoting@uestc.edu.cn](mailto:xiaoting@uestc.edu.cn);

**吴热冰** (1976–), 男, 博士, 1993–1998年本科就读于清华大学电机系, 1998–2004年就读于清华大学自动化系获博士学位, 2005–2008年在普林斯顿大学化学系从事博士后研究工作, 2008年至今在清华大学自动化系工作, 现为副教授, 研究方向包括量子控制、非线性控制、量子机器学习等, E-mail: [rbwu@tsinghua.edu.cn](mailto:rbwu@tsinghua.edu.cn).