

量子机器学习算法综述

黄一鸣 雷 航 李晓瑜

(电子科技大学信息与软件工程学院 成都 610054)

摘 要 机器学习在过去十几年里不断发展,并对其他领域产生了深远的影响.近几年,研究人员发现结合量子计算特性的新型机器学习算法可实现对传统算法的加速,该类成果引起了广泛的关注和研究.因此,文中对近十年的量子机器学习算法进行总结、梳理.首先,介绍了量子计算和机器学习的基本概念;其次,从四个方面分别介绍了量子机器学习,分别是量子无监督聚类算法、量子有监督分类算法、量子降维算法、量子深度学习;同时,对比分析量子机器学习算法与传统机器学习算法的区别和联系;最后,总结该领域存在的问题及挑战,并对量子机器学习未来的工作进行展望.

关键词 量子机器学习;量子计算;大数据;人工智能;量子深度学习

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2018.00145

A Survey on Quantum Machine Learning

HUANG Yi-Ming LEI Hang LI Xiao-Yu

(School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054)

Abstract With tremendous development of classical machine learning in past decades, it has influenced on many different fields as an essential part of data science. Recent years, as an innovative type of machine learning algorithms, quantum machine learning plays high parallel performance based on quantum mechanics. Consequently, it speeds up some of conventional algorithms. Researchers in quantum machine learning focus on improving performance of classical machine learning through quantum computation and exploring the possibility of combination of machine learning with quantum mechanics, and presenting some new algorithms. The framework of quantum machine learning mainly contains three steps: (1) Load and input classical data. It is converting classical information into quantum information. Making full use of high performance of quantum computation, it is necessary to encode classical data to quantum data according to quantum representation. (2) Construct series of unitary operators and use them to process quantum data. The basic principle of quantum computation is that all the operations must follow the unitary characteristics. Thereby, it is not possible to revise all classical algorithms to quantum environment, and not all the quantum algorithms can provide exponential speed-up. (3) Read and output the learning result. After the first two steps, the quantum machine learning results are quantum states. In order to read and output the useful learning result, it needs some necessary measurements to extract relevant information. There are three kinds of research in this field: (1) Use quantum computing accelerates the costly part of traditional algorithm. This work is the hybrid of classical algorithm and quantum parts. There is no change of the main idea of learning algorithm. But complex parts are running on quantum devices, and it makes full use of

收稿日期:2016-05-14;在线出版日期:2017-05-19. 本课题得到国家自然科学基金(61502082)、中央高校基本科研业务费基础研究项目(ZYGX2014J065)资助. 黄一鸣,男,1991年生,博士研究生,主要研究方向为数据分析、数据挖掘和量子机器学习. E-mail: yiminghwang@gmail.com. 雷 航,男,1960年生,博士,教授,中国计算机学会(CCF)会员,主要研究领域为大数据挖掘和嵌入式系统. 李晓瑜(通信作者),女,1984年生,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为数据分析、量子计算和量子机器学习.

quantum computing to speed up counterparts. (2) Borrow the physics concept to propose new algorithms. It is based on the mathematical framework of quantum mechanics. And researchers develop more efficient methods to solve optimization problem. (3) Use machine learning methods to solve some physics issues, such as quantum tomography by using compress sensing. With the power of machine learning, researchers can explore quantum world from another perspective, vice versa. Quantum machine learning solves classical data mining and data analysis problems from another perspective. There are some difficulties in building general quantum computer, researchers still make solid progress. With the development of quantum techniques, practical quantum machine learning algorithms will increasingly be presented. Therefore, some typical algorithms of quantum machine learning are surveyed in this paper. Firstly, it introduces the fundamental concepts of quantum computation and classical machine learning, including quantum states, quantum gate, and measurement. Secondly, it summarizes the quantum machine learning algorithms from four aspects including quantum unsupervised clustering algorithm, quantum supervised classification algorithm, quantum dimension reduction algorithm, quantum deep learning. It collects some typical algorithms for every part. Meanwhile, it analyzes and makes a comparison of computational complexity between classical machine learning and quantum machine learning. Finally, it reconfirms the problems, challenges and future research works for quantum machine learning along with an insight of quantum computation for big data.

Keywords quantum machine learning; quantum computation; big data; artificial intelligence; quantum deep learning

1 引 言

近十几年时间里,机器学习快速崛起,已经成为大数据时代下的技术基石.机器学习根据已有数据进行学习策略的探索和潜在结构的发现,依据所得模型进行预测及分析^[1].机器学习源于人工智能和统计学,其应用极其广泛.从数据挖掘到人脸识别,从自然语言处理到生物特征识别,从垃圾邮件分类到医学诊断,社会生活的各个方面都被机器学习技术所影响.随着信息技术不断发展,信息化将各行业紧密联系起来,产业数据成爆炸式增长.这种增长不仅是数据量的增长,还包括数据种类、结构和产生速度上的增长.最近几年全球数据量的增长率接近24%^[2].以Google为首,凭借数据服务为核心、机器学习技术为支撑的一大批IT公司占领数据挖掘与信息化的市场.他们掌握海量数据,使用机器学习技术挖掘潜在价值信息,提供数据服务,改变社会生活各个方面.数据的增长不仅带来丰厚的利润,同时也带来技术的挑战.不少传统机器学习算法已无法应对大数据时代海量数据的处理和分析,所以不得不寻找新的方法来解决.

最近不少研究机构及大型IT公司都将目光集

中到了量子计算上,想通过量子计算的独特性质,解决传统算法的运算效率问题.传统电子计算机存储电平的高低,每次只能处理一个比特的状态数据.量子计算机存储量子比特,一个量子比特可表示量子态 $|0\rangle$ 和 $|1\rangle$ 的叠加,一次运算就可同时处理两个状态的信息.以此类推,经典计算机对 2^n 比特的数据执行相同计算需要 2^n 次操作,而量子计算机只需要对 n 个量子比特进行一次操作即可.正因如此,量子计算不管在数据存储能力还是数据处理能力上都理论上远超经典计算.

早在1982年,Feynman指出基于量子力学建造的计算机对特定问题的求解是传统计算机无法比拟的^[3].1994年,Shor提出了一种里程碑式的量子因子分解算法^[4],称为Shor算法.计算步骤上,传统大数因子分解的最佳算法的时间复杂度随问题规模成指数倍增加,而Shor算法可在多项式时间内完成.1997年,Grover提出一种量子搜索算法^[5],该算法相比传统无序数据库搜索算法有着平方级效率的提升.现有的量子算法^①,大多相较于对应的经典算法有着明显提速效果.由此,研究学者猜想既然量子计算对特定经典问题的求解有显著提速,是否可将

① <http://math.nist.gov/quantum/zoo/>

其应用到机器学习领域,解决目前处理大数据时计算效率低的问题。

近年来出现越来越多结合量子计算和机器学习的研究^[6-8]。研究人员一方面希望通过量子计算解决机器学习的运算效率问题;另一方面探索使用量子力学的性质,开发更加智能的机器学习算法。虽然量子计算机的研究不断深入,但依然有不少问题有待解决^[9],例如如何保持量子的相干性、如何降低苛刻的环境条件等。同时,量子计算机的建造以及量子算法的提出都涉及诸多领域要素,量子机器学习理论现在依旧处于一个起步阶段,至今没有形成如经典机器学习领域完备的理论体系,大多研究还处于探索实验阶段。

量子机器学习领域的研究最早可追溯到1995年对量子神经网络的研究:Kak最先提出量子神经计算的概念^[10]。随后研究人员提出了各类量子神经网络模型,如Behrman等人提出的基于量子点神经网络模型^[11],Toth等人提出了量子细胞神经网络^[12],Ventura等人提出的使用量子叠加态表示网络^[13],Matsui等人提出了通过量子门电路实现的神经网络^[14],Zhou等人提出了量子感知机^[15],Schuld等人提出了由量子随机行走构建神经网络^[16],等等。这些不同种类的量子神经网络模型的构建,均是基于量子力学特性与不同数据结构的结合。该领域研究的一个难点在于如何将神经网络的非线性映射结构同量子计算的线性变换结合在一起,即如何实现神经网络的量子演化。

研究人员发现量子特性有助于研究无监督聚类问题,故提出了量子无监督聚类算法。2001年,Horn等人最早将量子力学特性引入传统聚类算法,将其薛定谔方程与Parzen窗估算量的极大值求解联系起来,用于发现数据的聚类中心^[17]。2013年,Lloyd等人将量子态的叠加性应用到经典向量表示上,提出量子K-means算法,该算法理论上能够实现海量数据的高效聚类^[18]。同年,Aïmeur等人提出了量子分裂聚类算法,其借助Grover变体算法进行子过程中最大距离的快速搜索^[19-20]。类似还有研究人员结合监督分类算法和量子计算,提出量子有监督分类算法,例如2014年,微软的Wiebe等人提出了用量子态的概率幅表示经典向量,并通过比较两个量子态间距离完成量子最近邻算法^[21]。同年,Rebentrost等人首次提出使用量子系统的密度矩阵进行支持向量机中核矩阵的表示^[22]。2016年,微软的Wiebe等人提出了量子深度学习的概念,首次将量子计算同

深度学习相结合,其通过量子采样实现受限玻尔兹曼机的梯度估计,旨在加速深度网络的训练^[23]。上述量子机器学习算法的核心大多还是与传统算法相同,主要区别在于通过量子计算的高并行性去处理计算耗时的子步骤。

其他关于量子计算及机器学习核心问题的研究也进一步推动量子机器学习的发展。首先,Giovannetti等人于2008年提出了量子随机存取存储器(Quantum Random Access Memory, QRAM)^[24-25],随后许多量子机器学习算法相继产生,如2014年Lloyd等人基于QRAM提出量子主成分分析(Quantum Principle Component Analysis, QPCA)算法等^[26]。其次,Harrow等人于2009年,提出用量子算法解决线性系统的方程问题^[27],常被研究人员成为HHL算法;2015年,Childs等人也对该问题进行了相关研究^[28],进一步拓展了量子算法对线性系统问题的解决能力。很多传统机器学习问题最终与最优化问题的求解相关,而最优化问题常涉及线性方程组的求解,所以通过该技术可有助经典机器学习中最优化步骤的提速。例如,Rebentrost等人于2014年提出的量子支持向量机就用到了量子线性方程求解算法^[22]。很多算法是以QRAM物理实现为前提,利用QRAM实现任意量子态的制备,继而进行后续量子态计算。

对于量子机器学习的可学习性及其与经典算法的比较,也是研究人员的关注点之一。2004年,Servedio等人对传统机器学习算法的可学习性与量子算法的可学习性进行了分析与比较^[29]。随后,2006年,Aïmeur等人提出了在量子环境下完成机器学习任务的猜想^[30]。Yoo等人从二分类问题上对量子机器学习与传统机器学习进行了比较,指出量子的叠加性原理使得量子机器学习算法运算效率明显优于传统算法,并从学习的接受域上进行了比较,发现量子机器学习的接受域较大,从而决定了学习效率优于传统算法^[31]。随着大数据时代的到来,传统算法对于海量数据的处理能力,也日益捉襟见肘。这就进一步促使研究人员考虑量子机器学习对大数据问题的解决能力和可行性。首先,2015年初陆朝阳教授研究小组的相关研究开始关注量子机器学习与大数据的结合^[32]。随后,年底王书浩和龙桂鲁教授对量子机器学习及大数据领域的应用做了综述性介绍^[33]。这些都进一步推动了量子计算在数据挖掘和数据分析方面的研究和应用。

以D-Wave及Google为首的公司对该领域也

进行了研究. 2008 年, Google 的 Neven 及 D-Wave 公司的 Rose 等人, 在其研发的超导绝热量子处理器上使用量子绝热算法解决图像识别问题^[34-35], 此后他们又做了一系列将量子绝热算法应用到人工智能领域的研究^[36-39]. 这一系列量子绝热算法没有通过量子门电路进行量子计算, 而是运行在 D-Wave 研发的特定量子芯片上, 并且其运行的环境条件也相对苛刻. 目前他们研究的算法还有很多限制, 其商业领域的实际应用还有一段距离, 不过已经向量子机器学习的产业化应用迈出了坚实的一步.

在量子机器学习的物理实验验证方面, 全球的研究人员也在努力不懈. 2015 年, 潘建伟教授团队首次在小量子计算机上实现并验证了 Lloyd 提出的 K-means 算法^[40]. Li 等人也于同年实验验证了量子支持向量机算法, 并进行了手写数字的二分类实验验证^[32]. 虽然实验的数据规模受限于当前量子计算机的量子比特数, 但足以证明量子机器学习算法的可行性, 且鼓舞我们在该领域进行更加深入的研究, 以获得解决海量数据分析处理的有用工具.

相信随着制造工艺的不断提升, 量子计算及量子机器学习理论不断完善, 更多巧妙结合物理原理和机器学习理论的新型算法将被提出. 量子机器学习将极大促进现有机器学习的发展, 产生出更加高效、强大的学习算法.

2 量子计算与机器学习

2.1 量子计算

所有信息计算的本质都是一个物理过程^[41]. 计算设备的物理特性决定其计算能力. 根据 Moore 定律^[42], 集成电路上可容纳的晶体管数目约每隔 24 个月增加一倍. 按此速度发展, 一方面芯片元件集成度的提高会导致单位体积内散热增加, 由于材料散热速度有限, 就会出现计算速度的上限, 产生“热耗效应”; 另一方面元件尺寸的不断缩小, 在纳米甚至埃尺度下经典世界的物理规则不再适用, 出现“尺寸效应”^[43]. 综上, 随着集成技术的提升, 为解决“热耗效应”和“尺寸效应”不得不重新考虑能适用于微观世界的新型计算机模型. 由此量子计算应运而生. 它最早的想法源于 Benioff 提出的基于量子力学原理的可逆计算模型^[44], 后来 Feynman 进行了扩展.

2.1.1 量子态

不同于经典电子计算机, 量子计算机操作的基本数据单元是量子比特 (qubit). 物理上, 量子比特可以有不同方法的实现, 可由两能级原子系统来表

示, 也可由光的不同偏振方向来表示. 一个量子态通常用狄拉克符号 $|\cdot\rangle$ 来表示. 数学上, 它为一个 n 维希尔伯特空间中的复向量, 如式 (1) 所示.

$$|\varphi\rangle = \sum_{i=1}^n \alpha_i |i\rangle \quad (1)$$

式 (1) 中 $|i\rangle$ 为基态, α_i 为复数, 是每个态的概率幅. 按照量子力学原理, 如果测量该量子态 $|\varphi\rangle$, 最后会以 $|\alpha_i|^2$ 的概率波包塌缩到基态 $|i\rangle$ 上, 因此每个基态的概率幅 α_i 必须满足式 (2).

$$\sum_{i=1}^n |\alpha_i|^2 = 1 \quad (2)$$

多个量子比特组成的集合常称为量子寄存器, 除了使用狄拉克符号也可以使用向量的形式表示. 例如可使用 n 维向量表示一个 n 位量子寄存器

$$|\varphi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle, \text{ 如式 (3) 所示.}$$

$$|\varphi\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{2^n-1} \end{pmatrix} \quad (3)$$

其中每个基态 $|i\rangle$ 都可以展成二进制形式, 如 $|2^n-1\rangle = \underbrace{|111, \dots, 1\rangle}_n$. $|111, \dots, 1\rangle$ 中 1 对应一位量子比特.

2.1.2 量子门

与经典计算机相同, 量子计算机对量子比特的任何操作都可通过量子门实现. 不同于经典计算机进行门电路操作后产生的能量耗散, 例如 0 和 1 的与运算这是个不可逆且能量耗散的过程, 量子计算机可通过幺正变换实现可逆计算, 从而解决能耗上的问题. 所以全部量子门都对应一个幺正算子. 任何 n 位量子比特门都可用 $n \times n$ 的矩阵 U 来表示, 如式 (4).

$$U = (u_{ij})_{i,j=1}^{2^n} = \begin{bmatrix} u_{11} & \cdots & u_{12^n} \\ \vdots & \ddots & \vdots \\ u_{2^n 1} & \cdots & u_{2^n 2^n} \end{bmatrix} \quad (4)$$

因为 U 实质为一个幺正变换, 则必须满足式 (5) 条件.

$$UU^\dagger = U^\dagger U = I \quad (5)$$

其中 U^\dagger 为 U 的共轭转置. 对量子态 $|\varphi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle$ 进行 U 变换, 可得式 (6).

$$U|\varphi\rangle = \sum_{j=0}^{2^n-1} u_{ij} \alpha_j |j\rangle = \sum_{i=0}^{2^n-1} \beta_i |i\rangle$$

$$= \begin{bmatrix} u_{11} & \cdots & u_{12^n} \\ \vdots & \ddots & \vdots \\ u_{2^n-1} & \cdots & u_{2^n-1} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{2^n-1} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \quad (6)$$

2.1.3 测量

量子计算的结果依旧处于一个叠加态,为了得到该结果,必须对该量子态进行测量,使叠加态波包塌缩到一个基态.定义一组测量算子 $\{M_m\}$,该组测量算子还需满足完备性.

$$\sum M_m^\dagger M_m = \mathbf{I} \quad (7)$$

m 对应测量的可能结果,若使用该测量算子对量子态 $|\varphi\rangle$ 进行测量,最终得到 m 的概率为

$$p(m) = \langle \varphi | M_m^\dagger M_m | \varphi \rangle \quad (8)$$

测量后的量子态为

$$\frac{M_m |\varphi\rangle}{\sqrt{\langle \varphi | M_m^\dagger M_m | \varphi \rangle}} \quad (9)$$

2.2 机器学习与深度学习

机器学习属于人工智能领域,主要利用计算机分析处理已知数据,从数据中自动“学习”潜在规则,并将此规则用于对未知数据的预测.例如对手写字符的识别,需要将字符图像信息转换成特征数据集 X , $X = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)})$.数据集总共有 m 个样本, $\mathbf{x}^{(i)}$ 为第 i 个样本. $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_j^{(i)}, \dots, x_n^{(i)})$,每个样本有 n 个特征, $x_j^{(i)}$ 为第 j 个特征.然后机器使用学习算法对手写字符集 X 进行潜在规则的提取,最后使用该规则进行字符的判断.传统机器学习主要分为监督学习、无监督学习、半监督学习及增强学习.监督学习的训练样本都有人为标注分类结果的标签.输入训练样本,通过最小化代价函数,找出样本特征与标签之间的映射关系(学习模型).新的未标注数据通过训练得到的学习模型进行分类.无监督学习,没有标注标签的训练集,学习算法自行寻找数据集当中的潜在规则,该类以聚类算法最为常见.半监督学习是介于监督与无监督之间的一种学习算法,为解决大数据量及监督学习标记代价高的问题,其主要使用较少的带标签及较多的未带标签数据进行训练分类.增强学习源于心理学的奖励与惩罚,通过刺激进而逐步优化,通过不断的反馈达到最优.

机器学习相较于以往基于人工规则的学习方法,显示出强大处理分析能力.各类学习算法不断被提出,例如通用的逻辑回归^[45](Logistic Regression, LR)、支持向量机^[46](Support Vector Machine, SVM)、关联规则挖掘^[47](Associated Rules Mining, ARM)

算法等;随着研究的逐步深入,研究人员发现以上学习算法具有一定的局限性.相对于深度学习,上述算法的输入特征需要使用专业领域知识进行筛选提取,同时该学习算法均使用单层非线性变换进行数据分析处理.这种单层非线性变换结构,制约了其对现实中更加复杂高维的函数的表示能力.为了达到更好描述复杂世界,解决实际问题的目的,学习算法需要拥有大量的训练数据作为基础,使用专业的领域知识来提取优质的特征,通过强大的计算能力求解问题.通常训练样本无法保证充足,同时特征的提取也直接影响了学习算法的表现能力.因此,深度学习应运而生.

深度学习的概念最早由 Hinton 等人在 2006 年提出,文章指出多层的神经网络能更好地表达数据本质,并且可以通过逐层初始化的方法解决多层复杂网络训练困难的问题^[48].之后,不管是学术还是工业界,深度学习都引起了广泛的关注.大量的实验证明深度学习相对于浅层学习,精度上有明显的提高.不管从语音识别还是图像识别领域,深度学习都能更好地刻画和承载海量数据中更丰富的信息.典型的深度学习结构分为以下 3 种^[49]. (1) 生成型,如深度信念网络^[50].该类模型能表达输出与输入样本数据间的联合概率分布,解决网络初始权重设置问题,提升模型性能及收敛速度^[51]; (2) 区分型,如卷积神经网络^[52].该类模型使用权值共享,同一层神经元共享一个权值集合,所以网络中的自由参数少于传统网络^[53]; (3) 混合型,该模型通过受限玻尔兹曼机预训练产生,再使用反向传播算法(Backpropagation Algorithm, BP)优化深度信念网络的权值,这样就比单用 BP 算法对卷积神经网络进行训练更优.

3 量子机器学习

量子机器学习借助量子计算的高并行性,实现进一步优化传统机器学习的目的. Servedio, Aaronson 以及 Cheng 等人对量子态的可学习性进行了研究^[29,54-55]. Servedio 等人从信息论的角度指出量子信息和经典信息的可学习性是等价的,这也促使了量子机器学习研究的发展^[56-58]. Cheng 在 Servedio 的基础上提出量子学习算法主要分为两大类研究,一类是量子计算学习(Quantum Computational Learning, QCL),另一类是量子概率学习(Quantum Statistical Learning, QSL)^[29]. QCL 主要研究两点. (1) 如何利用量子力学原理改进传统机器学习过

程,提高计算效率,降低计算复杂度^[18,23,58-65]; (2) 利用统计学习理论去解决量子状态层析(Quantum State Tomography, QST)、寻找量子系统的隐藏结构等问题^[66-69]. 至今绝大部分前人工作都集中于解决第一种计算复杂性问题,故本文主要介绍该类量子机器学习算法。

并非所有经典学习算法的步骤都可通过量子特性进行加速. 量子力学公设指出封闭量子系统通过酉变换来刻画量子态的演化^[70],即量子态演化需要由酉算子来实现. 近年来,龙桂鲁教授研究小组提出使用酉变换的线性组合,也可实现量子计算^[71]. 量子机器学习算法的基本原理是需将经典学习算法中的信息映射成量子态,然后对量子态进行酉演化操作,进而达到计算的目的. 因此,只有满足以上计算条件的经典学习步骤才能实现加速. 为使用量子特性达到加速的目的,就必须对传统算法进行相应的改写,使其满足上述量子计算的基本要求. 当前的研究大多集中于用量子算法替代原有经典算法的特定子过程,进而达到降低计算复杂度,提高算法效率的目的。

量子机器学习算法一般需要经过以下步骤:

步骤 1. 将经典信息转换成量子信息. 为了发挥量子计算机的高并行特性,必须对经典信息进行编码,将其转换成量子信息,这就好比将一门语言翻译为另外一门语言. 合适、巧妙的编码将更加有效的利用量子计算的潜力。

步骤 2. 传统机器学习算法的量子版转换. 由于量子计算机和经典电子计算机的操作单元不同,无法将所有的经典计算机的方法都移植到量子计算机上,并且不是所有在量子计算机上的操作都会有指数性的加速. 所以设计出适用于量子计算机的算法将十分重要. 量子机器学习算法的设计,既要结合经典算法的数据结构、数据库等技术,同时,还要不断设计出更多适合量子理论的算法模型. 这个建模的过程,也是步骤 2 的重点和难点。

步骤 3. 提取最终计算结果. 由于计算结果为量子态无法直接使用,需要经过量子测量操作,使量子叠加态波包塌缩至经典态,将经典信息提取出来。

已有的量子机器学习主要可以分为以下 3 类,分别是:

(1) 第一类量子机器学习. 该类算法将机器学习中复杂度较高的部分替换为量子版本进行计算,从而提高其整体运算效率,如图 1 所示. 该类量子机器学习算法整体框架沿用原有机器学习的框架. 其主体思想不变,不同点在于将复杂计算转换成量子

版本运行在量子计算机上,从而得到提速. 该类研究的代表性成果有:Lloyd 教授提出的 QPCA、QSVM (Quantum Support Vector Machine)^[18,22]等。

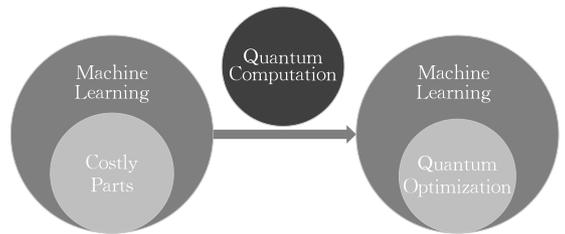


图 1 第一类量子机器学习

(2) 第二类量子机器学习. 该类算法的特点是寻找量子系统的力学效应、动力学特性与传统机器学习处理步骤的相似点,将物理过程应用于传统机器学习问题的求解,产生出新的机器学习算法,如图 2 所示. 该类算法与第一类不同,其全部过程均可在经典计算机上进行实现. 在其他领域也有不少类似思路的研究,如退火算法、蚁群算法等. 该类量子机器学习算法的代表性研究有 Horn 等人的基于量子力学的聚类算法^[17].

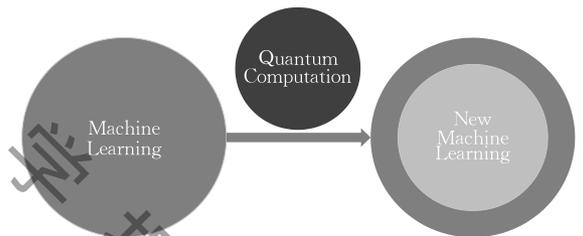


图 2 第二类量子机器学习

(3) 第三类量子机器学习. 该类算法主要借助传统机器学习强大的数据分析能力,帮助物理学家更好的研究量子系统,更加有效的分析量子效应,作为物理学家对量子世界研究的有效辅助,如图 3 所示. 该类算法的提出将促进我们对微观世界进一步的了解,并解释量子世界的奇特现象. 该类算法的代表性研究有:Gross 的基于压缩感知的量子断层分析^[68-69]等。

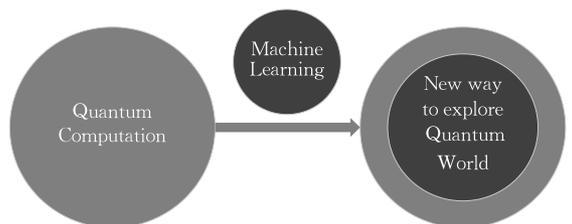


图 3 第三类量子机器学习

由于量子机器学习的大多数研究集中于第一类算法,第二类算法的研究还较少,第三类量子机器学

习算法主要应用于物理领域. 所以, 本部分后续小节将着重介绍第一类量子机器学习算法. 本文分别从无监督量子聚类算法、有监督量子分类算法、量子降维算法、量子深度学习等几个方面进行, 最后会对时下主流的量子机器学习算法进行比较和分析.

3.1 无监督量子聚类算法

3.1.1 量子 K -means 算法

聚类算法是无监督学习算法中较为重要的一类. 以最普遍使用的 k -means 算法为例, 该算法以距离的远近作为样本相似性指标, 即两个样本的距离越近相似程度就越高^[72]. 当遇到海量样本数据时, k -means 算法存在时间开销较大的缺点.

Lloyd 等人于 2013 年提出了量子版最近中心算法^[18]. 量子最近中心算法可视为经典 k -means 聚类算法的子过程, 使用该方法可对经典算法进行指数级提速^[18]. 该算法的加速来源于量子计算机具有高维空间中对向量和张量操作的优势, 这是由量子力学自身特性所决定的. 量子力学使用希尔伯特空间中的向量来描述物理系统, 而希尔伯特空间本身就是完备的线性空间, 对量子态的操作即在线性空间中对向量的操作. 同时, 希尔伯特空间中的量子态满足叠加性原理, 对多个态可实行并行操作, 故计算效率远超经典计算. 量子计算满足以上量子力学规律及数学特性, 故量子机器学习能有效处理高维数据. 量子最近中心算法的核心思想是对实向量间的距离作比较, 通过寻找向量 \mathbf{u} 与集合 $\{\mathbf{v}\}$ 中的哪个量的距离最近, 从而不断分类, 进而不断自动获得聚类结果, 即 $\arg \min_c \left| \mathbf{u} - m^{-1} \cdot \sum_{j=1}^m \mathbf{v}_j^c \right|$. Lloyd 通过式(10)将实向量 $\mathbf{u} = (u_0, u_1, \dots, u_n)$ 转换成量子态 $|u\rangle$, 从而将向量的距离比较转换成量子态间的距离比较.

$$|u\rangle = \frac{\mathbf{u}}{|\mathbf{u}|} \quad (10)$$

同时定义纠缠态 $|\varphi\rangle$:

$$|\varphi\rangle = \frac{1}{\sqrt{2}} \left(|u\rangle |0\rangle + \frac{1}{\sqrt{m}} \sum_{j=1}^m |\mathbf{v}_j^c\rangle |j\rangle \right) \quad (11)$$

以及态 $|\phi\rangle$:

$$|\phi\rangle = \frac{1}{\sqrt{2}} \left(|\mathbf{u}| |0\rangle - \frac{1}{\sqrt{m}} \sum_{j=1}^m |\mathbf{v}_j^c\rangle |j\rangle \right) \quad (12)$$

其中, $|\mathbf{u}|$ 为实向量 \mathbf{u} 的范数, $|\mathbf{v}_j^c\rangle$ 为 c 簇的第 j 个向量, m 为样本总数, $Z = |\mathbf{u}|^2 + (n^{-1} \sum_{j=1}^n |\mathbf{v}_j^c|^2)$ 为归一化系数. 这两个量子态的距离可通过式(13)得到.

$$D_i = \sqrt{2|\langle \phi | \varphi \rangle|^2 Z} \quad (13)$$

$|\langle \phi | \varphi \rangle|^2$ 通过 controlled-SWAP 电路实现^[73], 如图 4 所示.

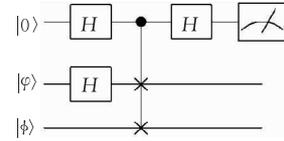


图 4 controlled-SWAP 量子电路实现

图 4 中 H 代表 Hadamard 门, 电路使能与否由第一个量子比特控制, 即电路最上面的量子线对应的是控制比特, 下面的两个量子线对应的是受控量子比特. controlled-SWAP 电路^[73] 作用于输入量子态 $|\varphi\rangle$ 和 $|\phi\rangle$ 后, 得到式(14).

$$\frac{1}{2} |0\rangle \langle \phi | \varphi \rangle + |\varphi\rangle \langle \phi \rangle + \frac{1}{2} |1\rangle \langle \phi | \varphi \rangle - |\varphi\rangle \langle \phi \rangle \quad (14)$$

对第一位量子比特进行测量, 得 $|0\rangle$ 的概率为

$$P(|0\rangle) = \frac{1}{2} + \frac{1}{2} |\langle \phi | \varphi \rangle|^2,$$

那么 $|\langle \phi | \varphi \rangle|^2 = 2P(|0\rangle) - 1$.

k -means 算法中初始点的选择十分重要, 好的初始点将极大地缩短迭代次数. 初始点选择的是否足够好, 评判的一个标准是, 需要尽可能将初始点稀疏分布于整个样本空间中, 即各点之间距离尽可能大. 为了优化初始点, Lloyd 使用量子绝热演化算法来实现 k -means 算法最优初始点的选择^[74]. 其基本思想为, 首先初始化 Hamiltonian 量, 如式(15)所示.

$$H_0 = \mathbf{I} - |\psi\rangle \langle \psi| \quad (15)$$

令 $|\psi\rangle = \underbrace{|\varphi\rangle \otimes \dots \otimes |\varphi\rangle}_k$, k 为初始点个数, $|\varphi\rangle =$

$(1/\sqrt{m}) \sum_{j=1}^m |j\rangle$. $|\varphi\rangle$ 为样本标签编码后的归一化量子态, m 为样本总数. 然后定义末态 Hamiltonian 如式(16)所示. 末态 Hamiltonian 的基态即是满足稀疏分布的初始点, 最后进行测量操作, 提取基态信息即可.

$$H_s = \sum_{j_1 \dots j_k} \sum_{i, i'=1}^k -|\vec{v}_{j_i} - \vec{v}_{j_{i'}}|^2 |j_1\rangle \langle j_1| \otimes \dots \otimes |j_k\rangle \langle j_k| \quad (16)$$

在量子机器学习算法的物理实验方面, 我国的研究团队处于世界领先水平. 2015 年, 潘建伟院士团队以小型光量子计算机为实验平台, 首次对 Lloyd 教授于 2013 年提出的量子 K -means 算法进行了物理实验验证^[40]. 该实验将单光子作为一位量子比特, 使用自发参量下转换技术制备纠缠的光子对, 并使用干涉仪将参考向量即测试向量编码到量

子态上. 实验分别对 4、6、8 量子比特规模的算法进行了验证, 结果指出传统机器学习中普遍存在的高维向量间距离和内积的计算可在量子计算机上实现, 这表明量子计算机上实现的该部分运算其效率将优于传统算法, 也预示着量子机器学习算法具有拥面向大数据分析处理的优势.

3.1.2 量子分裂聚类

分裂分层聚类为聚类算法中简单常用的一类算法. 该类算法, 首先将所有数据点视为一个类, 然后不断分裂类簇, 直至每个类簇不能再分裂^[75]. 其算法流程, 通常初始选取所有数据点中尽可能远的两个数据点, 作为初始分裂的两个子类的代表, 如式(17)所示; 其次, 计算其余所有点与这两个点的距离, 根据离这两个点的远近进行分配; 以此类推, 直到每个类内部点的两两距离都小于某阈值, 结束分裂.

$$D_{\max}(C_x, C_y) = \max_{x \in C_x, y \in C_y} \|x - y\| \quad (17)$$

D_{\max} 表示类别 C_x 及 C_y 之间的最大距离. 若有 n 个样本, 每个样本维度为 m , 则传统算法计算该步骤的时间复杂度为 $O(n^2)$. 显然, 若处理大量样本, 且维数较高的数据集时, 经典算法的计算效率就会较低. 2007 年, Aïmeur 等人提出了通过结合量子计算, 实现快速寻找最大距离点这个子步骤^[19]. 该方法的核心步骤是使用 Durr 等人于 1999 年提出的 Grover 变体算法^[76], 通过该算法提高寻找数据集中距离最远的两点的计算效率. 以下为该算法的核心流程思想.

Quantum divisive clustering(D)

输入: 数据集 D

输出: 聚类后数据集 D'

1. If (D 内数据点均满足距离相似性条件)
2. Return D
3. Else{
4. 初始化 $d_{\max} = 0$
5. Do
6. Grover 变体算法寻找最远距离点 a, b
7. If $d(a, b) > d_{\max}$
8. $d_{\max} = d(a, b)$
9. While(存在 $d(a, b) \geq d_{\max}$)
10. For each $x \in D$
11. 将 x 分配到对应的类簇 a 或类簇 b
12. End for
13. 将 D 分裂形成类簇 D_a 和 D_b
14. Quantum divisive clustering(D_a)
15. Quantum divisive clustering(D_b)
16. End if

该量子分裂聚类算法主要依靠 Grover 变体算法来解决数据集的最值问题, 相比于对应的传统算法, 其提速效果主要源于此部分. 该算法对传统分裂聚类算法的主体框架并未进行改变, Aïmeur 等人只是将传统分裂聚类算法中计算复杂性较高的部分单独提取出来, 通过量子搜索算法进行提速. 这种各取所长的有机结合思路, 也为后面众多的量子机器学习算法提供了很好的借鉴意义.

3.2 有监督量子分类算法

3.2.1 量子最近邻算法

最近邻算法为传统机器学习算法中关于分类的一种简单且有效的算法. 该类算法核心思想是, 对一个样本 x , 在特征空间中找出与该样本最邻近的 k 个点, 然后根据这 k 个点的类别, 使用分类决策规则(例如多数表决), 决定 x 所属类别.

相较于传统最近邻分类算法, Wiebe 等人采用与 Lloyd 相似的研究思路, 提出了量子最近邻算法^[21]. 首先他将经典信息中的非零数据编码到量子态的概率幅值上, 通过 controlled-SWAP 电路计算相应两量子态的内积来确定量子态之间的距离. Wiebe 等人通过式(18)、(19)对向量 u 和向量 v 进行编码, 其中向量中非零特征使用负指数形式表示 $v_{j_i} = |v_{j_i}| e^{i\phi_{j_i}}$, $u_{j_i} = |u_{j_i}| e^{i\phi_{j_i}}$, v 为已知分类结果的向量, d 为非零特征数.

$$|u\rangle = \frac{1}{\sqrt{d}} \sum_{u_{j_i} \neq 0} |i\rangle \left[\sqrt{1 - \frac{|u_{j_i}|^2}{r_{\max}^2}} e^{-i\phi_{j_i}} |0\rangle + \sqrt{\frac{|u_{j_i}|^2}{r_{\max}^2}} |1\rangle \right] |1\rangle \quad (18)$$

$$|v\rangle = \frac{1}{\sqrt{d}} \sum_{v_{j_i} \neq 0} |i\rangle |1\rangle \left[\sqrt{1 - \frac{|v_{j_i}|^2}{r_{\max}^2}} e^{-i\phi_{j_i}} |0\rangle + \sqrt{\frac{|v_{j_i}|^2}{r_{\max}^2}} |1\rangle \right] \quad (19)$$

r_{\max} 为特征值上界, 对 $|u\rangle$ 和 $|v\rangle$ 进行 controlled-SWAP 操作后即可得式(20).

$$|\langle u | v \rangle|^2 = (2P(|0\rangle) - 1)d^2 r_{\max}^2 \quad (20)$$

Wiebe 等人也尝试使用两量子态的欧式距离来判定其分类规则(详见文献[21]), 但实验结果显示该方法相比于内积方法, 迭代次数更多, 精度也低于内积方法. 故而, 量子欧式距离分类算法并未得到推广.

3.2.2 量子支持向量机

支持向量机(Support Vector Machine, SVM), 在传统机器学习算法中, 是一种重要的监督类线性分类算法, 其思想是通过寻找最大间隔分类超平

面进行分类^[77], 寻找最大间隔超平面问题可等价于式(21).

$$\arg \max_{\mathbf{w}, b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_i [t_i (\mathbf{w}^\top \phi(\mathbf{x}_i) + b)] \right\} \quad (21)$$

上式中 \mathbf{w} 是超平面的法向量, b 是偏置, \mathbf{x}_i 为第 i 个训练样本, $t_i \in \{-1, 1\}$ 为该训练样本的标签, $\|\mathbf{w}\|^{-1} t_i (\mathbf{w}^\top \phi(\mathbf{x}_i) + b)$ 为样本点到超平面的距离. 若分类正确, 则 $t_i (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) > 0$, 否则 $t_i (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) < 0$.

其中, 最接近分类超平面的样本点 \mathbf{x}_i , 这些点也称为支持向量. 针对于这些支持向量, 需要寻找一个超平面, 使得这些支持向量到该超平面的距离最大. 通过超平面的尺度变换可将问题转化为式(22).

$$\arg \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (22)$$

$$\text{s. t. } t_n (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1$$

寻找距超平面距离最大的支持向量, 即一个优化求解问题, 故而可通过 Lagrangian 方法解决. 通常引入核方法, 将线性不可分的低维特征 $(\mathbf{x}, \mathbf{x}')$ 变换成线性可分的高维特征 $k(\mathbf{x}, \mathbf{x}')$, 再对高位特征进行分类^[77], 特征空间映射示意如图 5 所示.

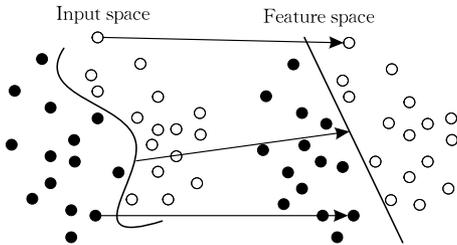


图 5 SVM 特征空间映射示意图^①

关于量子支持向量机算法的有关研究, 最早是 Anguita 等人于 2003 年使用量子计算的方法, 解决了 SVM 的训练效率问题^[78]. 随后, Reberstrost 等人在 2014 年提出量子版本的 SVM, 其核心思想是利用量子算法解决训练数据的内积运算问题(核方法)^[22]. Reberstrost 等人首先将特征向量的各维特征编码至量子态概率幅 $|x_i\rangle = |\mathbf{x}_i|^{-1} \cdot \sum_{j=1}^N (\mathbf{x}_i)_j |j\rangle$,

其中 N 为特征维度, $(\mathbf{x}_i)_j$ 为第 i 个特征向量的第 j 个特征, $|\mathbf{x}_i|$ 为特征向量范数. 其次, 制备如下量子态, 如式(23)所示.

$$|\chi\rangle = (\sqrt{N_\chi})^{-1} \cdot \sum_{i=1}^M |x_i\rangle |i\rangle |x_i\rangle \quad (23)$$

其中, $N_\chi = \sum_{i=1}^M |\mathbf{x}_i|^2$, \mathbf{x}_i 为第 i 个训练样本. Reberstrost 等人之后将核矩阵和量子系统的密度矩阵联系起来. 由于核矩阵的每个元素为向量间的内积 $K_{ij} =$

$\mathbf{x}_i \cdot \mathbf{x}_j$, 且 $\langle x_j | x_i \rangle = (|\mathbf{x}_i|^{-1} \mathbf{x}_i) \cdot (|\mathbf{x}_j|^{-1} \mathbf{x}_j)$, 此时通过求密度矩阵 $|\chi\rangle\langle\chi|$ 的偏迹即可得到归一化核矩阵.

$$\begin{aligned} \text{tr}_\chi (|\chi\rangle\langle\chi|) &= \frac{1}{N_\chi} \sum_{i,j=1}^M |\mathbf{x}_i| |\mathbf{x}_j| \langle x_i | x_j \rangle |i\rangle\langle j| \\ &= \frac{K}{\text{tr}(K)} \end{aligned} \quad (24)$$

其中, $\mathbf{x}_i \cdot \mathbf{x}_j = |\mathbf{x}_i| |\mathbf{x}_j| \langle x_i | x_j \rangle$. 通过该方法, 就使量子系统同传统机器学习的核矩阵联系起来, 由于量子态之间的演化运算具有高并行性, 通过该方法即可完成传统机器学习中对核矩阵计算的加速.

Reberstrost 等人也提出了量子版本最小二乘法支持向量机(Least-Squares Support Vector Machine, LS-SVM). 他们将传统 SVM 分类问题转化为求解式(25)的问题.

$$\begin{aligned} \hat{F}|b, \mathbf{d}\rangle &= |y\rangle \\ \text{s. t. } \|F\| &\leq 1 \end{aligned} \quad (25)$$

其中, b 为偏置, 向量 \mathbf{d} 的每个元素为支持向量距最优超平面的距离, $|b, \mathbf{d}\rangle$ 即表示分类超平面, 且 $(b, \mathbf{d})^\top = F^{-1}(0, \mathbf{y})^\top$, $\mathbf{y} = (y_1, y_2, \dots, y_M)$ 为训练样本的标签, 求解过程类似 HHL 算法^[27]. $F = \begin{pmatrix} 0 & \mathbf{1}^\top \\ \mathbf{1} & \mathbf{K} + \gamma^{-1} \mathbf{1} \end{pmatrix}$, \mathbf{K} 为核矩阵, γ 为可接受误差权重.

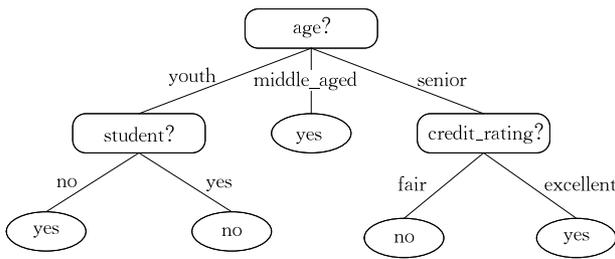
最后, 分类任务即转化为使用 controlled-SWAP 操作来比较 $|b, \mathbf{d}\rangle$ 和输入样本 $|x\rangle$ 的距离, 从而得到 $|x\rangle$ 所属类别.

量子支持向量机算法的物理实验也有一些进展. 最近 Li 等人通过核磁共振的方法, 在物理上实现了 4 量子比特的量子 SVM^[32], 并对最基本的手写数字 6 和 9 进行识别, 实验结果显示识别精度高达 99%. 虽然实验样本较小, 但该实验显示出量子理论与机器学习算法结合的可行性.

3.2.3 量子决策树算法

决策树模型是一种描述对象属性或特征, 并与对象所属类别之间进行关系映射, 所形成的树形结构模型^[79]. 树中每个节点代表一个对象, 分为内部节点和叶节点(即最后一层节点)两种. 内部节点代表对象的属性值, 叶节点代表对象的类别. 决策树分类过程, 如图 6 所示. 分类, 首先从根节点开始, 对输入实例的特征进行判断, 并根据判别结果将实例分配至相应子节点, 以此类推, 直到对象到达叶节点. 最终得到该实例所在类别. 为提高决策树学习效率, 常使用信息增益来选择特征.

① <http://www.statsoft.com/Textbook/Support-Vector-Machines>

图 6 决策树分类过程实例图^[80]

最早 Farhi 等人于 1997 年提出量子计算与决策树模型相结合的研究,其指出量子方法实现的决策树算法,在部分情况下相比于传统决策树算法更加高效^[81]. 2014 年 Lu 和 Braunstein 提出了量子决策树分类器^[80].

Lu 和 Braunstein 首先将样本数据 $\{\mathbf{x}^{(k)}, \mathbf{y}^{(k)}\}_{k=1}^m$ 转换成量子态 $\{|x_i\rangle, |y_i\rangle\}_{i=1}^m$, 其中 $\mathbf{x}^{(k)} \in R^d$ 为特征向量, d 为特征维数, $\mathbf{y}^{(k)} \in \langle C_j \rangle_{j=1}^m$ 为所属类别.

他们在选取最优特征时,将冯诺依曼熵作为判断依据,如式(26)所示.

$$S(\boldsymbol{\rho}) = -\text{tr}(\boldsymbol{\rho} \log \boldsymbol{\rho}) \quad (26)$$

其中, $\boldsymbol{\rho} = \sum_{i=1}^n p_i |y_i'\rangle \langle y_i'|$ 为第 t 个节点的对应类别量子态的密度矩阵, $|y_i'\rangle$ 为在第 t 个节点第 i 类的量子态, n 表示类的总数. 然后,通过计算该节点 t 的冯诺依曼熵的期望值,如式(27)所示.

$$S_e(\rho_i^{(t)}) = \sum_{j=1}^{t_i} p_j S(\rho_{i,j}^{(t)}) \quad (27)$$

其中, t_i 表示该节点的子节点数. 每个节点会产生不同的期望值,即 $\{S_e(\rho_{i,j}^{(t)})\}_{j=1}^{t_i}$, 最后选取其中期望值最小的 $S_e(\rho_{i,j}^{(t)})$, 所对应的特征作为最优特征. 算法具体步骤详见文献^[80].

该量子算法与传统决策树算法的不同之处在于,其使用量子信息中的冯诺依曼熵,来替换经典信息中的香农熵,通过计算其期望值,进而计算特征值.

3.2.4 量子神经网络

人工神经网络是一种仿生计算模型,通过模拟生物神经网络的结构和功能而得名^[82]. 人工神经网络也是一种非线性的数据建模工具. 该模型由大量节点构成,这些节点也称为神经元. 层与层之间神经元根据不同的权重相连接,形成网状结构的模型,每层的神经元都含有一个激活函数. 网络的第一层为输入层,最后一层为输出层,中间层为隐含层.

如图 7 所示,该图为神经网络的一部分,是第 i 层节点到第 $i+1$ 层第 j 个节点的连接示意图,网络中其他点的连接情况类似. 图 7 中左侧节点 $\{x_j^i\}_{j=1}^n$

为第 i 层的神经元,它们通过权重 $\{\omega_{kj}^i\}_{k=1}^n$ 与 $i+1$ 层的第 j 个节点相连,且 $x_j^{i+1} = \sum_{k=1}^n \omega_{kj}^i x_k^i$. f 为激活函数,通常为非线性函数,例如常见的 Sigmoid 函数. 图 7 的输出函数可表示为式(28).

$$x_j^{i+1} = f(\mathbf{w}_{ij} \mathbf{x}^i + b) \quad (28)$$

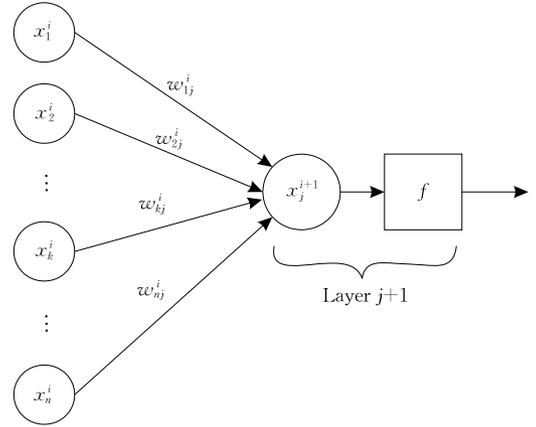


图 7 神经网络示意图

神经网络的训练过程是,将特征向量输入网络,根据网络处理后的输出结果,优化以网络权重为参数的损失函数. 其目的是,经过训练网络输出结果与标签的误差最小. 神经网络常使用反向传播(Backpropagation, BP)算法进行训练. 该方法主要包含两个阶段: (1) 前向传播. 根据式(28)计算规则,由输入层向输出层逐层计算; (2) 反向传播. 计算输出层与标签的误差,对损失函数使用梯度下降进行最优化,从输出层向输入层反向更新网络中各层权重. 每一个训练样本均进行一次向前计算和反向更新的操作,最终至网络收敛.

相关研究指出,人脑的信息处理过程与量子效应相关,并且生物神经网络的动力学特征与量子系统的动力学特征相似,故产生了量子理论与生物神经网络相结合的研究^[83]. Kak 于 1995 年,将神经网络和量子计算的概念相结合,首次提出量子神经网络计算^[10]. 同年, Menneer 等人提出了量子衍生神经网络,传统神经网络需使用数据集对同一网络进行训练,从而找到适合不同模式的网络参数. 而他则利用量子叠加性原理,对同一模式训练多个同构网络,得到不同模式对应的同构网络的量子叠加^[84]. Behrman 等人于 1996 年,首先从数学形式上提出了量子点神经网络的概念^[11]. 他们发现基于量子点的时间演化模型能够完成神经网络中的前向或反向计算,之后他们从不同方面对量子神经网络做了一系列研究^[85-88]. 同年, Toth 等人提出了量子细胞神经网络,其将网络中每个细胞视为一个量子系统,并使

用含时薛定谔方程来描述该量子系统的演化过程^[12]. 1998年, Ventura等人结合量子理论中的叠加性原理, 将网络的权重向量替换为希尔伯特空间中的量子态表示, 对神经网络的训练, 即对应这些量子态的演化^[13], 同年他又提出了量子联想存储算法(Quantum Associative Memory, QUAM), 其相比于传统存储有着指数级提升^[89]. 1998年, Menneer的博士论文以多宇宙理论为切入点, 系统地研究了量子神经网络的相关问题^[90]. 1999年, Pribram等人指出可以使用量子理论来解释人脑的处理过程^[91]. 同年, Zhou等人将量子神经网络应用到了手写数字识别领域^[92]. 2000年, Narayanan等人对不同结构的量子神经网络进行了研究, 指出量子神经网络相比传统神经网络训练更加高效, 但全量子结构的网络模型并非更优^[93]. 同年, Ezhov等人对量子神经网络进行了综述性研究^[94]; 同时 Matsui首次提出了通过量子比特作为神经元, 使用量子旋转门和受控非门构造神经网络^[14]. 2001年, Gupta发现可使用阈值电路(threshold circuits)构建量子神经网络, 该方案相比传统方案需要更少的资源, 同时他也证明该网络的计算本质上等同传统方法^[95]. 同年, Altaisky提出了一种物理上简单可行的量子神经网络, 该量子神经网络的输入输出可用偏振光实现, 网络的权重可用平面分束器实现^[96]. 2002年, 开始不断出现基于量子神经网络应用研究, 例如 Kouda等人使用三层量子神经网络应用于图像压缩, 该方法相较于传统方法其压缩效率更优^[97]. 2005年, 他们又通过4位和6位的奇偶校验法验证了基于量子比特的神经网络优于传统神经网络^[98]. 2006年, Zhou等人提出了仅适用单个神经元即可实现传统异或操作的量子神经网络, 解决传统神经网络需两层才能解决的线性不可分问题^[15]. 2007年, 李盼池等人提出了一种量子自组织特征映射神经网络模型^[99]. 该模型有两层, 包含输入层与竞争层, 神经元由量子态表示, 其通过计算样本量子态与相应权值量子态的相似系数来提取聚类样本所隐含的模式特征. 2008年, Silva等人做了一系列无权重神经网络模型的不同实现方案^[59-62, 100-101], 同时也在2013年给出了不同类型量子神经网络模型的分析比较^[102]. Schuld等人对现有量子神经网络做了系统的比较分析^[103]. 他们指出量子神经网络的框架应满足3个条件: (1) 神经网络的输入和输出都必须是经过编码的数据, 通常将输入输出的二进制串, 编码至量子态上; (2) 量子神经网络必须表达基本神经计算的原理及结构; (3) 量子神经网络中的演

化过程, 必须服从量子力学规律, 如叠加性、纠缠性等. 但 Schuld指出现今的量子神经网络方法无法完全满足上述条件, 并提出了使用量子随机游走的方式构建人工神经网络的方法^[16].

虽然量子神经网络模型的实现方式有很多, 但大致可分为以下几类: 基于测量的量子神经网络; 基于量子点的神经网络; 基于量子电路的神经网络; 基于量子随机行走的神经网络; 基于量子比特的神经网络等.

(1) 基于测量的量子神经网络

基于测量的量子神经网络最先由 Kak提出, 他将网络视为一个量子系统 $|\psi\rangle$, 并将网络的权重视为该系统的可观测量 $\hat{\omega}$ (可观测量由厄米算符表示), 则有该算符对应的本征方程 $\hat{\omega}|\psi\rangle = \lambda|\psi\rangle$. 网络权值的更新过程, 即为对系统不断进行测量操作, 找出本征值 $\lambda=1$ 所对应 $\hat{\omega}$ 的过程. 另外一种基于测量的方法来自于 Menneer等人, 其将整个神经网络的权值(存储多种模式), 视为多个网络权值(存储一种模式)的叠加, 识别某个特定模式即为通过测量操作, 使权值的叠加态塌缩到该模式的对应权值上. 由于Kak的研究使用线性算符 $\hat{\omega}$ 来表征系统演化, 而传统神经网络的映射是非线性的, 这两者是否等价还有待商榷.

(2) 基于量子点的神经网络

基于量子点神经网络的相关研究中, 以 Behrman的研究最具代表性. Behrman提出的神经网络与其他神经网络不同, 整个网络只有一个神经元, 网络权值的更新是通过该神经元在不同时间的状态变化来进行, 并且可通过该方式实现传统神经网络的前向及反向传播计算.

(3) 基于量子门的神经网络

基于量子门电路的神经网络, 其代表性方法是由 Matsui和 Maeda等人提出的^[14, 104]. Matsui等人使用一位量子旋转门及受控非门来构建神经网络, 以量子态作为输入, 根据传统前向传播算法提出了复数形式下的网络权值更新过程, 其网络结构如图8所示. 由于使用通用门来进行网络的构建, 当网

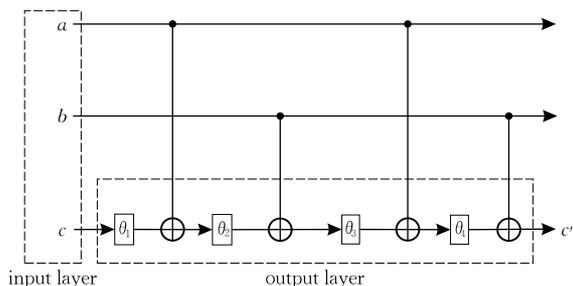


图8 基于量子门的神经网络

络复杂后,通用门的数量也会成倍增加,其并未考虑该构建方法是否为最优的量子门电路综合算法。

(4) 基于量子比特的神经网络

Kouda 等人提出了使用量子比特 $|1\rangle$ 和 $|0\rangle$ 分别作为处于激活与未激活状态的神经元^[98],并将神经网络结构定义为图 9 所示,Zhou 等人也在该网络的构建方法的基础上进行了研究^[15].图 9 中 $\arg(u)$ 为 u 的相位角, $g(\delta)$ 为 Sigmoid 函数, $f(\theta_j) = \cos(\theta_j) +$

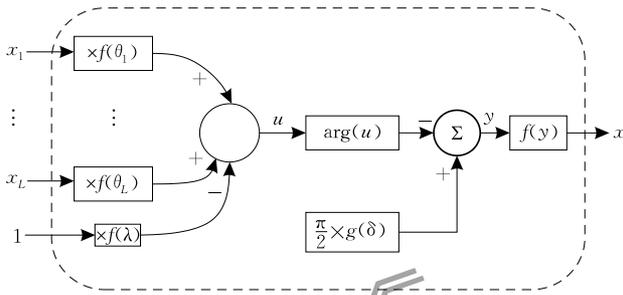


图 9 基于量子比特的神经元^[98]

$i \sin(\theta_j) = e^{i\theta_j}$. 虽然 Kouda 等人的实验表明该模型好于传统神经网络以及复数神经网络,他认为该优势来自于量子的特有性质(如叠加性等),但他并未给出相应的数学证明。

3.3 量子降维算法

机器学习需要处理的数据特征通常是高维,特别是计算机视觉及自然语言处理领域.原始特征不但高维,而且包含了大量与学习问题相关度小、甚至不相关的信息.部分算法会随着维度的提高,出现高维空间中无法工作的情况.因此,降维操作在机器学习中十分重要.降维操作降低了数据复杂程度,剔除了不必要的无用信息,使模型能在降维之后进行有效的学习.经典机器学习常用的降维操作包括主成分分析(Principle Component Analysis, PCA)^[105]、线性判别分析(Linear Discriminant Analysis, LDA)^[106]和流行学习等,其中常用到的是 PCA,该方法示意如图 10 所示。

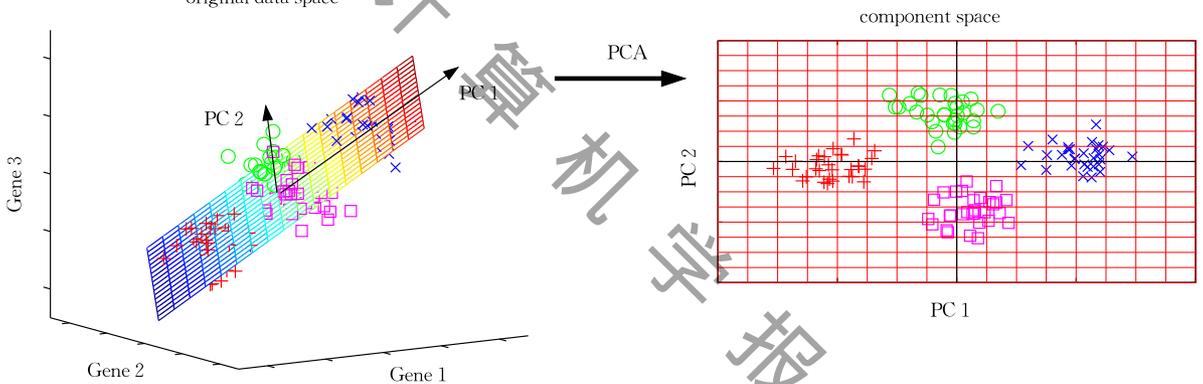


图 10 主成分分析示意图^①

近年来,也有研究人员着眼于将量子的叠加性、纠缠性原理同降维算法相结合,试图借助量子计算对高维数据处理的优点,以降低传统降维算法的计算复杂度.国外有 Lloyd 教授团队首次提出的量子主成分分析、Cong 等人提出的量子线性判别分析,国内杨国武教授团队以及我们的研究小组,也在尝试通过量子方法解决高维数据降维问题。

3.3.1 量子主成分分析

Lloyd 等人在 2014 年提出量子主成分分析算法(Quantum Principal Component Analysis, QPCA).由于量子系统的密度矩阵均为厄米矩阵,其可表示成 Gram 矩阵的形式,可看作是一组向量的协方差矩阵^[26],因此该方法使用量子系统的多个密度矩阵副本构造对应特征值较大的特征向量。

Lloyd 等人指出 QPCA 也适用于量子态的判别.有两组量子态集合 $\{|\phi_i\rangle\}_{i=1}^m$ 和 $\{|\varphi_i\rangle\}_{i=1}^m$, 两组

量子态的密度矩阵为 $\rho = m^{-1} \sum_i |\phi_i\rangle\langle\phi_i|$ 和 $\sigma = m^{-1} \sum_i |\varphi_i\rangle\langle\varphi_i|$. 判别新量子态 $|\chi\rangle$ 归属于哪组,可对量子态 $|\chi\rangle$ 进行密度矩阵连乘(该技术涉及到 Hamiltonian Simulation, 详见文献[26])及量子相位估计操作,如式(29)所示。

$$|\chi\rangle|0\rangle \rightarrow \sum_j \chi_j |\xi_j\rangle |x_j\rangle \quad (29)$$

其中, $|\xi_j\rangle$ 为 $\rho - \sigma$ 的特征向量, x_j 则为对应的特征值. 对 $\sum_j \chi_j |\xi_j\rangle |x_j\rangle$ 的第二个量子比特进行测量. 测量结果若为正,则判定为 $\{|\phi_i\rangle\}_{i=1}^m$ 集合,反之则为 $\{|\varphi_i\rangle\}_{i=1}^m$. 虽然 QPCA 对传统 PCA 具有一定加速作用,但是其假设使用 QRAM 对量子态进行制备^[26],而 QRAM 至今被认为仅限于理论模型,并未出现真正可靠的物理实现方案。

① http://www.nlpc.a.org/fig_pca_principal_component_analysis.png

3.3.2 量子线性判别分析

线性判别分析也可称为 Fisher 线性判别分析, 该方法最先由 Fisher 于 1936 年提出^[107], 通过计算类间数据散布度与类内数据散布度的比值, 寻找最优区分数据的投影方向. 其类间散布度 S_b 及类内散布度 S_w 如式(30)、(31)所示.

$$S_b = \sum_{i=1}^{N_c} (\mu_i - \bar{x})(\mu_i - \bar{x})^T \quad (30)$$

$$S_w = \sum_{i=1}^{N_c} \sum_{x \in C_i} (x - \mu_i)(x - \mu_i)^T \quad (31)$$

其中, N_c 为类别个数, μ_i 为第 i 类数据的中心, \bar{x} 为数据集的中心, x 属于类 C_i . Fisher 判别函数为

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}} \quad (32)$$

最佳投影方向为 $J(\mathbf{w})$ 最大的那个方向, 即类间数据散布度尽可能的大, 类内散布度尽可能的小. 其等价于目标函数式(33), 通过对该最优化问题的求解即可得到最优区分数据的投影方向 \mathbf{w} .

$$\begin{aligned} \max \quad & \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}} \\ \text{s. t.} \quad & \|\mathbf{w}^T S_w \mathbf{w}\| = 1 \end{aligned} \quad (33)$$

2015 年, Cong 等人受到 HHL 算法的启发, 提出通过厄米矩阵连乘, 来实现线性方程组求解的问题, 并将其与线性判别分析结合, 提出量子线性判别分析^[108]. 该方法同 QPCA 一样, 都假设已有可靠的 QRAM 物理实现为基础, 并以其能高效制备量子态为前提.

Cong 等人提出的方法, 首先使用 QRAM 制备态 $|\phi_1\rangle$ 和 $|\phi_2\rangle$.

$$\begin{aligned} |\phi_1\rangle &= \text{QRAM} \left(\frac{1}{\sqrt{N_c}} \sum_{i=1}^{N_c} |i\rangle |0\rangle |0\rangle \right) \\ &= \frac{1}{\sqrt{N_c}} \sum_{i=1}^{N_c} \|\mu_i - \bar{x}\| |i\rangle \|\mu_i - \bar{x}\rangle |\mu_i - \bar{x}\rangle \end{aligned} \quad (34)$$

$$\begin{aligned} |\phi_2\rangle &= \text{QRAM} \left(\frac{1}{\sqrt{M}} \sum_{j=1}^M |j\rangle |0\rangle |0\rangle \right) \\ &= \frac{1}{\sqrt{M}} \sum_{j=1}^M \|x_j - \mu_{c_j}\| |j\rangle \|x_j - \mu_{c_j}\rangle |x_j - \mu_{c_j}\rangle \end{aligned} \quad (35)$$

然后通过这两个量子态的密度矩阵来表示 S_b 和 S_w , 即

$$\begin{aligned} S_b &= \frac{1}{A} \sum_{i=1}^{N_c} \|\mu_i - \bar{x}\|^2 |\mu_i - \bar{x}\rangle \langle \mu_i - \bar{x}|, \\ A &= \sum_{i=1}^{N_c} \|\mu_i - \bar{x}\|^2 \end{aligned} \quad (36)$$

$$\begin{aligned} S_w &= \frac{1}{B} \sum_{i=1}^{N_c} \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 |x_j - \mu_i\rangle \langle x_j - \mu_i|, \\ B &= \sum_{j=1}^M \|x_j - \mu_{c_j}\|^2 \end{aligned} \quad (37)$$

由于式(37)可通过拉格朗日乘子法求解后可得 $S_w^{-1} S_b \mathbf{w} = \lambda \mathbf{w}$. 令 $\mathbf{w} = S_b^{-1/2} \mathbf{v}$, 通过量子相位估计得到 \mathbf{v} 的特征值及特征向量. 最后, 通过厄米矩阵连乘求解得出最优投影方向 \mathbf{w} , 具体步骤详见文献[108].

该算法与 QPCA 算法类似, 都将原问题数据的协方差矩阵与量子系统的密度矩阵联系起来, 之后对密度矩阵的特征值与特征向量进行研究, 从而找出最优投影方向或主要特征向量.

3.4 量子深度学习

量子深度学习最早由 Wiebe 等人提出^[23], 他们实现的是量子版本的玻尔兹曼机 (Restricted Boltzmann Machine, RBM). 受限玻尔兹曼机最早由 Hinton 等人在 1985 年发明^[109], 并提出了该网络的快速训练算法^[110], 该算法使用能量函数来确定可视层和隐含层的联合概率分布. 受限玻尔兹曼机结构如图 11 所示.

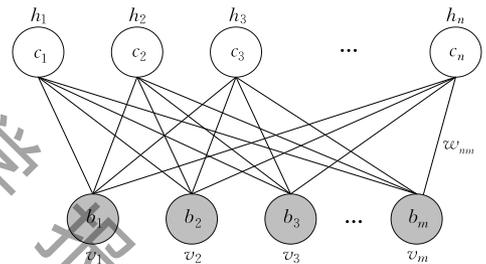


图 11 RBM 网络结构图^①

网络中 w_{ij} 为可视层的第 i 个节点到隐含层第 j 个节点的权重, $\mathbf{h} = \{h_i\}_{i=1}^n$ 为隐藏层节点, $\mathbf{v} = \{v_j\}_{j=1}^m$ 为可视层节点, $\mathbf{c} = \{c_i\}_{i=1}^n$ 为隐藏层偏置, $\mathbf{b} = \{b_j\}_{j=1}^m$ 为可视层偏置.

网络中输入输出的联合概率分布, 服从 Gibbs 分布, 如式(38)所示.

$$P(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v}, \mathbf{h})}}{Z} \quad (38)$$

其中, \mathbf{v} 代表可视层单元, \mathbf{h} 代表隐含层单元, $Z = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}$ 为归一化因子. 式(38)中的能量 E 由易辛模型 (Ising Model) 给出, 如式(39)所示.

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{i=1}^n \sum_{j=1}^m w_{ij} h_i v_j - \sum_{j=1}^m b_j v_j - \sum_{i=1}^n c_i h_i \quad (39)$$

学习问题为寻找网络参数 $\{\mathbf{w}, \mathbf{b}, \mathbf{c}\}$, 在特定输入

① http://image.diku.dk/shark/sphinx_pages/build/html/_images/rbm_graph

为 $\{v_j\}_{j=1}^m$ 的情况下的最大化代价函数(40).

$$L(\mathbf{w}, \mathbf{b}, \mathbf{c}, \mathbf{v}) = \operatorname{argmax}_{\mathbf{w}, \mathbf{b}, \mathbf{c}} \prod_{i=1}^N P(\mathbf{v}^{(i)}) \quad (40)$$

$\mathbf{v}^{(i)}$ 代表第 i 个样本且假设服从独立同分布, 由于处理乘积较为复杂, $P(\mathbf{v})$ 为似然函数, 故可将问题转化成等价形式, 见式(41).

$$\ln L(\mathbf{w}, \mathbf{b}, \mathbf{c}, \mathbf{v}) = \operatorname{argmax}_{\mathbf{w}, \mathbf{b}, \mathbf{c}} \sum_{i=1}^N \ln \prod_{i=1}^N P(\mathbf{v}^{(i)}) - \frac{\lambda}{2} \mathbf{w}^T \mathbf{w} \quad (41)$$

$\frac{\lambda}{2} \mathbf{w}^T \mathbf{w}$ 为正规化项, 作用是避免训练后出现过拟合问题. 当网络为受限玻尔兹曼机时, 代价函数对网络权重的梯度为

$$\frac{\partial \ln L(\mathbf{w}, \mathbf{b}, \mathbf{c}, \mathbf{v})}{\partial \mathbf{w}_{ij}} = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}} - \lambda \mathbf{w}_{ij} \quad (42)$$

其中, $\langle f(\mathbf{v}, \mathbf{h}) \rangle_{\text{data}}$, $\langle f(\mathbf{v}, \mathbf{h}) \rangle_{\text{model}}$ 分别为

$$\langle f(\mathbf{v}, \mathbf{h}) \rangle_{\text{data}} = \frac{1}{N} \sum_{\mathbf{v} \in v_{\text{train}}} \sum_{\mathbf{h}} \frac{f(\mathbf{v}, \mathbf{h}) e^{-E(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}} \quad (43)$$

$$\langle f(\mathbf{v}, \mathbf{h}) \rangle_{\text{model}} = \sum_{\mathbf{v}, \mathbf{h}} \frac{f(\mathbf{v}, \mathbf{h}) e^{-E(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}} \quad (44)$$

对于求解 RBM 代价函数的梯度估计及训练问题^[20], Wiebe 等人提出使用量子采样 (Gradient Estimation via Quantum Sampling, GEQS) 及量子概率幅值估计^[111] (Gradient Estimation via Quantum Amplitude Estimation, GEQAE) 两种算法. 接下来, 将针对于这两种算法详细介绍.

3.4.1 GEQS 算法

量子采样梯度估计算法, 首先制备与网络输入及输出的概率分布相似的量子分布. 使用平均场近似 $Q(\mathbf{v}, \mathbf{h})$ 有效地替代 Gibbs 采样. 然后对制备的量子态进行测量操作, 得到 $\langle v_i h_j \rangle_{\text{data}}$ 及 $\langle v_i h_j \rangle_{\text{model}}$. 最后根据式(42)计算得到权重梯度. 算法中 Gibbs 分布的近似分布为式(45).

$$P^{MF}(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v}, \mathbf{h})}}{Z^{MF}} \quad (45)$$

令 $P_{\text{norm}}(\mathbf{v}, \mathbf{h}) = \frac{P^{MF}(\mathbf{v}, \mathbf{h})}{k Z^{MF} Q(\mathbf{v}, \mathbf{h})}$, k 为固定常数项, 则有

$$P_{\text{norm}}(\mathbf{v}, \mathbf{h}) Q(\mathbf{v}, \mathbf{h}) \propto P(\mathbf{v}, \mathbf{h}) \quad (47)$$

使用辅助量子比特将 $P_{\text{norm}}(\mathbf{v}, \mathbf{h}) Q(\mathbf{v}, \mathbf{h})$ 编码至量子态的概率幅上, 其中最后一位量子比特变换如式(48)所示.

$$|0\rangle \rightarrow (\sqrt{1 - P_{\text{norm}}(\mathbf{v}, \mathbf{h})} |0\rangle + \sqrt{P_{\text{norm}}(\mathbf{v}, \mathbf{h})} |1\rangle) \quad (48)$$

之后进行投影测量最后一位量子比特, 若结果为 $|1\rangle$, 则波包塌缩后到近似 Gibbs 分布. 使用上述方法作用于模型的输入与输出, 即可得到模型的 Gibbs 分布, 最后计算可得模型的梯度.

3.4.2 GEQAE 算法

量子概率幅值梯度估计算法与 GEQS 大体思路相同, 不同之处是: 该方法量子态的制备使用 QRAM 来实现^[24-25]. 算法首先利用 QRAM 将训练数据转为量子态表示.

$$U \left(\frac{1}{N} \sum_{i=1}^N |i\rangle |0\rangle \right) = \frac{1}{N} \sum_{i=1}^N |i\rangle |v_i\rangle \quad (49)$$

随后对式(49)所得量子态进行如式(48)的操作, 可得式(50).

$$\frac{1}{N} \sum_{i=1}^N |i\rangle |v_i\rangle \sum_{\mathbf{h}} \sqrt{Q(\mathbf{v}_i, \mathbf{h})} |\mathbf{h}\rangle |P_{\text{norm}}(\mathbf{v}_i, \mathbf{h})\rangle |\chi(\mathbf{v}_i, \mathbf{h})\rangle \quad (50)$$

最后 $\langle v_i h_j \rangle_{\text{data}}$ 则可通过式(51)得到, $\langle v_i h_j \rangle_{\text{model}}$ 同理.

$$\langle v_i h_j \rangle_{\text{data}} = P(v_i = h_j = 1 | \chi = 1) \quad (51)$$

式(51)中各项概率值, 均可通过对式(50)相应的量子寄存器, 进行量子概率幅值估计得到.

最近 Wiebe 等人在通过量子方法解决 RBM 训练问题中得到启发, 提出了新的量子 RBM 训练方法: Instrumental Rejection Sampling 算法^[112]. 该算法相较于传统对比散度算法 (Contrastive Divergence, CD), 拥有更加精确的梯度近似值^[112]. Instrumental Rejection Sampling 算法用于目标函数的梯度进行近似. 其思想源于 GEQS 算法, 既然 GEQS 算法使用量子采样, 是否存在对应的经典模拟, 且尽可能的保留原有量子算法的优点. 据此 Wiebe 提出了 GEQS 的经典模拟算法 Instrumental Rejection Sampling 进行玻尔兹曼机的训练. 这表示我们不仅可以通过量子计算机实现对传统机器学习的加速, 还可借用量子物理的概念, 抽象出数学模型, 并将其与传统机器学习相结合, 产生新的学习算法, 如 Horn 等人在 2007 年提出的量子聚类算法^[17].

Adcock 等人在量子机器学习的研究报告中指出, 除了现有方法, 还有不少研究人员正尝试通过改变 RBM 的物理模型来解决问题^[7]. 例如将 RBM 使用的易辛模型换做海森堡模型, 这是因为海森堡模型在考虑 z 轴自旋时, 还分别考虑了 x 轴和 y 轴的自旋, 这就使得模型更加具有鲁棒性; 把 Gibbs 分布换做 Bose-Einstein 分布或 Fermi-Dirac 分布. 这是因为 Gibbs 分布为经典机器学习系统中的分布, 在量子效应未占主导时可用其进行分析, 但当量子效应占主导则需考虑 Bose-Einstein 分布和 Fermi-Dirac 分布.

3.5 算法比较

本节对具有代表性的量子机器学习算法, 从时间复杂度、数据是否量子化、是否具有物理实现等方面进行对比分析, 详见表 1 所示. 表项中的量子化表

示输入输出是否为量子态,是否涉及经典数据与量子数据的转换.表 1 列举出了目前量子机器学习中较为有代表性的工作.其中无监督学习算法和监督学习算法中大多都使用了 Grover 的量子搜索相关算法,例如 K -medians、Quantum SVM 等.此外,本领域研究论文很少提及相应算法在真实物理条件下的表现.其原因,一是,现有大多方法假设在已有足够量子比特的环境下运行,只是理论上的分析,但目前物理实现上还不具备实验条件.二是,不少研究基于 QRAM 进行初态的制备,但至今还未出现完美

的 QRAM 物理实现,故无法给出算法的真实表现.量子机器学习算法对经典算法的加速主要来源于两类:(1)经典算法思路不变,通过将处理数据量子化,借助量子计算机的高并行性进行快速处理.首先将经典信息转为量子态,其次输入量子计算机,之后运行相应量子算法实现,即可对其相应的经典步骤进行量子加速;(2)借助量子原理,从算法思路上进行创新,借鉴物理演化过程,改变原有算法,对传统机器学习算法进行加速.

表 1 量子算法对比

算法名称	量子算法复杂度	经典算法复杂度	量子化	物理实现	参数说明
Quantum Weightless Neural Network ^[62]	—	$O(\text{poly}(N))$	是	有	N 为训练样本数
Quantum k -nearest Neighbours ^[21]	$O(\log(nm))$	$O(nm)$	否	否	n 为训练向量维度 m 为训练样本数
Quantum Principle Component Analysis ^[26]	$O(\log(d))$	$O(d)$	是	否	d 为数据空间维度
Quantum Deep Learning (Amplitude Estimation) ^[23]	$O(\sqrt{NE}^2(\sqrt{k} + \max_x \sqrt{k_x}))$	$O(NEk)$	是	否	N 为训练样本数 E 为边的个数 x 为训练集
Quantum Deep Learning (Quantum Sampling) ^[23]	$O(\sqrt{NE}(\sqrt{k} + \max_x \sqrt{k_x}))$	$O(NEk)$	是	否	N 为训练集大小 E 为边的个数 x 为训练集
Quantum K -means ^[18]	$O(k \log(kmn))$	$O(mn)$	是	否	n 为训练向量维度 m 为训练样本数 k 为聚类簇总数
Quantum SVM ^[22,32,78]	$O(\log(mn))$	$O(mn)$	是	有	n 为训练向量维度 m 为训练样本数
Quantum Decision Tree Classifier ^[80]	$\frac{(\sqrt{n}-1)\sqrt{kd}}{\sqrt{n}(\sqrt{k}-1)}(T_1+T_2)$	$O(mn \log(n))$	否	否	n 为训练向量大小 m 为训练向量大小 T_1, T_2 为两个子过程时间,详见文献 ^[91]
Quantum K -medians ^[19]	$O\left(\frac{n^2}{k}\right)$	$O\left(\frac{n^2}{\sqrt{k}}\right)$	否	否	k 为聚类簇总数
Quantum divisive clustering ^[20]	$O(n \log n)$	$O(n^2)$	否	否	n 为训练样本数

4 总结与展望

量子机器学习是结合量子力学原理和机器学习的交叉研究领域,其研究意义和价值为:第一,利用量子计算的高并行性,提高机器学习面对大数据的处理、分析和挖掘能力,如 Lloyd 等人的研究^[18];第二,借鉴量子力学原理,促进新型机器学习算法的产生,如 Wiebe 等人的工作^[112];第三,借鉴传统机器学习算法,提出量子力学领域新的研究方法,如提出新的量子断层分析方法^[68].量子机器学习的研究与进展,给众多的研究者带来了无限的希望和憧憬.让大家感受到了量子计算的巨大威力,虽然这仅仅是量子计算能力的冰山一角,但是我们需要更加需要

理性的看待它.因为,在通用量子计算机建造成功之前,量子算法还很难在实际应用中展现出其数据处理方面强大的能力.目前,量子机器学习还有如下问题需要突破:

(1)量子机器学习算法的发展还处于一个相当初级的阶段,并没有如传统机器学习领域较为完备的理论框架加以指导研究.使用量子特性进行机器学习算法的研究大多还处于理论阶段,只有少部分论文提及具体算法的实现及验证.虽然理论上已经证明量子算法对部分经典计算问题有着可喜的提速效果,但是大多算法并未在量子环境下运行,只能通过模拟的方法实现,故难以展现其真正的数据处理能力.

(2)经典信息到量子信息的转换研究中,大多

数工作基于 Lloyd 等人提出的 QRAM, 将经典信息编码到量子概率幅。但 Adcock 等人指出 QRAM 在具体物理实现上仍有许多问题未能解决, 例如量子的退相干效应, 所以这种读写方案在实际中是很难实现的^[7]。虽然有研究人员对 QRAM 进行了分析与改进, 从不同方面进行了相关尝试^[113-114], 但依旧未做出能够有效存储任意量子态的 QRAM 方法的物理实现。

(3) 在量子机器学习算法的设计上, Adcock 也指出将数据编码到概率幅上会导致产生巨大 Hilbert 空间^[7]。由于从量子信息中提取出经典信息主要通过量子测量操作, 巨大的 Hilbert 空间将导致量子测量难以提取出经典信息。

近年来量子机器学习吸引着越来越多的科研团队(如 MIT Seth Lloyd, Oxford University Quantum Optimization and Machine Learning Group 等)、企业(如 Google Quantum A. I. Lab, Quantum Architectures and Computation Group 等)和个体研究者, 该领域的研究不仅可以推动机器学习发展, 提高机器学习的学习效率和学习精度, 也可促进量子世界的繁荣发展。加之, 大数据时代来临、人工智能方兴未艾, 这些都是推动量子机器学习发展的内在动力。综上所述, 未来量子机器学习的研究令人振奋, 同时也充满挑战。

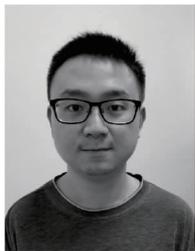
参 考 文 献

- [1] Kohavi R, Foster P. Special issue on applications of machine learning and the knowledge discovery process. *Journal of Machine Learning*, 1998, 30(2): 271-274
- [2] Hilbert M, López P. The world's technological capacity to store, communicate, and compute information. *Science*, 2011, 332(6025): 60-65
- [3] Feynman R P. Simulating physics with computers. *International Journal of Theoretical Physics*, 1982, 21(6): 467-488
- [4] Shor P W. Algorithms for quantum computation: Discrete logarithms and factoring//*Proceedings of the 35th Annual Symposium on Foundations of Computer Science*. Santa Fe, USA, 1994: 124-134
- [5] Grover L K. A fast quantum mechanical algorithm for database search//*Proceedings of the 28th Annual ACM Symposium on Theory of Computing*. Philadelphia, USA, 1996: 212-219
- [6] Wittek P. *Quantum Machine Learning: What Quantum Computing Means to Data Mining*. Cambridge, USA: Academic Press, 2014
- [7] Adcock J, Allen E, Day M, et al. Advances in quantum machine learning. arXiv:1512.02900, 2015
- [8] Schuld M, Sinayskiy I, Petruccione F. An introduction to quantum machine learning. *Contemporary Physics*, 2015, 56(2): 172-185
- [9] Shor P W. Why haven't more quantum algorithms been found?. *Journal of the ACM*, 2003, 50(1): 87-90
- [10] Kak S. On quantum neural computing. *Information Sciences*, 1995, 83(3-4): 143-160
- [11] Behrman E C, Niemel J, Steck J E, et al. A quantum dot neural network//*Proceedings of the 4th Workshop on Physics of Computation*. Portland, USA, 1996: 22-24
- [12] Toth G, Lent C S, Tougaw P D, et al. Quantum cellular neural networks. *Superlattices and Microstructures*, 1996, 20(4): 473-478
- [13] Ventura D, Martinez T. An artificial neuron with quantum mechanical properties//*Proceedings of the International Conference on Artificial Neural Nets and Genetic Algorithms*. Norwich, UK, 1998: 482-485
- [14] Matsui N, Takai M, Nishimura H. A network model based on qubitlike neuron corresponding to quantum circuit. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 2000, 83(10): 67-73
- [15] Zhou R, Qin L, Jiang N. Quantum perceptron network//*Proceedings of the 16th International Conference on Artificial Neural Networks-Volume Part I*. Athens, Greece, 2006: 651-657
- [16] Schuld M, Sinayskiy I, Petruccione F. Quantum walks on graphs representing the firing patterns of a quantum neural network. *Physical Review A*, 2014, 89(3): 032333
- [17] Horn D, Gottlieb A. Algorithm for data clustering in pattern recognition problems based on quantum mechanics. *Physical Review Letters*, 2001, 88(1): 018702
- [18] Lloyd S, Mohseni M, Rebentrost P. Quantum algorithms for supervised and unsupervised machine learning. arXiv:1307.0411, 2013
- [19] Aïmeur E, Brassard G, Gambs S. Quantum clustering algorithms//*Proceedings of the 24th International Conference on Machine Learning*. Corvallis, USA, 2007: 1-8
- [20] Aïmeur E, Brassard G, Gambs S. Quantum speed-up for unsupervised learning. *Machine Learning*, 2013, 90(2): 261-287
- [21] Wiebe N, Kapoor A, Svore K M. Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning. *Quantum Information & Computation*, 2015, 15(3-4): 316-356
- [22] Rebentrost P, Mohseni M, Lloyd S. Quantum support vector machine for big data classification. *Physical Review Letters*, 2014, 113(13): 130503
- [23] Wiebe N, Kapoor A, Svore K M. Quantum deep learning. arXiv:1412.3489, 2014
- [24] Giovannetti V, Lloyd S, Maccone L. Quantum random access memory. *Physical Review Letters*, 2008, 100(16): 160501
- [25] Giovannetti V, Lloyd S, Maccone L. Architectures for a quantum random access memory. *Physical Review A*, 2008, 78(5): 052310

- [26] Lloyd S, Mohseni M, Rebentrost P. Quantum principal component analysis. *Nature Physics*, 2014, 10(9): 631-633
- [27] Harrow A W, Hassidim A, Lloyd S. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 2009, 103(15): 150502
- [28] Childs A M, Kothari R, Somma R D. Quantum linear systems algorithm with exponentially improved dependence on precision. arXiv:1511.02306, 2015
- [29] Servedio R A, Gortler S J. Equivalences and separations between quantum and classical learnability. *SIAM Journal on Computing*, 2004, 33(5): 1067-1092
- [30] Aïmeur E, Brassard G, Gambs S. Machine learning in a quantum world//Proceedings of the 19th International Conference on Advances in Artificial Intelligence: Canadian Society for Computational Studies of Intelligence. Quebec City, Canada, 2006: 431-442
- [31] Yoo S, Bang J, Lee C, et al. A quantum speedup in machine learning: finding an N-bit Boolean function for a classification. *New Journal of Physics*, 2014, 16(10): 103014
- [32] Li Z, Liu X, Xu N, et al. Experimental realization of a quantum support vector machine. *Physical Review Letters*, 2015, 114(14): 140504
- [33] Wang Shu-Hao, Long Gui-Lu. Big data and quantum computation. *Chinese Science Bulletin*, 2015, 60(5-6): 499-508(in Chinese)
(王书浩, 龙桂鲁. 大数据与量子计算. *科学通报*, 2015, 60(5-6): 499-508)
- [34] Pudenz K L, Lidar D A. Quantum adiabatic machine learning. *Quantum Information Processing*, 2013, 12(5): 2027-2070
- [35] Neven H, Rose G, Macready W G. Image recognition with an adiabatic quantum computer I. Mapping to quadratic unconstrained binary optimization. arXiv:0804.4457, 2008
- [36] Neven H, Denchev V S, Rose G, et al. Training a binary classifier with the quantum adiabatic algorithm. arXiv:0811.0416, 2008
- [37] Neven H, Denchev V S, Rose G, et al. Training a large scale classifier with the quantum adiabatic algorithm. arXiv:0912.0779, 2009
- [38] Neven H, Denchev V S, Drew-Brook M, et al. NIPS 2009 demonstration: Binary classification using hardware implementation of quantum annealing//Proceedings of the Neural Information Processing Systems 2009. Vancouver, Canada, 2009: 1-17
- [39] Neven H, Denchev V S, Rose G, et al. QBoost: Large scale classifier training with adiabatic quantum optimization//Proceedings of the 4th Asian Conference on Machine Learning. Singapore, 2012: 333-348
- [40] Cai X D, Wu D, Su Z E, et al. Entanglement-based machine learning on a quantum computer. *Physical Review Letters*, 2015, 114(11): 110504
- [41] Landauer R. The physical nature of information. *Physics Letters A*, 1996, 217(4/5): 188-193
- [42] Moore G E. Cramping more components onto integrated circuits. *Electronics*, 1965, 38(8): 114-117
- [43] Zhou Zheng-Wei, Tu Tao, Gong Ming, et al. Advances and prospects in research on quantum computation. *Progress in Physics*, 2009, 29(2): 127-165(in Chinese)
(周正威, 涂涛, 龚明等. 量子计算的进展和展望. *物理学进展*, 2009, 29(2): 127-165)
- [44] Benioff P. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *Journal of Statistical Physics*, 1980, 22(5): 563-591
- [45] Bammann K. Statistical models: Theory and practice. *Biometrics*, 2006, 62(3): 943-943
- [46] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*, 1995, 20(3): 273-297
- [47] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases//Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. Washington, USA, 1993: 207-216
- [48] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313(5786): 504-507
- [49] Sun Zhi-Jun, Xue Lei, Xu Yang-Ming, et al. Overview of deep learning. *Application Research of Computers*, 2012, 29(8): 2806-2810(in Chinese)
(孙志军, 薛磊, 许阳明等. 深度学习研究综述. *计算机应用研究*, 2012, 29(8): 2806-2810)
- [50] Bengio Y. Learning deep architectures for AI. *Machine Learning*, 2009, 2(1): 1-127
- [51] Larochelle H, Erhan D, Courville A, et al. An empirical evaluation of deep architectures on problems with many factors of variation//Proceedings of the 24th International Conference on Machine Learning. Corvallis, USA, 2007: 473-480
- [52] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, Second Edition, 1995, 3361(10): 1995
- [53] Liu Jian-Wei, Liu Yuan, Luo Xiong-Lin. Research and development on deep learning. *Application Research of Computers*, 2014, 31(7): 1921-1930(in Chinese)
(刘建伟, 刘媛, 罗雄麟. 深度学习研究进展. *计算机应用研究*, 2014, 31(7): 1921-1930)
- [54] Aaronson S. The learnability of quantum states. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 2007, 463(2088): 3089-3114
- [55] Cheng H C, Hsieh M H, Yeh P C. The Learnability of Unknown Quantum Measurements. arXiv: 1501.00559, 2015
- [56] Servedio R A, Gortler S J. Quantum versus classical learnability //Proceedings of the 16th Annual IEEE Conference on Computational Complexity. Chicago, USA, 2001: 138-148
- [57] Servedio R A. Separating quantum and classical learning//Proceedings of the 28th International Colloquium on Automata, Languages and Programming. Crete, Greece, 2001: 1065-1080
- [58] Atici A, Servedio R A. Improved bounds on quantum learning algorithms. *Quantum Information Processing*, 2005, 4(5): 355-386

- [59] Oliveira W, Silva A J, Ludermir T B, et al. Quantum logical neural networks//Proceedings of the 10th Brazilian Symposium on Neural Networks. Salvador, Brazil, 2008: 147-152
- [60] Silva A, de Oliveira W, Ludermir T. A weightless neural node based on a probabilistic quantum memory//Proceedings of the 11th Brazilian Symposium on Neural Networks. Sao Paulo, Brazil, 2010: 259-264
- [61] da Silva A J, Ludermir T B, de Oliveira W R. On the universality of quantum logical neural networks//Proceedings of the 2012 Brazilian Symposium on Neural Networks. Curitiba, Brazil, 2012: 102-106
- [62] Da Silva A J, De Oliveira W R, Ludermir T B. Classical and superposed learning for quantum weightless neural networks. *Neurocomputing*, 2012, 75(1): 52-60
- [63] Tucci R R. Quantum bayesian nets. *International Journal of Modern Physics B*, 1995, 9(03): 295-337
- [64] Monras A, Beige A, Wiesner K. Hidden quantum Markov models and non-adaptive read-out of many-body states. arXiv:1002.2337, 2010
- [65] Clark L A, Huang W, Barlow T M, et al. Hidden quantum Markov models and open quantum systems with instantaneous feedback//Proceedings of the Interdisciplinary Symposium on Complex Systems 2014. Florence, Italy, 2015: 143-151
- [66] Guță M, Kotłowski W. Quantum learning: Asymptotically optimal classification of qubit states. *New Journal of Physics*, 2010, 12(12): 123032
- [67] Bisio A, Chiribella G, D'Ariano G M, et al. Optimal quantum learning of a unitary transformation. *Physical Review A*, 2010, 81(3): 032324
- [68] Gross D, Liu Y K, Flamia S T, et al. Quantum state tomography via compressed sensing. *Physical Review Letters*, 2010, 105(15): 150401
- [69] Flamia S T, Gross D, Liu Y K, et al. Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 2012, 14(9): 095022
- [70] Nielsen M A, Chuang I L. *Quantum Computation and Quantum Information*. Cambridge, UK: Cambridge University Press, 2000
- [71] Long G L. Duality quantum computing and duality quantum information processing. *International Journal of Theoretical Physics*, 2011, 50(4): 1305-1318
- [72] MacQueen J. Some methods for classification and analysis of multivariate observations//Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, USA, 1967: 281-297
- [73] Buhrman H, Cleve R, Watrous J, et al. Quantum fingerprinting. *Physical Review Letters*, 2001, 87(16): 167902
- [74] Arthur D, Vassilvitskii S. *k-means++*: The advantages of careful seeding//Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms. New Orleans, USA, 2007: 1027-1035
- [75] Murtagh F. A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 1983, 26(4): 354-359
- [76] Durr C, Hoyer P. A quantum algorithm for finding the minimum. arXiv quant-ph/9607014, 1996
- [77] Bishop C M, Nasrabadi N M. *Pattern recognition and machine learning*. *Journal of Electronic Imaging*, 2007, 16(4): 9901
- [78] Anguita D, Ridella S, Rivieccio F, et al. Quantum optimization for training support vector machines. *Neural Networks*, 2003, 16(5): 763-770
- [79] Li Hang. *Method of Statistical Learning*. Beijing: Tsinghua University Press, 2012(in Chinese)
(李航. 统计学习方法. 北京: 清华大学出版社, 2012)
- [80] Lu S, Braunstein S L. Quantum decision tree classifier. *Quantum Information Processing*, 2014, 13(3): 757-770
- [81] Farhi E, Gutmann S. Quantum computation and decision trees. *Physical Review A*, 1998, 58(2): 915
- [82] Zeidenberg M. *Neural Networks in Artificial Intelligence*. Upper Saddle River, USA: Ellis Horwood, 1990
- [83] Perus M. Neuro-quantum parallelism in brain-mind and computers. *Informatica (Ljubljana)*, 1996, 20(2): 173-184
- [84] Menneer T, Narayanan A. Quantum-inspired neural networks. Department of Computer Science, University of Exeter, Exeter, UK: Technical Report 329, 1995
- [85] Behrman E C, Chandrashekar V, Wang Z, et al. A quantum neural network computes entanglement. arXiv quant-ph/0202131, 2002
- [86] Behrman E C, Steck J E, Skinner S R. A spatial quantum neural computer//Proceedings of the 1999 International Joint Conference on Neural Networks. Washington, USA, 1999: 874-877
- [87] Behrman E C, Nash L R, Steck J E, et al. Simulations of quantum neural networks. *Information Sciences*, 2000, 128(3): 257-269
- [88] Behrman E C, Steck J E. A quantum neural network computes its own relative phase//Proceedings of the 2013 IEEE Symposium on Swarm Intelligence. Singapore, 2013: 119-124
- [89] Ventura D, Martinez T. Quantum associative memory with exponential capacity//Proceedings of the 1998 IEEE World Congress on Computational Intelligence. Anchorage, USA, 1998: 509-513
- [90] Menneer T S I. *Quantum Artificial Neural Networks*[Ph. D. dissertation]. University of Exeter, Exeter, UK, 1998
- [91] Pribram K H. Quantum holography: Is it relevant to brain function? *Information Sciences*, 1999, 115(1-4): 97-102
- [92] Zhou J, Gan Q, Krzyzak A, et al. Recognition of handwritten numerals by quantum neural network with fuzzy features. *International Journal on Document Analysis and Recognition*, 1999, 2(1): 30-36
- [93] Narayanan A, Menneer T. Quantum artificial neural network architectures and components. *Information Sciences*, 2000, 128(3): 231-255
- [94] Ezhov A A, Ventura D. Quantum neural networks. *Future Directions for Intelligent Systems and Information Sciences*, 2000, 45: 213-235

- [95] Gupta S, Zia R K P. Quantum neural networks. *Journal of Computer and System Sciences*, 2001, 63(3): 355-383
- [96] Altaisky M V. Quantum neural network. arXiv quant-ph/0107012, 2001
- [97] Kouda N, Matsui N, Nishimura H. Image compression by layered quantum neural networks. *Neural Processing Letters*, 2002, 16(1): 67-80
- [98] Kouda N, Matsui N, Nishimura H, et al. Qubit neural network and its learning efficiency. *Neural Computing & Applications*, 2005, 14(2): 114-121
- [99] Li Pan-Chi, Li Shi-Yong. A quantum self-organization feature mapping networks and clustering algorithm. *Chinese Journal of Quantum Electronics*, 2007, 24(4): 463-468(in Chinese)
(李盼池, 李士勇. 一种量子自组织特征映射网络模型及聚类算法. *量子电子学报*, 2007, 24(4): 463-468)
- [100] da Silva A J, de Oliveira W R, Ludermit T B. Training a classical weightless neural network in a quantum computer// *Proceedings of the 2014 European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Bruges, Belgium, 2014: 523-528
- [101] da Silva A J, de Oliveira W R, Ludermit T B. Weightless neural network parameters and architecture selection in a quantum computer. *Neurocomputing*, 2016, 183: 13-22
- [102] de Paula Neto F M, da Silva A J, Ludermit T B, et al. Analysis of quantum neural models// *Proceedings of the 11th Brazilian Congress on Computational Intelligence*. Ipojuca, Brazil, 2013
- [103] Schuld M, Sinayskiy I, Petruccione F. The quest for a quantum neural network. *Quantum Information Processing*, 2014, 13(11): 2567-2586
- [104] Maeda M, Suenaga M, Miyajima H. A learning model in qubit neuron according to quantum circuit// *Proceedings of the 1st International Conference on Advances in Natural Computation-Volume Part I*. Changsha, China, 2005: 283-292
- [105] Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 1987, 2(1-3): 37-52
- [106] Izenman A J. *Linear Discriminant Analysis, Modern Multivariate Statistical Techniques*. New York, USA: Springer, 2013: 237-280
- [107] Fisher R A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 1936, 7(2): 179-188
- [108] Cong I, Duan L. Quantum discriminant analysis for dimensionality reduction and classification. *New Journal of Physics*, 2016, 18(7): 073011
- [109] Hinton G E, Sejnowski T J. Learning and relearning in Boltzmann machines// *Rumelhart D E, McClelland J L eds. Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, USA: MIT Press, 1986: 282-317
- [110] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. *Neural computation*, 2006, 18(7): 1527-1554
- [111] Brassard G, Hoyer P, Mosca M, et al. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 2002, 305: 53-74
- [112] Wiebe N, Kapoor A, Granade C, et al. Quantum inspired training for Boltzmann machines. arXiv:1507.02642, 2015
- [113] Hong F Y, Xiang Y, Zhu Z Y, et al. Robust quantum random access memory. *Physical Review A*, 2012, 86(1): 010306
- [114] Arunachalam S, Gheorghiu V, Jochym-O'Connor T, et al. On the robustness of bucket brigade quantum RAM. *New Journal of Physics*, 2015, 17(12): 123010



HUANG Yi-Ming, born in 1991, Ph.D. candidate. His research interests include data analysis, data mining and quantum machine learning.

LEI Hang, born in 1960, Ph.D., professor, Ph.D. supervisor. His research interests include big data mining and embedded system.

LI Xiao-Yu, born in 1984, Ph.D., associated professor. Her research interests include data analysis, quantum computation and quantum machine learning.

Background

Quantum machine learning is an emerging field between machine learning and quantum computation. In the past tens of years, researchers try to investigate whether quantum computation can do some help for improving classical machine learning algorithms. Their works include (1) Take advantage of the high parallel of quantum computation to speed up some conventional machine learning algorithms. (2) Use machine

learning algorithms to overcome fundamental quantum problems. (3) Propose a new idea of quantum machine learning which is derived from principle of quantum mechanics. Before the universal quantum computer comes into being, it is hard to highlight the enormous power of quantum algorithms. In addition, there are still some problems and challenges that need to be solved.