物理学报Acta Physica Sinica





Institute of Physics, CAS

机器学习与物理专题编者按

Preface to the special topic: Machine learning and physics

引用信息 Citation: Acta Physica Sinica, 70, 140101 (2021) DOI: 10.7498/aps.70.140101 在线阅读 View online: https://doi.org/10.7498/aps.70.140101 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

电介质材料和物理专题编者按 Preface to the special topic: Dielectric materials and physics 物理学报. 2020, 69(12): 120101 https://doi.org/10.7498/aps.69.120101

柔性电子专题编者按

Preface to the special topic: Flexible electronics 物理学报. 2020, 69(17): 170101 https://doi.org/10.7498/aps.69.170101

光学超构材料专题编者按 Preface to the special topic: Optical metamaterials 物理学报. 2020, 69(15): 150101 https://doi.org/10.7498/aps.69.150101

太赫兹自旋光电子专题编者按 Preface to the special topic: Terahertz spintronic optoelectronics 物理学报. 2020, 69(20): 200101 https://doi.org/10.7498/aps.69.200101

低维材料非线性光学与器件专题编者按 Preface to the special topic: Nonlinear optics and devices of low-dimensional materials 物理学报. 2020, 69(18): 180101 https://doi.org/10.7498/aps.69.180101

探索凝聚态中的马约拉纳粒子专题编者按 Preface to the special topic: Majorana in condensed matter 物理学报. 2020, 69(11): 110101 https://doi.org/10.7498/aps.69.110101 专题: 机器学习与物理

机器学习与物理专题编者按

DOI: 10.7498/aps.70.140101

机器学习,尤其是深度学习,在很多方面取得了令人瞩目的成就,是当前科学技术领域最为热门、发展最快的方向之一.其与物理的结合是最近几年新兴的交叉前沿领域,受到了广泛关注.一方面,运用机器学习的方法可以解决一些复杂的、传统方法很难或无法解决的物理问题;另一方面,物理中的一些概念、理论和方法也可以用于研究机器学习.二者的交叉融通带来了新的机遇与挑战,将极大地促进两个领域的发展.

本专题邀请了若干活跃在该新兴领域的专家撰稿,重点介绍机器学习与物理交叉方向的部分国际前沿课题和最新研究进展.内容涵盖了量子人工智能中的对抗学习,量子生成模型,基于波动与扩散的机器学习,自动微分,绝热量子算法设计,量子机器学习中的编码与初态制备,以及基于自旋体系的量子机器学习实验进展等.

希望本专题能够帮助读者了解机器学习与物理交叉方向的研究内容,基本思想与方法,最新进展情况,以及面临的挑战与机遇.同时,也希望这个专题能够激发读者的兴趣,吸引更多的研究人员加入到此交叉领域的研究中.

(客座编辑:邓东灵 清华大学)

SPECIAL TOPIC—Machine learning and physics

Preface to the special topic: Machine learning and physics

DOI: 10.7498/aps.70.140101

Machine learning, especially deep learning, has achieved remarkable success in a wide range of applications. It is one of today's most rapidly growing fields in science and technology. In recent years, the interplay between machine learning and physics has attracted tremendous attention, giving rise to a new interdisciplinary research frontier. On the one hand, we may utilize machine learning methods to tackle certain intricate physical problems that are beyond the capability of traditional approaches. On the other hand, certain concepts, ideas, and methods originated in physics can also be exploited to enhance the study of machine learning. Without a doubt, the fusion of machine learning and physics will bring us new opportunities and challenges, and significantly advance the studies in both fields.

This special topic contains several review papers written by experts working actively in this emergent interdisciplinary field. These papers review a number of hot topics and some latest progresses, covering adversarial learning in quantum artificial intelligence, quantum generative models, machine learning based on waves and diffusions, automatic differentiation, machine learning assisted quantum adiabatic algorithm design, state preparation in quantum machine learning, experimental progress of quantum machine learning based on spin systems, etc.

We hope this special topic can help readers gain a primary picture of the research content, basic ideas and methods, the latest developments, and the challenges and opportunities faced in the intersection of machine learning and physics. Meanwhile, we also hope this special topic can provide some inspiration to readers, and attract more researchers to join this exciting interdisciplinary field.

> Deng Dong-Ling Tsinghua University, China

物理学报Acta Physica Sinica





Institute of Physics, CAS

量子人工智能中的对抗学习

沈培鑫 蒋文杰 李炜康 鲁智德 邓东灵

Adversarial learning in quantum artificial intelligence Shen Pei-Xin Jiang Wen-Jie Li Wei-Kang Lu Zhi-De Deng Dong-Ling 引用信息 Citation: Acta Physica Sinica, 70, 140302 (2021) DOI: 10.7498/aps.70.20210789 在线阅读 View online: https://doi.org/10.7498/aps.70.20210789 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

利用超导量子电路模拟拓扑量子材料

Topological quantum material simulated with superconducting quantum circuits 物理学报. 2018, 67(22): 220302 https://doi.org/10.7498/aps.67.20181857

基于人工神经网络在线学习方法优化磁屏蔽特性参数

Online learning method based on artificial neural network to optimize magnetic shielding characteristic parameters 物理学报. 2019, 68(13): 130701 https://doi.org/10.7498/aps.68.20190234

人工带隙材料的拓扑性质

Topological properties of artificial bandgap materials 物理学报. 2017, 66(22): 224203 https://doi.org/10.7498/aps.66.224203

HgTe/CdTe量子阱中自旋拓扑态的退相干效应

Dephasing effect of quantum spin topological states in HgTe/CdTe quantum well 物理学报. 2019, 68(22): 227301 https://doi.org/10.7498/aps.68.20191072

一维扩展量子罗盘模型的拓扑序和量子相变

Topological orders and quantum phase transitions in a one-dimensional extended quantum compass model 物理学报. 2018, 67(19): 190301 https://doi.org/10.7498/aps.67.20180855

马约拉纳零能模的非阿贝尔统计及其在拓扑量子计算的应用

Non-abelian statistics of Majorana modes and the applications to topological quantum computation 物理学报. 2020, 69(11): 110302 https://doi.org/10.7498/aps.69.20200812 专题: 机器学习与物理

量子人工智能中的对抗学习*

沈培鑫1) 蒋文杰1) 李炜康1) 鲁智德1) 邓东灵1)2)†

1) (清华大学交叉信息研究院,北京 100084)

2) (上海期智研究院,上海 200232)

(2021年4月25日收到; 2021年5月26日收到修改稿)

量子人工智能是一个探究人工智能与量子物理交叉的领域:一方面人工智能的方法和技术可以用来解 决量子科学中的问题;另一方面,量子计算的发展也可能为人工智能,尤其是机器学习,提供新的范式,极大 促进人工智能的发展.然而,量子机器学习和经典学习系统对于对抗样本同样具有脆弱性:在原始数据样本 上添加精心制作的微小扰动将很可能导致系统做出错误的预测.本文介绍经典与量子对抗机器学习的基本 概念、原理、以及最新进展.首先从经典和量子两个方面介绍对抗学习,通过二维经典伊辛模型和三维手征 拓扑绝缘体的对抗样本揭示出经典机器学习在识别物质相时的脆弱性,同时利用手写字体的对抗样本直观 展示出量子分类器的脆弱性.随后从理论层面上分别阐述经典与量子的"没有免费午餐"定理,并探讨了量子 分类器的普适对抗样本.最后,分析并讨论了相应的防御策略.量子人工智能中对抗学习的研究揭示了量子 智能系统潜在的风险以及可能的防御策略,将对未来量子技术与人工智能的交叉产生深刻影响.

关键词: 量子人工智能,量子对抗学习,量子分类器,拓扑物质相 **PACS:** 03.67.-a, 03.67.Ac, 05.30.Rt, 07.05.Mh **DOI:** 10.7498/aps.70.20210789

1 引 言

在过去的十年里,无论是从图像识别^[1]到自然 语言处理^[2],还是从医学诊断^[3]到自动驾驶^[4],人 工智能领域都取得了巨大的成功,引发了现代社会 诸多领域的技术革命^[5,6].特别地,基于人工神经网 络的 AlphaGo^[7,8] 与 AlphaFold^[9]分别在围棋和预 测蛋白质结构方面取得了里程碑式的突破.人工智 能主要有三条发展路线:符号主义、连接主义与行 为主义^[10].基于人工神经网络的机器学习属于连 接主义,它是实现人工智能的一个重要途径,近年 来发展非常迅速^[11].与此同时,近期实验上展现的 量子优越性也标志着量子计算领域^[12]取得了开拓 性的进展^[13,14].这两个快速发展的领域的交叉融通 催生了一个新的研究前沿——量子人工智能^[15-17]. 一方面,可以利用人工智能技术来解决量子科学中的难题,例如量子多体问题^[18]、量子态层析^[19]、拓 扑量子编译^[20]、无序材料中的结构转变^[21]、量子非 定域性探测^[22]以及物质相分类^[23-33]等.另一方面, 直接运行在量子计算机上的量子算法具有巨大的 潜力,可以增强、加速甚至革新^[34-41]某些经典人工 智能算法.其中具有代表性的算法包括 Harrow-Hassidim-Lloyd算法^[34]、量子主成分分析^[35]、量子 生成模型^[38-40]和量子支持向量机^[42]等.毫无疑问, 人工智能和量子物理之间的相互融通将极大促进 两个领域的发展.当前,量子物理与人工智能的交 叉研究主要集中在与机器学习的交叉方面,与人工 智能另外两条路线(即符号主义与行为主义)之间 的交叉研究还相对缺乏.

在经典人工智能领域,近期有许多工作揭示出 基于深度人工神经网络的分类器在对抗场景中的

^{*} 国家自然科学基金(批准号: 12075128)、清华大学启动经费(批准号: 53330300320)和上海期智研究院资助的课题.

[†] 通信作者. E-mail: dldeng@mail.tsinghua.edu.cn

^{© 2021} 中国物理学会 Chinese Physical Society

脆弱性[43-45]: 向原始数据添加精心制作的微量 (甚 至是人眼无法察觉的)噪声,可能会导致分类器以 非常高的置信度做出错误的预测. Szegedy 等^[46,47] 利用一个著名的例子直观地展示了深度学习的脆 弱性: 在初始的熊猫图像添加了肉眼不可察觉的对 抗噪声后,分类器将其错误地标识为长臂猿,且置 信度大于 99% (图 1). 这些精心设计用来欺骗分类 器的输入样本被称为对抗样本.目前,人们普遍认 为对抗样本在经典机器学习中广泛存在——无论 输入数据类型和神经网络的细节如何,几乎所有学 习模型都易遭受对抗攻击[43-45].从理论计算机科 学的角度来看, 经典分类器关于对抗微扰的脆弱性 与"没有免费午餐"定理之间存在深刻的联系. 需要 指出的是,对抗学习中的对抗样本与生成对抗网 络 (generative adversarial networks, GAN)^[48]产 生的样本存在本质区别. 前者目的在于攻击已经训 练好的学习模型,后者目的在于模拟目标数据集以 生成新的数据样本.



图 1 量子与经典对抗学习示意图 输入的原始熊猫图 像样本可以编码为经典或量子数据,分类器 (包含变分量 子线路或人工神经网络)能够以非常高的准确率识别出熊 猫;但添加少量精心制作的噪声后,同一分类器将以非常 高的置信度把轻微修改过的熊猫图像错误分类为长臂猿 Fig. 1. A schematic illustration of quantum and classical adversarial learning. The image of a panda can be encoded as classical or quantum data. A classifier, which uses either variational quantum circuits or classical artificial neural networks, can successfully identify the image as a panda with the state-of-the-art accuracy. However, adding a small amount of carefully crafted noise will cause the same classifier to misclassify the slightly modified image into a gibbon with a notably high confidence.

随着嘈杂中型量子 (noisy intermediate-scale quantum, NISQ) 时代的到来,量子计算有望在近期的科研应用领域中大放异彩^[49]. 然而,尽管量子

计算机能够在某些方面展示出超越经典计算机的 性能,但在对抗学习中也会展现出脆弱性,事实上, 量子分类器的脆弱性最近开始受到广泛关注[50-52]. 也因此衍生了一个全新研究方向——量子对抗机 器学习. 文献 [53] 在理论上表明, 添加一个随着量 子分类器比特数目指数减小的对抗微扰 (adversarial perturbation) 就足以影响学习模型的输出. 这表明量子分类器的稳健性 (robustness) 与高维 希尔伯特 (Hilbert) 空间潜在的量子优势之间存在 竞争关系: 展示量子优势需要高维的希尔伯特空 间,但量子分类器的稳健性随空间维度的增加而减 弱. 最近文献 [50] 在量子机器学习的框架下探究了 不同的对抗场景,其诸多数值实例表明量子分类器 同样很容易受到精心制作的对抗样本的攻击.目前 这个新兴的研究方向正在迅速增长,吸引了越来越 多来自不同领域研究者的关注,但它仍处于起步阶 段,许多重要问题仍待探讨.

本文首先介绍近期经典对抗机器学习的研究 进展和技术方法,探讨其在识别物质相任务中得到 的对抗样本实例.随后介绍经典机器学习中的"没 有免费午餐"定理及其在量子情况下的推广,还回 顾了对量子分类器脆弱性的相关研究.有工作表 明,研究者可以生成一个通用的对抗样本来迷惑一 组不同量子分类器,同时也能够利用一个普适的对 抗微扰将不同的初始样本全部变成对抗样本^[54]. 基于这些研究结果,人们可以针对性地开发出实用 的防御策略以对抗相应的攻击.通过对抗训练,可 以使得分类器针对特定类型的对抗扰动的稳健性 有显著提高.最后,对量子人工智能中的对抗学习 做出总结与展望,希望能给 NISQ 时代的量子机器 学习提供有价值的指导.

2 经典对抗机器学习

经典对抗机器学习的早期探索可以追溯到垃 圾邮件过滤 (spam filtering) 问题,即垃圾邮件的 发送方与抵制方的博弈.一般来说,用户的邮箱地 址为外界得知后,一些恶意的群体便可能为了商业 利益等向这个邮箱发送广告邮件甚至电脑病毒.为 了抵御这种侵犯行为,人们开发了邮件过滤器以区 分正常邮件与恶意邮件并对后者加以拦截.这些早 期的邮件过滤器可以看作是线性的分类器,通过对 邮件中的词汇与已采集到的恶意邮件的特征词汇 相对比来做出分类的判断. 而作为恶意邮件的发送 方, 为了躲过邮件过滤器的检测, 便会采取一系列 的手段, 如修改恶意邮件的特征词汇、增加正常词 汇的比例等. 以上便是对抗机器学习中防御与攻击 的例子. 需要说明的是, 在实际的对抗机器学习的 运行中, 防御与攻击的发展往往是一个迭代的过 程, 即攻击方与防御方都会根据对方技术的演变而 做出相应的调整与改进, 以提高己方成功的概率.

在上述邮件过滤的对抗学习问题中, Dalvi 等^[55] 以及 Lowd 和 Meek^[56] 在此背景下研究了线 性分类器对于对抗样本的脆弱性.他们发现,在不 影响邮件包含的目的信息下只对邮件内容做小幅 度精心设计的修改,就可以顺利地通过过滤器的筛 选.Barreno等^[57] 在 2006 年首先针对对抗机器学 习做出了广泛的讨论,并给出了不同的攻击分类. 此后对抗机器学习得到了全方位的发展,研究内容 包括攻击策略与攻击背景、对抗攻击的防御手段、 以及机器学习的安全性评估等^[44,45,58,59].以下将对 对抗机器学习做一个宏观的介绍,以及从不同角度 对对抗机器学习做出分类.

在对抗过程中, 攻击者能够获得的攻击目标 的模型信息(如训练数据、模型结构等)在不同 的情形下有不同的等级. 根据对目标的信息由完全 了解到完全不了解,可以将攻击划分为白盒攻击 (white-box attack)、灰盒攻击 (gray-box attack) 和黑盒攻击 (black-box attack). 对于攻击者的能 力来说,我们根据其可以同时操纵训练集和测试集 或只能操纵测试集将其分为毒药攻击 (poisoning attack) 和躲避攻击 (evasion attack). 前文提到的 绕过垃圾邮件过滤器便是躲避攻击的一个例子,即 通过技术手段修改测试集并让所攻击的模型给出 攻击者所期望的结果. 而通过与用户进行交互对话 来改善表现的语音模型则属于毒药攻击中的一类, 比如某些恶意用户带有歧视性的言论将会使这个 模型有潜在的风险.此外,可以将攻击的目标分为 无差别攻击 (untargeted attack) 和针对性攻击 (targeted attack). 顾名思义, 无差别攻击是指攻击 的目的为使分类器分类错误,而不关注将样本判错 成哪个错误标签. 针对性攻击则希望样本被错判成 某个特定标签.

对于防御者而言,一个直接的防御方式是根据 攻击方已有的攻击手段来重新设计和训练机器学 习模型,从而使其具有抵御这一类攻击的能力.对 于躲避攻击中提到的在某一类数据微扰得到的 对抗样本,可以将其连同原始的标签加入到训练 集进行训练,以此来增加模型面对攻击时的抵抗 能力——这便是所谓的对抗训练 (adversarial training)^[60].除此之外,一个更加理论化的策略是 稳健优化 (robust optimization)^[61].这个方法的思 想在于把防御的过程看成一个极小化极大问题 (minimax problem),即通过扰动训练集以最大化 损失函数,并通过训练模型来最小化损失函数,以 此让重新训练的模型获得较好的防御能力 (详见 第5节的讨论).

近年来深度学习得到了长足的发展,其在人脸 识别、自动驾驶等领域的应用受到了广泛的关 注[1,4]. 然而, 研究者们发现深度学习同样也存在着 被对抗样本攻击的威胁[46,48]. 在一个已经训练好的 可以正确识别熊猫的深度学习模型中,即使添加一 个肉眼难以察觉的扰动,也很可能会使这个模型给 出的预测结果变为长臂猿 (图 1). 如果这类攻击没 有得到解决而被恶意者利用,将会使深度学习的实 际应用中存在严重的安全隐患. 例如在自动驾驶汽 车上,如果前方一个停车告示牌被贴上一层精心设 计的扰动薄膜,被汽车的识别程序识别为常速行 驶,便可能引发安全事故[44].因此深度学习中的对 抗攻击和防御策略也受到了广泛的关注.目前已经 被提出的较为著名的攻击算法有快速梯度符号 法[62]、基本迭代法[63]、动量迭代法[64]、投影梯度下 降法 [62] 等, 这些算法通过精心设计的扰动来使模 型无法给出正确的预测.为了防御这些攻击,同样 有很多的策略被提出. 除了上文提到的稳健优化方 法,在深度学习中,随机化和去噪往往也有着不错 的效果. 这是由于深度学习有着较强的表达能力 以及对随机噪声、去噪操作有着较好的稳健性,而 这些操作能够有效地消除对抗样本的扰动. 在对抗 攻击与防御策略两者的快速发展和博弈中,这些成 果必将为机器学习的安全性和稳定性提供有益的 指导.

3 机器学习物质相中的对抗样本

在凝聚态物理中,识别不同的物质相并分辨出 他们之间的相变点是一个重要且有趣的问题.近年 来,机器学习在处理这类问题上取得了一系列令人 瞩目的进展^[24-33,65-70],其中包括监督学习和无监督 学习等方法.与此同时,一个亟待回答的问题是, 这些通过机器学习方法所获得的结论是否可靠? 它们能否抵御得住来自对抗样本的攻击?针对这 个问题,文献 [71] 进行了一系列的探索,其研究的 内容包含经典二维伊辛 (Ising) 模型和三维手征拓 扑绝缘体不同物质相的分类.

经典二维伊辛模型的哈密顿量可以表示为

$$H_{\rm Ising} = -J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z, \tag{1}$$

其中在 i点的 z方向自旋 $\sigma_i^z = \pm 1$,最近邻自旋的 耦合强度设为单位能量 (J = 1).为了测试机器学 习关于对抗样本的脆弱性,采用一个全连接的前馈 神经网络来训练由蒙特卡罗模拟生成的经典自旋 构型数据,其在测试集的准确率能够达到 97%.在 对抗攻击的阶段,随机选取了一个铁磁相样本,运 用对抗攻击的算法对其迭代地增加扰动,最终该模 型以 88% 的高置信度将这个受到扰动的样本错误 地分类为顺磁相.图2展示了一个更为极端的例 子:尽管修改前的未扰动数据(图 2(a))与对抗样 本(图 2(b))只相差了一个自旋,分类模型却给出 了截然不同的预测结果.

对于拓扑相数据,考虑一个三维手征拓扑绝缘体^[72,73]:

$$\boldsymbol{H}_{\mathrm{TI}} = \sum_{\boldsymbol{k} \in \mathrm{BZ}} \boldsymbol{\Psi}_{\boldsymbol{k}}^{\dagger} \boldsymbol{H}_{\boldsymbol{k}} \boldsymbol{\Psi}_{\boldsymbol{k}}, \qquad (2)$$

其中 $\Psi_{k}^{\dagger} = \left(c_{k,1}^{\dagger}, c_{k,0}^{\dagger}, c_{k,-1}^{\dagger}\right), c_{k,\mu}^{\dagger}$ 表示在动量 $k = (k_{x}, k_{y}, k_{z})$ 处在 $\mu = -1, 0, 1$ 态上的的费米产生算符, 求 和遍历整个布里渊区 (Brillouin zone, BZ). 单体 动量空间的哈密顿量 $H_{k} = \lambda_{4} \sin k_{x} + \lambda_{5} \sin k_{y} + \lambda_{6} \sin k_{z} - \lambda_{7} (\cos k_{x} + \cos k_{y} + \cos k_{z} + h), 其中 \lambda_{4,5,6,7}$ 代表四个无迹的 Gell-Mann 矩阵, 0 < |h| < 1、 1 < |h| < 3、 |h| > 3分别对应着三种不同的拓扑物



图 2 机器学习物质相中的对抗样本 (a) 一个原始的经典二维伊辛模型铁磁相的自旋构型; (b) 被分类器错误识别成顺磁相的 对抗样本,其相对于 (a) 只改变了一个自旋; (c) 一个原始的三维手征拓扑绝缘体的拓扑相样本; (d) 被分类器错误识别成其他相 的对抗样本,其相对于 (c) 只有肉眼难以识别的细微差别

Fig. 2. Adversarial examples in machine learning phases of matter: (a) A legitimate sample of the spin configuration in the ferromagnetic phase of the two-dimensional (2D) classical Ising model; (b) an adversarial example misclassified as the paramagnetic phase, which only differs from the original legitimate one shown in (a) by a single pixel; (c) a legitimate sample of the topological phase of three-dimensional (3D) chiral topological insulators; (d) an adversarial example misclassified as the other phase, which only differs from the original legitimate one shown in (c) by a tiny amount of noises that are imperceptible to human eyes. 质相^[72]. 首先使用一个三维卷积神经网络来进 行初步的训练和分类, 随后采用动量迭代法生成 对抗微扰. 与二维伊辛模型类似, 对抗样本在这 个卷积神经网络模型以极高的概率被误分类成其 他相. 为了在视觉上更加直观, 对处于拓扑相的 图 2(c) 运用快速梯度符号法生成其相应的对抗样 本 (图 2(d)). 尽管二者通过肉眼观察难以分辨, 但神 经网络却会把添加扰动的对抗样本识别成其他相.

通过这些数值实验,可以观察到在识别物质相 及其相变的问题上,机器学习在对抗攻击面前有其 脆弱性,为此需要制定相应的防御策略.特别地, 本文以三维卷积神经网络训练拓扑相数据的例 子来演示对抗训练方法. 在通过快速梯度符号法和 投影梯度下降法等方式获得对抗样本后,将其与原 始数据一起作为训练集重新训练神经网络,所获得 的新模型在对抗样本上也具有很高的识别准确率. 需要指出的是,上述的防御手段具有一定的局限 性: 对抗训练后的新模型一般只能抵御特定的攻击 方法 (训练用的对抗样本由此攻击方法产生), 当面 对黑盒攻击或其他类型的对抗攻击时,需要根据具 体的情形再次重新训练网络. 以上工作揭示了机器 学习应用于物质相识别时潜在的脆弱性,同时对机 器学习在其他物理学问题上的应用也提供了有益 参考.

4 量子对抗机器学习

4.1 "没有免费午餐"定理

基于人工智能的技术早已应用在各个领域,从 日常生活到科学探究,从休闲娱乐到工业生产,逐 渐深入人类社会的各个方面^[10].然而在诸如金融、 国防、医疗等对安全性有着严格要求的领域,对抗 机器学习的发展使得人们难以忽略其背后潜藏的 风险^[74].正如前文提到,在经典人工智能领域中有 很多攻击算法可以用于构造对抗样本,这使得精心 训练的人工智能模型面临重大的安全挑战^[75].从 理论上阐明对抗学习和对抗样本是一件充满挑战 但却意义重大的事情.对这一问题的研究不仅可以 帮助我们明确经典人工智能算法面临的安全性问 题,更可以指引我们理解量子人工智能算法的相关 性质^[15-17].我们已经见证了中等规模量子信息处 理技术的突破:谷歌公司 53 个超导量子比特的"悬 铃木"量子处理器被成功应用于随机线路取样^[13]; 我国 76 个光子的量子计算原型机"九章"在 200 s 内成功获得 5000 万个样本的高斯玻色取样^[14],均 标志着量子技术取得突破性进展.将量子计算技术 应用于人工智能领域是一个前沿交叉领域,近年来 正蓬勃发展^[17].从理论上剖析量子人工智能的原 理不仅可以帮助研究人员挖掘其潜力,更能明确其 限制所在,这是当下此领域最为核心的研究内容之 一.量子对抗学习就是将对抗学习的思想和技术应 用在量子人工智能领域,这可以帮助研究人员更好 地理解其内在的性质及可能存在的限制.

很多人工智能算法的实质是通过有限的数据 拟合未知的函数,以期对没有学习过的数据也能做 出精准的预测[76]. 在实际应用中, 人工智能算法可 以用于处理多种多样的问题,比如在机器翻译问题 中利用机器学习寻找两种语言之间的映射,从而使 得计算机可以自动地将一种语言翻译成为另一种 语言[77]. 未知的函数可能是非常复杂的, 比如上面 提到的机器翻译问题中,语言所包含的信息量非常 庞大,语言之间的映射是非常复杂且一般难以直接 计算的[78],因此需要使用复杂的学习模型从大量 数据中寻找可能的规律. Wolpert 与 Macready^[79] 早在1997年就提出,在没有对未知函数做出限制 的情况下,所有的学习算法都是无法区分优劣的, 这是"没有免费午餐"定理的体现.因此,对于待解 决问题的先验知识决定了对学习算法的选择.另一 方面,在训练学习模型的时候,还需要大量的数据 以提高学习模型预测新数据的准确率.一般情况 下,更大的数据规模意味着更准确的模型,而数据 规模的限制往往会使得学习模型预测准确率降低. 事实上,数据规模与模型准确率的关系也可以通过 "没有免费午餐"定理来定量刻画[80]. 假设未知函 数f是从有限集合 χ 到有限集合 γ 的映射,S是包 含 N个数据的训练集, h_s 是学习算法基于已知的 数据集 S学习得到的模型, $R_{f}(h_{S})$ 为模型预测错 误的概率,则有如下关系:

$$\mathbb{E}_{f}\left[\mathbb{E}_{S}\left[R_{f}(h_{S})\right]\right] \geqslant \left(1 - \frac{1}{|\mathcal{Y}|}\right) \left(1 - \frac{N}{|\mathcal{X}|}\right), \quad (3)$$

其中|X|, |ン|分别代表集合元素的个数.(3)式的直 观理解为:训练数据集包含有关未知函数的知识, 学习算法的核心是从中发现并学习其中的知识,从 而应用到新的数据预测.因此,如果训练数据过于 稀疏,其中包含的信息非常稀少,学习算法一般是 无法学习到足够的知识的.需要指出的是,上述结 论是针对最一般情况的平均结果.具体到某些特殊的问题,基于先验知识仍可能通过较少的数据量做出较为准确的预测.

量子机器学习使用量子数据作为学习算法的 输入与输出.量子数据一般表示为量子比特的物理 状态,因此全部的量子数据构成一个维度指数增长 的希尔伯特空间.量子学习算法的目标是从已知的 数据集中学习从输入空间到输出空间的未知幺正 映射.经典人工智能的学习效果用学习模型预测错 误的概率来表示,在量子机器学习中,可以使用学 习得到的幺正映射与未知的目标映射之间的距离 *R_U(V_S)*来表征学习效果.与经典"没有免费午餐" 定理相对应的量子"没有免费午餐"定理有如下表 述^[81]:

$$\mathbb{E}_{U}\left[\mathbb{E}_{S}\left[R_{U}\left(V_{S}\right)\right]\right] \ge 1 - \frac{1}{d\left(d+1\right)}\left(N^{2} + d + 1\right), \quad (4)$$

其中 U是未知的目标映射, S是包含 N个量子数 据的训练集, Vs是基于数据集 S得到的学习结果, d是输入空间的维度. 与经典结论相比, 可以看到 有限数据集的大小被希尔伯特空间的维度所替代. 在训练集规模与空间指数维度相当的时候, 一般可 以获得比较好的学习效果. 同样需要指出的是, 这 个结论是针对全部可能幺正映射而言, 对于某些特 殊的幺正映射, 通过较少的数据量做出较为准确的 预测是可能实现的.

"没有免费午餐"定理给出了训练集规模与学 习效果之间的关系.在应用学习算法处理实际问题 的时候,由于问题本身的复杂性,往往只能获得一 部分的数据.另一方面,即使获得了大量的数据, 由于现行计算能力的限制,学习模型的参数规模 往往受到一定限制.因此在大部分情况下,只能获 得对于目标问题的近似解,这就意味着预测一般 难以做到完全准确,这是对抗攻击能够成功的重要 原因.

4.2 量子分类器的脆弱性

分类任务是一类常见的人工智能任务^[10],日 常生活的人脸识别^[82]、自动驾驶^[4]中的信号检测、 物理研究中的物质相识别^[24-33,65-70]等问题,其内 核都是分类任务.在量子人工智能领域,分类任务 同样广泛存在^[17].不幸的是,无论是经典分类器还 是量子分类器,在面对精心设计的对抗样本时,其 预测准确率都会显著下降^[50,60].即使对抗样本与原 来的样本相差无几,甚至仅仅改变图片的一个像素 值,都会使得分类器分类错误——这就是分类器的 脆弱性.

在前文中已经看到,在使用学习模型识别物质 相的任务中[71],可以使用多种攻击算法,让原本分 类正确的样本在经过微小的扰动之后被错误识别. 并且随着扰动强度的增加,能够被正确识别的样本 数会迅速下降.实际上,这一现象是广泛存在的, 这是"没有免费午餐"定理与高维数据独特的统计 性质相结合的结果. 学习任务的全部数据的集合构 成一个概率空间,任意数据出现概率即是其在这个 空间中的体积.因此,全部数据组成的集合是以概 率为测度的测度空间.同时还需要衡量数据受扰动 的强度,因此我们引入距离函数来衡量扰动前后数 据之间的差别. 装备了距离函数的高维测度空间有 一个非常重要的现象叫做测度集中^[83].物理上有 许多这样的例子,比如对于平衡状态下的自由粒子 集合. 经典统计力学告诉我们, 这些系统的状态总 可以使用一些简单的宏观量来描述,这些宏观量就 是系统微观状态的平均.进一步的计算可以证明, 随着系统粒子数的增加,平衡状态下系统偏离这些 统计平均值的概率迅速下降,即是在相空间中,系 统微观状态的概率展现了集中的性质[84].

在经典人工智能领域,"没有免费午餐"定理告 诉我们,在实际情况下难以得到预测完全正确的学 习结果.因此在数据组成的测度空间中,总是能够 找到被分类错误的数据集合 B. 假设数据集所在的 空间具备概率集中性质,大部分数据点存在于错误 分类的集合 B附近,从而微小的扰动就可以使得 原本分类正确的样本得到错误的分类结果.使用 k代表分类的类别,h代表学习算法给出的映射, acc_ϵ代表在强度 ϵ 的扰动下分类器的准确率,那么 可以得到如下关系^[85]:

$$acc_{\epsilon}(h|k) \leq \min\left(acc(h|k), e^{-\frac{1}{2\sigma_{k}^{2}}(\epsilon-\epsilon_{0})^{2}}\right),$$
 (5)

其中σ_k是与数据空间几何性质相关的常数, ε₀是 与分类器准确率相关的常数.可以看到,在存在扰 动的情况下,分类器的准确率随扰动强度的增加 而指数下降.这一结论在使用卷积神经网络识别 MNIST 数据集的实验中得到了验证^[85].

在量子人工智能领域,可以通过在SU(d)中选择幺正变换作用在固定初始态以制备量子数据,

因此可以将对量子态的分类转换为对幺正变换的 分类,其中 *d* 是对应希尔伯特空间的维度.装备了 Haar 测度和 Hilbert-Schmidt 距离的 SU(*d*)是一 个具有测度集中性质的空间,因此有类似的关 系^[53]:

$$acc_{\epsilon} \leqslant \frac{2}{acc} \mathrm{e}^{-\frac{d}{4}\epsilon^{2}},$$
 (6)

其中 ε 为对数据扰动的强度, d 为希尔伯特空间维度, acc 为分类器的准确率, acc ε 为允许强度为 ε 的 扰动下分类器的准确率. 可以发现, 量子分类器的 准确率同样随扰动强度的增加而指数下降.

分类器的脆弱性是一个内禀的性质,不依赖于 学习模型或算法的具体细节."没有免费午餐"定理 表明,在很多情况下会得到问题的近似解,总是存 在一部分数据被分类错误,这是其脆弱性的根本原 因.人工智能算法一般用于处理复杂的高维数据, 而高维数据所表现出的测度集中的性质是分类器 准确率急剧下降的重要原因.大部分数据分布在错 误分类的数据附近,一些微小的扰动就可以使其分 类错误.这两个原因使得分类器的脆弱性成为一个 难以避免的问题.

4.3 量子分类器的对抗样本

接下来介绍如何在不同模式下生成攻击量子 分类器的对抗样本.根据攻击者对量子分类器和学 习算法的掌握情况,分为白盒攻击和黑盒攻击两个 模式.白盒攻击模式假设攻击者掌握量子分类器和 学习算法的全部信息.在白盒攻击模式下,探讨针 对性攻击和无差别攻击两种情况.在黑盒攻击模式 下,攻击者掌握极少(甚至没有)关于量子分类器 和训练算法的信息.

4.3.1 白盒攻击模式: 无差别攻击

无差别攻击以分类器识错样本为目的,在多分 类问题中,攻击者并不在意将样本判错成哪个错误 标签.在经典对抗学习领域有一个著名的例子:攻 击者戴上一个精心设计的眼镜便可以迷惑面部识 别系统从而假装成其他人^[86].考虑到在白盒攻击 模式下,攻击者掌握量子分类器的结构和学习算 法,他们可以先用某个数据集先训练量子分类器, 使得该量子分类器达到很高的准确率,然后固定分 类器里已训练好的量子门参数,将施加在样本数据 上的微扰视为优化参数,希望求得最优的扰动函 数,使得量子分类器最大概率识错该数据集上的对 抗样本,即最大化如下损失函数:

$$U_{\delta} \equiv \operatorname*{argmax}_{U_{\delta} \in \Delta} L\left(h\left(U_{\delta} | \psi \rangle_{\mathrm{in}}; \Theta^{*}\right), a\right), \qquad (7)$$

其中Θ*是量子分类器中已训练好的参数, |ψ⟩_{in}是 输入数据样本, U_δ是施加在|ψ⟩_{in}上的、限制在Δ以 内的对抗微扰, a是样本对应的正确标签, h是 分类器基于已知的数据集学习得到的模型, L是常 用的交叉熵损失函数.图 3(a) 展示了在手写字体 MNIST 数据集上的无差别攻击结果.可以发现, 在所有原始样本上添加很小的、肉眼几乎不可分辨 的微扰, 就能以很大概率获得被分类器识别错的对 抗样本.并且当对抗样本数据和原始数据之间相似 度降低到 73% 时, 所有的对抗样本都会被识别错, 这说明了量子分类器的脆弱性^[50].



图 3 量子分类器在识别 MNIST 中手写字体图片时的对 抗样本 (a) 经过无差别攻击,量子分类器以极高置信度 将数字 7,9 分别识别成 9,7,即使对抗样本和初始样本的 差别非常微小;(b) 通过针对性攻击,量子分类器将把对抗 样本预测为给定错误标签,尽管对抗样本和初始样本相差 无几

Fig. 3. Adversarial examples in quantum learning of MNIST hand-written images: (a) After untargeted attacks, the quantum classifier will misclassify the images of digit 7 (9) as digit 9 (7) with notably high confidence, although the differences between the adversarial and legitimate images are tiny; (b) after targeted attack, the quantum classifier will misclassify the adversarial examples into the category with the targeted label, even though the adversarial and legitimate images only differ slightly from each other.

4.3.2 白盒攻击模式:针对性攻击

针对性攻击目的是让分类器将带有某种标签 的样本数据错判成攻击者期望的标签.比如攻击者 可以试图迷惑带有面部识别系统的电子产品并将 他识别成该产品的实际拥有者,从而取得该电子产 品的控制权^[86].由于要针对性地让对抗样本被预 测为某一个特定标签,考虑最小化如下损失函数:

$$U_{\delta}^{(t)} \equiv \operatorname*{argmin}_{U_{\delta}^{(t)} \in \Delta} L\left(h\left(U_{\delta}^{(t)}|\psi\rangle_{\mathrm{in}};\Theta^{*}\right), a^{(t)}\right), \qquad (8)$$

其中 $a^{(t)}$ 是攻击者期望的标签 (与初始正确的标签 a 不同), $U^{(t)}_{\delta}$ 是施加在 $|\psi\rangle_{in}$ 上的对抗微扰.在图 3(b)中,从 MNIST 数据集中随机选取数字 1, 3, 7, 9的原始样本各一个,采用基本迭代法^[63]最小化 (8) 式以得到相应的对抗样本.尽管原始样本和相应的对抗样本之间的区别肉眼几乎无法察觉,但是添加微扰之后,数字 9, 1, 3, 7 将会被量子分类器分别识别成 1, 3, 7, 9. 这表明在针对性模式下,量子分类器也是易受攻击的^[50].

4.3.3 黑盒攻击模式

在黑盒攻击模式下,攻击者掌握极少(甚至没 有)关于量子分类器和训练算法的信息,这种情况 在实际情况中更为普遍,所以攻击者难以直接根据 分类器结构特征来生成对抗样本.但是由于对抗样 本的可传递性(能够欺骗某一个机器学习模型的对 抗样本,也会有一定的概率可以成功欺骗其他模 型),攻击者仍然能够在不掌握量子分类器的具体 信息,甚至没有量子资源的情况下,产生能够欺骗 量子分类器的对抗样本^[46,47,87].

文献 [50] 探究了对抗样本在经典机器学习模型之间,以及经典机器学习模型和量子机器学习模型之间的可传递性.作者先用 MNIST 数据集训练好两个经典分类器 (卷积神经网络和前馈神经网络),再在白盒无差别攻击模式下,采用不同的迭代方法分别生成它们的对抗样本,最后检验训练好的量子分类器识别这些对抗样本的准确率.结果表明,尽管量子分类器和经典分类器的结构差异巨大,但是基于经典机器学习模型所生成的对抗样本也可以用来有效攻击量子机器学习系统.

4.3.4 对抗微扰并不是随机噪声

以上揭示了量子机器学习系统在面对对抗微 扰时普遍存在的脆弱性.值得强调的是,对抗微扰 并不是随机噪声, 而是精心设计的扰动. 文献 [50] 分别将随机噪声和对抗微扰施加在原始数据上, 对 比它们对分类器识别准确率的影响. 结果表明, 当 施加非关联退相干噪声的时候, 识别准确率将随着 相似度 (原始样本和对抗样本之间的保真度) 的降 低而线性减小 ^[50]. 这种准确率和稳健性之间的折 衷关系在考虑完全未知的噪声时也有所体现 ^[51]. 但是在施加对抗微扰的情况下, 当相似度偏离 100% 时, 识别准确率将显著减少, 甚至当准确率为 0 的 时候, 相似度还可以维持在较高的水平 ^[50]. 这表明 了对抗微扰并不是随机噪声, 同时也体现了量子分 类器针对随机噪声具有很好的稳健性.

4.4 对抗样本的普适性

上述讨论了对于某一个特定分类器,在不同攻 击模式下生成对抗样本的过程.目前已经知道,添 加一个随着量子分类器比特数目 n 指数减小的对 抗微扰就足以得到对抗样本,即对抗微扰强度 e 的 下限随着量子线路规模的增大而指数降低^[53].文 献 [54] 在此基础上将以上结论推广到 k 个分类器: e ~ O(lnk/2ⁿ)的对抗微扰就足以生成一个以较大 概率同时欺骗 k 个分类器的对抗样本.这个结论根 本源于高维希尔伯特空间中的测度集中现象^[88], 与量子分类器的具体结构、训练算法以及数据集 无关.

此外, 文献 [54] 还证明了对于一个给定的量子 分类器, 将一个普适的对抗微扰施加在 m个不同 的初始样本上, 则量子分类器的判错率至少以 $1 - \delta (0 < \delta < 1)$ 的概率限制在内禀错误率附近^[89]:

$$|R_{\rm E} - \mu(\mathcal{E})| \leqslant \sqrt{\frac{1}{2m} \ln\left(\frac{2}{\delta}\right)},\tag{9}$$

其中, *R*_E是量子分类器的判错率, μ(ε)是给定量 子分类器在 Haar 测度下无对抗攻击时的内禀错误 率, ε是误分类的数据集. (9) 式说明当 *m* 越大时, 量子分类器的判错率越接近于内禀错误率. 当 *m* 非常大时, 可以用内禀错误率来估算量子分类器 的判错率.

进一步地, 文献 [54] 证明内禀错误率的期望值 满足如下不等式^[80,81,90]:

$$\mathbb{E}_{U}\left[\mathbb{E}_{S}\left[\mu\left(\mathcal{E}\right)\right]\right] \ge 1 - \frac{d'}{d\left(d+1\right)}\left(N^{2} + d + 1\right), \quad (10)$$

其中 U是未知的真实目标映射, S是规模为 N的

训练集, *d* 是输入空间的维度, *d'* 是输出标签的数 目. 从 (10) 式可以看出, 内禀错误率的期望值随着 输出标签数或训练集数目的增加而减小, 随着样本 空间的维度增大而增大. 当 *d* 趋于无穷时, 内禀错 误率会逼近 100%, 且与量子分类器的结构和训练 算法无关.

综上所述,尽管在所有可能的数据样本中添加 单一的对抗微扰平均而言并不会使对抗风险增大. 但是对仅包含 m个初始样本的有限集合而言,单 一的对抗微扰仍然可能会增加量子分类器的判错 率. 文献 [54] 通过详实的数值实验证实这种普适对 抗微扰的存在:将其添加到一组不同的初始样本 中,便能以极大概率生成一组对抗样本同时欺骗某 个特定的分类器.

5 防 御

通过以上的讨论可以清楚地得到, 经典和量子 的分类器都很容易受到对抗微扰的影响——对抗 样本普遍存在. 这可能引发人们对量子学习系统可 靠性的关切, 尤其在那些安全性至关重要的领域, 例如自动驾驶^[4]和医疗诊断^[74]等. 因此, 研究可能 的防御策略以提高量子分类器的稳健性具有重要 的理论与实际意义^[91].

实际上,完全消除对抗攻击带来的风险是非常 困难的.首先,很难为对抗学习过程创建一个精确 的理论模型.这是一个高度非线性且复杂的非凸优 化过程,目前缺乏适当的理论工具来分析它^[61].因 此,要从理论上分析并阐明一个特定的防御策略 能否防御对抗攻击是极其困难的.此外,要达到防 御对抗攻击的目的,我们希望分类器能够对每种 可能的输入产生正确的输出,其数量通常随着问 题的变大而指数增长.原则上这要求学习模型的复 杂度指数增长.在大多数情况下,机器学习模型只 能在所有可能输入的一小部分样本中具有良好表 现^[10,11,76].

尽管如此,近年来在经典对抗机器学习领域, 人们提出了多种防御策略来减小对抗样本带来的 影响,其中包括对抗训练^[60]、梯度隐藏^[92]、量子噪 声添加^[91]、防御性蒸馏^[93]和防御-生成对抗网络 (defense-generative adversarial network, defense-GAN)^[94]等.每种策略都有其自身的优点和缺点, 然而并没有一个策略能够应对所有类型的对抗攻 击.本节探讨如何提高量子分类器关于对抗攻击的 稳健性.数值实验表明,前文提到的对抗训练方法 可以显著提高量子分类器在防御特定对抗攻击时 的表现.

对抗训练的基本思想是通过将对抗样本注入 训练集中来增强模型的稳健性.这是一种非常简单 且直接的方法:给定多种攻击策略生成的对抗样 本,可以把这些对抗样本与初始样本结合成新的训 练集,并重新训练分类器.特别地,如果采用稳健 优化^[01]方法来训练量子分类器,那么可以将任务 归结于一个典型的极小化极大的优化问题:

$$\min_{\Theta} \frac{1}{N} \sum_{i=1}^{N} \max_{U_{\delta} \in \Delta} L\left(h\left(U_{\delta} |\psi\rangle_{\text{in}}^{(i)}; \Theta\right), y^{(i)}\right), \quad (11)$$

其中|ψ⟩_n⁽ⁱ⁾是受到攻击的第 i个样本, y(i)表示其原 始对应标签.重新训练量子分类器可以降低对抗攻 击的风险,该风险由输入样本最坏情况的平均损失 函数来描述.这种极小化极大问题在稳健优化领域 已经得到广泛的研究,存在诸多解决此类问题的方 法.其中一个卓有成效的方法是把上述表达式拆分 为两部分:外部最小化和内部最大化.内部最大化 与生成对抗样本的过程完全相同,已经在上述章节 进行了讨论.外部最小化则归结为最小化对抗样本 的损失函数,可以运用前文提到的迭代算法实现.

值得一提的是,虽然经过对抗训练的量子分类 器针对某类对抗扰动的稳健性确实会得到显著提 升,但是一般只能对由相同攻击方法生成的对抗样 本表现良好.当攻击者采用其他攻击策略时,分类 器的防御性能可能急剧下降.此外由于黑盒攻击的 梯度掩蔽 (gradient masking),对抗训练倾向于使 量子分类器在防御白盒攻击上相比于黑盒攻击更 加稳健^[87,92].实际上,尽管一种防御策略可以抵抗 针对量子分类器的一种攻击方式,但是其将不可避 免地会给知道并了解其防御机制的攻击者开放另 一种漏洞.

在经典对抗学习领域,另一种新的防御机制最 近受到了广泛关注,它可以有效地应对白盒和黑盒 攻击.这便是上文提到的 defense-GAN^[94].该机制 不直接把输入样本导入分类器,而是首先把样本导 入 GAN 的生成器中重新生成输入数据,然后再将 其输入分类器.该防御策略的核心在于利用 GAN 强大的表达能力减弱甚至过滤掉对抗微扰带来的 影响.最近 GAN 的量子版本 (QGAN) 已经在理 论上被提出^[39,95],同时 QGAN 的实验原型机也在 超导量子电路上得以实现^[40,96].研发一个 defense-QGAN 来增强量子分类器关于对抗微扰的稳健性 也是至关重要的,这将是未来一个有趣且充满前景 的研究方向.

6 总结与展望

本文系统地回顾了经典和量子机器学习在不同情况下关于对抗样本的脆弱性.由于测度集中现象^[88],对抗样本存在的普遍性是高维空间中量子机器学习应用的基本特征.无论是经典还是量子的分类器,在原始数据中添加精心设计的细微扰动都可能导致分类器以非常高的置信度做出错误的预测.本文总结了针对多种不同的任务和分类器生成对抗样本的通用方法,并回顾了在不同对抗环境中的具体示例.通过对抗训练,研究人员发现分类器对于特定类型的对抗扰动的脆弱性可以得到显著抑制.

值得强调的是,本文所讨论的量子对抗学习与 前面提到的QGAN存在本质区别^[39,40,43,95,97].QGAN 包含两个主要部分:生成器和判别器.它们通过对 抗博弈的方式进行交替训练:在每个学习回合中, 判别器都会优化其策略,以识别生成器产生的虚假 数据,而生成器会进一步更新其伪造数据的机制来 欺骗判别器.最终,这种动态博弈的训练过程将达 到纳什(Nash)均衡.理想情况下,生成器在达到纳 什均衡后生成的数据将与原始训练集中的真实数 据满足相同的统计规律,而辨别器将不能以大于一 半的概率分辨出伪造的数据.因此,QGAN的主要 目标是生成匹配原始训练数据的统计信息的新数 据(无论是经典的或量子的).与之不同的是,本文 介绍的对抗学习主要侧重如何生成对抗样本以及 如何防御对抗攻击.

本文仅揭示了对抗学习的冰山一角,该领域还 有许多重要问题值得进一步研究.本文的总结主要 集中在基于人工神经网络和变分量子线路的监督 学习上,但是无监督学习和强化学习也可能遭受脆 弱性问题的困扰^[59].因此,将对抗学习推广到其他 机器学习类型或许是可行的.另外,我们猜想深度 学习中对抗微扰的普遍存在性与量子多体物理 中的正交灾难现象(即添加任意弱的局部扰动后, 量子系统的基态在热力学极限下与原始基态正 交)^[98,99]之间可能存在深远的联系.最后,设计并进 行一个实验来展示对抗样本的普遍性和普适微扰 的存在性也亟待研究.这将是未来量子技术在人工 智能中实际应用的重要一步,尤其是在安全性至关 重要的领域,例如无人驾驶汽车、恶意软件检测、 生物识别和医疗诊断^[74]等.

感谢清华大学交叉信息研究院龚维元、蒋飔和马克斯-普朗克研究所陆思锐的有益讨论.

参考文献

- Krizhevsky A, Sutskever I, Hinton G E 2012 Proceedings of the 25th International Conference on Neural Information Processing Systems (Volume 1) New York, USA, December 3-8, 2012 p1097
- [2] Hinton G, Deng L, Yu D, Dahl G E, Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath T N, Kingsbury B 2012 *IEEE Signal Process. Mag.* 29 82
- [3] Kononenko I 2001 Artif. Intell. Med. 23 89
- [4] Grigorescu S, Trasnea B, Cocias T, Macesanu G 2020 J. Field Robot 37 362
- [5] LeCun Y, Bengio Y, Hinton G 2015 Nature 521 436
- [6] Jordan M, Mitchell T 2015 Science **349** 255
- [7] Silver D, Huang A, Maddison C J, et al. 2016 Nature 529 484
- [8] Silver D, Schrittwieser J, Simonyan K, et al. 2017 Nature 550 354
- [9] Senior A W, Evans R, Jumper J, et al. 2020 Nature 577 706
- [10] Russell S, Norvig P 2020 Artificial Intelligence: A Modern Approach (Hoboken: Pearson) pp1–61
- Bishop C 2006 Pattern Recognition and Machine Learning (New York: Springer-Verlag) pp225-284
- [12] Nielsen M A, Chuang I L 2010 Quantum Computation and Quantum Information (Cambridge: Cambridge University Press) pp171-352
- [13]~ Arute F, Arya K, Babbush R, et al. 2019 Nature~574~505
- [14] Zhong H S, Wang H, Deng Y H, Chen M C, Peng L C, Luo Y H, Qin J, Wu D, Ding X, Hu Y, Hu P, Yang X Y, Zhang W J, Li H, Li Y, Jiang X, Gan L, Yang G, You L, Wang Z, Li L, Liu N L, Lu C Y, Pan J W 2020 Science **370** 1460
- [15] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S 2017 Nature 549 195
- [16] Dunjko V, Briegel H J 2018 Rep. Prog. Phys. 81 074001
- [17] Das Sarma S, Deng D L, Duan L M 2019 Phys. Today 72 48
- [18] Carleo G, Troyer M 2017 Science 355 602
- [19] Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R, Carleo G 2018 Nat. Phys. 14 447
- [20] Zhang Y H, Zheng P L, Zhang Y, Deng D L 2020 Phys. Rev. Lett. 125 170501
- [21] Deringer V L, Bernstein N, Csányi G, Ben Mahmoud C, Ceriotti M, Wilson M, Drabold D A, Elliott S R 2021 Nature 589 59
- [22] Deng D L 2018 Phys. Rev. Lett. 120 240402
- [23] Deng D L, Li X, Das Sarma S 2017 Phys. Rev. B 96 195145
- [24] Zhang Y, Kim E A 2017 Phys. Rev. Lett. 118 216401
- [25] Carrasquilla J, Melko R G 2017 Nat. Phys. 13 431
- [26] van Nieuwenburg E P L, Liu Y H, Huber S D 2017 Nat. Phys. 13 435

- [27] Wang L 2016 Phys. Rev. B 94 195105
- [28] Broecker P, Carrasquilla J, Melko R G, Trebst S 2017 Sci. Rep. 7 8823
- [29] Ch'ng K, Carrasquilla J, Melko R G, Khatami E 2017 Phys. Rev. X 7 031038
- [30] Wetzel S J 2017 Phys. Rev. E 96 022140
- [31] Hu W, Singh R R P, Scalettar R T 2017 Phys. Rev. E 95 062122
- [32] Zhang Y, Mesaros A, Fujita K, Edkins S D, Hamidian M H, Ch'ng K, Eisaki H, Uchida S, Davis J C S, Khatami E, Kim E-A 2019 Nature 570 484
- [33] Lian W, Wang S T, Lu S, Huang Y, Wang F, Yuan X, Zhang W, Ouyang X, Wang X, Huang X, He L, Chang X, Deng D L, Duan L 2019 Phys. Rev. Lett. 122 210503
- [34] Harrow A W, Hassidim A, Lloyd S 2009 Phys. Rev. Lett. 103 150502
- [35] Lloyd S, Mohseni M, Rebentrost P 2014 Nat. Phys. 10 631
- [36] Dunjko V, Taylor J M, Briegel H J 2016 Phys. Rev. Lett. 117 130501
- [37] Amin M H, Andriyash E, Rolfe J, Kulchytskyy B, Melko R 2018 Phys. Rev. X 8 021050
- [38] Gao X, Zhang Z Y, Duan L M 2018 Sci. Adv. 4 eaat9004
- [39] Lloyd S, Weedbrook C 2018 Phys. Rev. Lett. **121** 040502
- [40] Hu L, Wu S H, Cai W, Ma Y, Mu X, Xu Y, Wang H, Song Y, Deng D L, Zou C L, Sun L 2019 Sci. Adv. 5 eaav2761
- [41] Schuld M, Killoran N 2019 Phys. Rev. Lett. 122 040504
- [42] Rebentrost P, Mohseni M, Lloyd S 2014 Phys. Rev. Lett. 113 130503
- [43] Chakraborty A, Alam M, Dey V, Chattopadhyay A, Mukhop adhyay D 2018 arXiv: 1810.00069 [cs, stat]
- [44] Biggio B, Roli F 2018 Pattern Recognit. 84 317
- [45] Miller D J, Xiang Z, Kesidis G 2019 arXiv: 1904.06292 [cs, stat]
- [46] Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R 2014 arXiv: 1312.6199 [cs]
- [47] Goodfellow I J, Shlens J, Szegedy C 2015 arXiv: 1412.6572 [cs, stat]
- [48] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y 2014 Adv. Neural Inf. Process. Syst. 27 2672
- [49] Preskill J 2018 Quantum 2 79
- [50]~ Lu S, Duan L M, Deng D L 2020 Phys. Rev. Res. 2 033212
- [51] Guan J, Fang W, Ying M 2020 arXiv: 2008.07230 [quant-ph]
- [52] Wiebe N, Kumar R S S 2018 New J. Phys. **20** 123019
- [53] Liu N, Wittek P 2020 Phys. Rev. A 101 062331
- [54]Gong W, Deng D L 2021 arXiv: 2102.07788 [cond-mat, physics: quant-ph]
- [55] Dalvi N, Domingos P, Mausam, Sanghai S, Verma D 2004 Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Seattle, USA, August 22-25, 2004 p99
- [56] Lowd D, Meek C 2005 Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining Chicago, USA, August 21–24, 2005 p641
- [57] Barreno M, Nelson B, Sears R, Joseph A D, Tygar J D 2006 Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security Taipei, China, March 21-24, 2006 p16
- [58] Huang L, Joseph A D, Nelson B, Rubinstein B I P, Tygar J D 2011 Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence Chicago, USA, October 21, 2011 p43
- [59] Vorobeychik Y, Kantarcioglu M 2018 Synth. Lect. Artif.

Intell. Mach. Learn. 12 1

- [60] Kurakin A, Goodfellow I, Bengio S 2017 arXiv: 1611.01236 [cs, stat]
- [61] Ben-Tal A, Ghaoui L E, Nemirovski A 2009 Robust Optimization (Princeton: Princeton University Press) pp1-146
- [62] Madry A, Makelov A, Schmidt L, Tsipras D, Vladu A 2018 International Conference on Learning Representations Vancouver, Canada, April 30–May 3, 2018 p1
- [63] Kurakin A, Goodfellow I, Bengio S 2017 arXiv: 1607.02533 [cs, stat]
- [64] Dong Y, Liao F, Pang T, Su H, Zhu J, Hu X, Li J 2018 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Salt Lake City, USA, June 18–23, 2018 p9185
- [65] Hsu Y T, Li X, Deng D L, Das Sarma S 2018 Phys. Rev. Lett. 121 245701
- [66] Rodriguez-Nieva J F, Scheurer M S 2019 Nat. Phys. 15 790
- [67] Zhang P, Shen H, Zhai H 2018 Phys. Rev. Lett. 120 066401
- [68] Huembeli P, Dauphin A, Wittek P 2018 Phys. Rev. B 97 134109
- [69] Rem B S, Käming N, Tarnowski M, Asteria L, Fläschner N, Becker C, Sengstock K, Weitenberg C 2019 Nat. Phys. 15 917
- [70] Bohrdt A, Chiu C S, Ji G, Xu M, Greif D, Greiner M, Demler E, Grusdt F, Knap M 2019 Nat. Phys. 15 921
- [71] Jiang S, Lu S, Deng D L 2019 arXiv: 1910.13453 [cond-mat, physics: quant-ph]
- [72] Neupert T, Santos L, Ryu S, Chamon C, Mudry C 2012 Phys. Rev. B 86 035125
- [73] Deng D L, Wang S T, Duan L M 2014 Phys. Rev. A 90 041601
- [74] Finlayson S G, Bowers J D, Ito J, Zittrain J L, Beam A L, Kohane I S 2019 Science 363 1287
- [75] Ren K, Zheng T, Qin Z, Liu X 2020 Engineering 6 346
- [76] Goodfellow I, Bengio Y, Courville A 2016 Deep Learning (Cambridge: The MIT Press) pp98–165
- [77] Sutskever I, Vinyals O, Le Q V 2014 Proceedings of the 27th International Conference on Neural Information Processing Systems (Volume 2) Montréal, Canada, December 8–13, 2014 p3104
- [78] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L, Polosukhin I 2017 arXiv: 1706.03762 [cs]
- [79] Wolpert D H, Macready W G 1997 IEEE Trans. Evol. Comput. 1 67
- [80] Shalev-Shwartz S, Ben-David S 2014 Understanding Machine Learning: From Theory to Algorithms (New York: Cambridge University Press) pp36-41
- [81] Poland K, Beer K, Osborne T J 2020 arXiv: 2003.14103 [quant-ph]
- [82] Schroff F, Kalenichenko D, Philbin J 2015 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Boston, USA, June 7–12, 2015 p815
- [83] Walters M 2015 arXiv: 1508.05448 [math-ph]
- [84] Schwabl F 2006 Statistical Mechanics (Berlin Heidelberg: Springer-Verlag) pp1-20
- [85] Dohmatob E 2019 arXiv: 1810.04065 [cs, stat]
- [86] Sharif M, Bhagavatula S, Bauer L, Reiter M K 2016 Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security Vienna, Austria, October 24–28, 2016 p1528
- [87] Papernot N, McDaniel P, Goodfellow I, Jha S, Celik Z B, Swami A 2017 Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security Abu

Dhabi, UAE, April 2–6 2017 p
506 $\,$

- [88] Ledoux M 2001 The Concentration of Measure Phenomenon (Providence: American Mathematical Society) pp1–21
- [89] Hoeffding W 1963 J. Am. Stat. Assoc. 58 13
- [90] Sharma K, Cerezo M, Holmes Z, Cincio L, Sornborger A, Coles P J 2020 arXiv: 2007.04900 [quant-ph]
- [91] Du Y, Hsieh M H, Liu T, Tao D, Liu N 2020 arXiv: 2003.09416 [quant-ph]
- [92] Tramèr F, Kurakin A, Papernot N, Goodfellow I, Boneh D, McDaniel P 2017 arXiv: 1705.07204 [cs, stat]
- [93] Papernot N, McDaniel P, Wu X, Jha S, Swami A 2016 37th IEEE Symposium on Security and Privacy San Jose, USA,

May 23-25, 2016 p582

- [94] Samangouei P, Kabkab M, Chellappa R 2018 arXiv: 1805.06605 [cs, stat]
- [95] Dallaire-Demers P L, Killoran N 2018 Phys. Rev. A 98 012324
- [96] Huang K, Wang Z A, Song C, Xu K, Li H, Wang Z, Guo Q, Song Z, Liu Z-B, Zheng D, Deng D L, Wang H, Tian J G, Fan H 2020 arXiv: 2009.12827 [quant-ph]
- [97] Zeng J, Wu Y, Liu J G, Wang L, Hu J 2019 Phys. Rev. A 99 052306
- [98] Anderson P W 1967 Phys. Rev. Lett. 18 1049
- [99] Deng D L, Pixley J H, Li X, Das Sarma S 2015 Phys. Rev. B 92 220201

SPECIAL TOPIC—Machine learning and physics

Adversarial learning in quantum artificial intelligence^{*}

Shen Pei-Xin¹⁾ Jiang Wen-Jie¹⁾ Li Wei-Kang¹⁾

Lu Zhi-De¹) Deng Dong-Ling^{1)²)^{\dagger}}

1) (Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China)

2) (Shanghai Qi Zhi Institute, Shanghai 200232, China)

(Received 25 April 2021; revised manuscript received 26 May 2021)

Abstract

Quantum artificial intelligence exploits the interplay between artificial intelligence and quantum physics: on the one hand, a plethora of tools and ideas from artificial intelligence can be adopted to tackle intricate quantum problems; on the other hand, quantum computing could also bring unprecedented opportunities to enhance, speed up, or innovate artificial intelligence. Yet, quantum learning systems, similar to classical ones, may also suffer adversarial attacks: adding a tiny carefully-crafted perturbation to the legitimate input data would cause the systems to make incorrect predictions at a notably high confidence level. In this paper, we introduce the basic concepts and ideas of classical and quantum adversarial learning, as well as some recent advances along this line. First, we introduce the basics of both classical and quantum adversarial learning. Through concrete examples, involving classifications of phases of two-dimensional Ising model and threedimensional chiral topological insulators, we reveal the vulnerability of classical machine learning phases of matter. In addition, we demonstrate the vulnerability of quantum classifiers with the example of classifying hand-written digit images. We theoretically elucidate the celebrated no free lunch theorem from the classical and quantum perspectives, and discuss the universality properties of adversarial attacks in quantum classifiers. Finally, we discuss the possible defense strategies. The study of adversarial learning in quantum artificial intelligence uncovers notable potential risks for quantum intelligence systems, which would have far-reaching consequences for the future interactions between the two areas.

Keywords: quantum artificial intelligence, quantum adversarial learning, quantum classifiers, topological phases of mater

PACS: 03.67.–a, 03.67.Ac, 05.30.Rt, 07.05.Mh

DOI: 10.7498/aps.70.20210789

^{*} Project supported by the National Natural Science Foundation of China (Grant No. 12075128), the Start-up Fund from Tsinghua University, China (Grant No. 53330300320), and the Shanghai Qi Zhi Institute, China.

[†] Corresponding author. E-mail: dldeng@mail.tsinghua.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

量子生成模型

孙太平 吴玉椿 郭国平

Quantum generative models for data generationSun Tai-PingWu Yu-ChunGuo Guo-Ping引用信息 Citation: Acta Physica Sinica, 70, 140304 (2021)DOI: 10.7498/aps.70.20210930在线阅读 View online: https://doi.org/10.7498/aps.70.20210930当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

周期驱动量子伊辛模型中非热统计的形成与抑制

Formation and suppression of nonthermal statistics in peridically driven quantum Ising models 物理学报. 2020, 69(14): 140501 https://doi.org/10.7498/aps.69.20191657

量子计算与量子模拟

Quantum computation and quantum simulation 物理学报. 2018, 67(12): 120301 https://doi.org/10.7498/aps.67.20180710

量子相干

Quantum coherence 物理学报. 2019, 68(3): 030304 https://doi.org/10.7498/aps.68.20181779

横场中具有周期性各向异性的一维XY模型的量子相变

Quantum phase transitions of one-dimensional period-two anisotropic XY models in a transverse field 物理学报. 2017, 66(18): 180302 https://doi.org/10.7498/aps.66.180302

基于量子隐形传态的量子保密通信方案

Quantum communication scheme based on quantum teleportation 物理学报. 2017, 66(23): 230303 https://doi.org/10.7498/aps.66.230303

Tavis-Cummings模型中的几何量子失协特性 Geometric quantum discord in Tavis-Cummings model 物理学报. 2018, 67(11): 110301 https://doi.org/10.7498/aps.67.20172699 专题: 机器学习与物理

量子生成模型*

孙太平1) 吴玉椿1)2)† 郭国平1)2)3)

(中国科学技术大学,中国科学院量子信息重点实验室,合肥 230026)
 2)(合肥综合性国家科学中心人工智能研究院,合肥 230088)
 3)(合肥本源量子计算科技有限责任公司,合肥 230026)
 (2021 年 5 月 17 日收到; 2021 年 6 月 30 日收到修改稿)

近年来,很多基于生成模型的机器学习算法,如生成对抗网络、玻尔兹曼机、自编码器等,在数据生成、 概率模拟等方面有广泛的应用.另一方面,融合量子计算和经典机器学习的量子机器学习算法也不断被提出. 特别地,量子生成模型作为量子机器学习的分支,目前已有很多进展.量子生成算法是一类量子-经典混合算 法,算法中引入参数量子线路,通过执行参数线路得到损失函数及其梯度,然后通过经典的优化算法来优化 求解,从而完成对应的生成任务.与经典生成模型相比,量子生成模型通过参数线路将数据流映射到高维希 尔伯特空间,在高维空间中学习数据的特征,从而在一些特定问题上超越经典生成模型.在中等规模含噪声 的量子体系下,量子生成模型是一类有潜力实现量子优势的量子机器学习算法.

关键词:量子机器学习,量子生成模型,量子优势 PACS:03.67.-a,03.67.Ac

DOI: 10.7498/aps.70.20210930

1 引 言

在经典计算领域,生成模型被广泛应用于各个领域,如图像生成^[1]、语音合成^[2]、药物设计^[3]、图像转换^[4]等.近来已有对经典生成模型的量子版本的研究,如量子生成对抗网络^[5,6]、量子玻恩机^[7,8]、量子玻尔兹曼机^[9-12]、量子自编码器^[13-15]以及量子贝叶斯网络^[16]等.

量子生成模型是受经典机器学习模型启发,使 用量子计算机来改进经典模型中的生成步骤,以完 成数据生成、概率模拟任务的一类算法.

这类算法的主要步骤是,通过量子线路生成量 子态,用量子态观测结果构建损失函数,再通过经 典的优化算法进行迭代优化,从而得到最优的参数 变量集合以完成相应的生成任务.因此,这些算法 被统称为量子-经典混合算法.这种量子-经典混合的闭环优化思想始于变分量子算法^[17],即通过量子计算机和参数化量子门操作集合获得分子的量子态表示,测量分子模型对应的哈密顿量的期望值,以此作为目标函数进行变分优化.

量子相位估计^[18]在大数分解^[18]、HHL^[19]等被 证明有量子优势的算法上具有广泛的应用,但基于 以下原因相位估计算法无法在当前的计算体系下 有效实现:1)当前硬件体系多集中于单量子门和 两比特纠缠门的实现,相位估计用到的受控幺正 门(*c* – *U^j*)必须分解为基本门的组合,此过程会 显著降低受控幺正门的保真度;2)相位估计用到 的量子傅里叶逆变换需要*O*(*n*²)(*n* 为量子比特个 数)数量级的量子门,使得深层量子线路的态保真 度无法达到量子计算要求.量子计算体系的噪声不 可避免,当前只有中等规模的量子比特、无法实现

© 2021 中国物理学会 Chinese Physical Society

^{*} 国家重点研究发展计划 (批准号: 2016YFA0301700)、国家自然科学基金 (批准号: 11625419)、中国科学院战略重点研究计划 (批 准号: XDB24030600) 和安徽省量子信息技术倡议 (批准号: AHY080000) 资助的课题.

[†] 通信作者. E-mail: wuyuchun@ustc.edu.cn

有效的量子纠错^[20]技术,这使量子-经典混合算法 受到广泛关注.

尽管已有多种量子生成任务在真实量子芯片 上完成,如利用量子生成对抗网络进行图像生成 与识别^[21],但对于参数量子线路(Parameterized Quantum Circuit, PQC,见附录)的表示能力与经 典网络相比是否具有优势还缺少理论的结果. Du 等^[22]尝试从内在的线路结构出发证明其结论, 而经典深度学习神经网络和量子线路的比较还 没有结论.同样的,量子生成网络作为量子机器学 习的一部分,如何将经典数据转化为量子态、如何 从测量结果获取经典信息仍需要更直观更普适的 方法.

本文主要分析一些常见量子生成模型的优势 和劣势,并对其扩展性展开讨论.量子生成模型的 分类与经典生成模型类似,是以任务为导向的.量 子生成模型基本沿袭与之对应的经典生成模型的 命名,如量子生成对抗网络是使用量子线路来完成 经典生成对抗任务的一种量子生成模型.对于利用 量子体系特有的性质以完成生成任务的,如量子玻 恩机,则是用量子专有的玻恩测量命名.还有一些 模型保留了原作者的命名方式,归并在"其他生成 模型"这一节进行介绍.一些量子生成模型如量子 生成对抗网络、量子玻恩机、量子自编码器等用到 的参数量子线路,在附录里会有详细的介绍.通过 对这些模型的比较,可以从不同角度理解量子生成 模型的构建思路,以此为不同的计算任务构建与之 适应的计算模型提供启发.

2 量子生成模型

2.1 量子生成对抗网络

经典的生成对抗网络^[23](generative adversarial Network, GAN) 被视为深度学习中的前沿进展. GAN 在生成图像^[1]、语音^[2]、文本^[24]等方面具有非常广泛的应用. 经典生成对抗网络是一类有两个玩家的博弈游戏,一个是生成器 (generator), 记为G, 另一个为鉴别器 (discriminator), 记为D. 对于已存在的源数据 $\vec{x} \sim p_{data}$, 生成器G生成新的数据以逼近真实的源数据分布, 使D无法区分数据的来源; 而鉴别器D对每个输入数据给出真或假的决策概率, 以最大程度地区分这两类数据.

如果 G 产生的数据足够逼近目标数据, D 无

论对于何种输入数据,都有 50% 的几率误判,这时 生成对抗网络就达到了 Nash 平衡点.设真实的源 数据标签为 1,由 G 产生的数据标签为 0,D 输出 (0,1)间的概率值,那么生成对抗网络的损失函数 可以被定义为:

$$\mathcal{L}(D,G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}}[\log D_{\boldsymbol{\theta}_{\text{D}}}(\boldsymbol{x})] \\ + \mathbb{E}_{\boldsymbol{z} \sim p(\boldsymbol{z})}[\log(1 - D_{\boldsymbol{\theta}_{\text{D}}}(G_{\boldsymbol{\theta}_{\text{G}}}(\boldsymbol{z}))].$$
(1)

生成对抗网络的训练策略是,更新其中一方的参数时,另一方的参数保持不变.对于上述的损失函数模型,优化任务为min_θ max_θ $\mathcal{L}(D,G)$.采用经典神经网络构建的 GAN 模型,可以用反向传播算法来实现梯度计算.

量子的生成对抗网络^[5,21,25,26](quantum GAN, QGAN)基本保留了经典生成对抗模型的特征,不 同之处在于量子生成对抗网络用到的数据集既 可以是量子的也可以是经典的,而且可选择地,生 成器与鉴别器也可以为经典神经网络或参数量子 线路.

对于数据是量子的, 生成器与鉴别器也都是量子的情况, Dallaire-Demers 等 ^[5]提出了这样的QGAN模型:源数据的概率分布模型为 $p_R(x)$, 由某个未知 (有可能很复杂)的过程 R 生成. 由参数 θ_G 确定的生成器 G 生成新的数据样本, 由参数 θ_D 确定的鉴别器 D 识别这两类数据. 用标签 λ 指定源数据中的类别, $R = R(|\lambda\rangle) = \rho_{\lambda}^{R}$,考虑噪声 $|z\rangle$, 可以按照 (1)式定义与参数有关的损失函数 $\mathcal{L}(\theta_D, \theta_G)$, 这里 D 的输入是 $R(|\lambda\rangle)$ 或 $G(\theta_G, |\lambda, z\rangle)$):

$$\mathcal{L}(\boldsymbol{\theta}_{\mathrm{D}}, \boldsymbol{\theta}_{\mathrm{G}}) = \frac{1}{2} + \frac{1}{2\Lambda} \sum_{\lambda=1}^{\Lambda} [\cos^{2}(\phi) \mathrm{tr}(Z\rho_{\lambda}^{DR}(\boldsymbol{\theta}_{\mathrm{D}})) - \sin^{2}(\phi) \mathrm{tr}(Z\rho_{\lambda}^{DG}(\boldsymbol{\theta}_{\mathrm{D}}, \boldsymbol{\theta}_{\mathrm{G}}, z))], \quad (2)$$

其中含 ϕ 的项是调整源数据和生成部分占整体的比例. 一般取 $\phi = \frac{\pi}{4}$. Dallaire-Demers 等构建的量子生成对抗网络线路图如图 1 所示.

量子生成对抗网络的任务就是通过参数优化 来实现 min_{θ_G} max_{θ_D} *L*(θ_D, θ_G),从而得到最接近于 源数据概率分布的分布模型与鉴别模型.量子线路 的梯度可以用数值差分^[27]、参数位移法则^[27]、以 及 Hadamard test 线路^[5] 来求. Dallaire-Demers 在数值模拟上实现了源数据为|0),|1)两个量子态的 生成对抗任务.

Lloyd 等^[6] 讨论了当数据是量子的, 鉴别器是

量子或经典的,生成器是经典的量子生成对抗模型. Lloyd 所述的量子数据是通过对一个量子体系进行固定的测量得到的数据分布.根据量子优越性^[19]的表述,经典的生成器不能有效地匹配量子数据分布,或者更准确地,除非有指数的资源可用,经典的生成器无法模拟量子数据的分布 *p*_R(*x*).对于数据是经典的,生成器和鉴别器都是量子的情况,Lloyd 通过对比得出,经典的生成器和鉴别器在计算梯度时需要*O*(*N*²)个比特和*O*(*N*²)步时间(*N*是经典数据的维度),而量子的生成器和鉴别器在计算协方差矩阵时,如果矩阵是低秩和稀疏的,需要*O*(log*N*)个量子比特和*O*(poly(log*N*))个逻辑门操作.



Fig. 1. The general structure of QGAN^[5].

Zeng 等^[26]结合量子玻恩机构建了经典数据-量子生成器-经典鉴别器的量子生成对抗模型,实现条带图 (bars and stripes, BAS)^[8,16,26,28]的概率 生成与识别 (见第 2 节). 类似的, Situ 等^[29]将 BAS 图的每个点都编码至量子比特上,通过参数量子线 路构建的生成器生成量子态,测量结果输入至经典 神经网络构建的鉴别器,也实现了 BAS 图的生成. Zouful 等^[25]通过测量量子态在计算基上的投影, 将其转化为数据对应的概率,以此为基础构建生成 对抗网络,实现了对数正态分布、双峰高斯分布等 概率分布的学习和生成.值得注意的是,连续数据 被截断并舍入到临近整数值,所以该模型仍是离散 数据的概率学习模型.

随着量子硬件操控水平的提高,在量子芯片上 实现量子生成对抗网络已经有相关的工作.Hu^[30] 等利用超导量子体系首次实现了具有量子数据、量 子生成器、量子鉴别器的量子生成对抗网络的实验 验证,生成保真度高达 98.8%.在 Hu 的实验中,量 子数据为单比特量子态,生成器参数为量子门旋转 角度. 在优化 D 或 G 时采用梯度下降算法, 量子 系统的操控和测量通过 FPGA(field programmable gate arrays) 实现, 线路梯度通过经典计算机 估算得到. 该实验对于 NISQ 体系下的量子机器学 习优势的探索具有重要意义. 此外, Huang 等^[21] 通过将测量量子生成器得到的数据作为经典图像 灰度值, 完成了手写数字集 0 和 1 的生成对抗任 务, 为量子比特资源受限条件下经典数据的生成提 供了一个思路.

2.2 量子玻恩机

量子玻恩机^[7,8](quantum circuit Born machine, QCBM) 是基于波函数的概率解释的一种模型.量 子玻恩机将概率分布编码至初态上,把量子态在计 算基上的投影作为数据的概率.量子玻恩机是一种 隐式的生成模型,因为我们实际上并不需要知道这 个纯态的密度矩阵,而只需要知道叠加态振幅的模 值.这样的设定对量子线路的结构限制也更少,所 以一般被认为具有更强的表示能力.

对目标概率分布 $\pi(x)$ 进行采样得到数据集 $\mathcal{D} = \{x\}$, QCBM 通过参数量子线路把初态 $|0\rangle$ 演 化为 $|\psi_{\theta}\rangle$,并在计算基上的测量得到生成数据 $x \sim p_{\theta}(x) = |\langle x | \psi_{\theta} \rangle|^2$ 的概率分布. 训练目标是让模型的 概率分布 p_{θ} 接近真实数据分布 $\pi(x)$. QCBM 采用 所谓的 MMD(Maximum Mean Discrepancy) 损失 函数:

$$\mathcal{L} = \left\| \sum_{x} p_{\theta}(x)\phi(x) - \sum_{x} \pi(x)\phi(x) \right\|^{2}, \quad (3)$$

其中 ϕ 将 x 映射到高维 Hilbert 空间. 进一步地, 定义核函数 $K(x,y) = \phi(x)^{T}\phi(y)$, (3) 式即可避免 在高维空间中进行运算,等价于

$$\mathcal{L} = \mathop{\mathbb{E}}_{x \sim p_{\theta}, y \sim p_{\theta}} [K(x, y)] - 2 \mathop{\mathbb{E}}_{x \sim p_{\theta}, y \sim \pi} [K(x, y)] + \mathop{\mathbb{E}}_{x \sim \pi, y \sim \pi} [K(x, y)].$$
(4)

核 函 数 可 以 取 为 高 斯 形 式 : $K(x,y) = \exp\left(-\frac{1}{2\sigma}|x-y|^2\right)$. 直接计算上述损失函数的梯度 比较困难, 但可以证明, 下式无偏地等价于梯度:

$$\frac{\partial \mathcal{L}}{\partial \theta_l^{\alpha}} = \underset{x \sim p_{\theta^+}, y \sim p_{\theta}}{\mathbb{E}} [K(x, y)] - \underset{x \sim p_{\theta^-}, y \sim p_{\theta}}{\mathbb{E}} [K(x, y)] - \underset{x \sim p_{\theta^+}, y \sim \pi}{\mathbb{E}} [K(x, y)] + \underset{x \sim p_{\theta^-}, y \sim \pi}{\mathbb{E}} [K(x, y)],$$
(5)

其中 $p_{\theta^+}(x)$ 和 $p_{\theta^-}(x)$ 是在参数 $\theta^{\pm} = \theta \pm \frac{\pi}{2} e_l^{\alpha}$ 下的 模型输出概率, e_l^{α} 是参数空间的第 (l,α) 个单位矢量 $(\alpha \pi l 分别表示迭代的步数与当前迭代的第 l 个参$ 数的下标).

对所有量子比特进行测量,会得到计算基上的 概率分布.源数据被映射为经典的比特流数据,与 量子态的计算基一一对应.这种映射是一种同构, 源数据空间和映射空间一样大,因此数据集的数 据必须是有限的.对于 N个数据的集合,需要n~ O(log(N))个量子比特.而生成线路中所用的量子 门个数一般为O(poly(n)).

量子玻恩机是一种隐式生成模型,模型本身并 不涉及经典数据的特征维度,甚至不涉及数据本身 的形态.在量子体系中不需要考虑数据的原本面貌 是什么.无论是图片、数字还是文字,都会被编码 为经典比特流数据.一旦数据同构的方式确定了, 基于量子线路的优化体系就与原本的数据形态无 关了.

利用量子玻恩机进行概率分布的学习和生成 在条带图 (bars and stripes, BAS)^[8,16,26,28] 任务中 取得很好的结果,对于理想的无误差测量,量子玻 恩机生成真实数据分布的比率为 95.7%^[8].此外通 过对量子玻恩机的非计算基测量,可以得到更多的 信息,这种方法可用来生成先验概率分布,如经 典生成对抗网络的先验分布^[31]. Cheng 等^[32] 还从 信息熵的角度比较了量子玻恩机与受限玻尔兹曼 机^[33,34],为理解基于概率的生成模型和基于能量的 生成模型提供了一个新的视角. Coyle 等^[35] 提出 一种量子玻恩机的子类——量子伊辛玻恩机—— 并引入量子核函数^[36] 来讨论量子学习优势.

2.3 量子玻尔兹曼机

与前述模型不同,量子玻尔兹曼机^[9-12] (quantum Boltzmann machine, QBM) 需要一种专用的 量子计算机——量子退火机.

经典的玻尔兹曼机被广泛用作机器学习的基础工具,例如深度信念网络^[37,38]和深度玻尔兹曼机^[37,39].玻尔兹曼机是二进制数节点构成的概率生成网络,由可见节点和隐藏节点组成,记为 z_a , $a = 1,2,\cdots,N$,其中N是总节点数.这里为了和后面的量子生成模型保持一致,取 $z_a \in \{+1,-1\}$ 而不是 $z_a \in \{+1,0\}$.为区别可见节点和隐藏节点,我们分

别记为 z_{ν}, z_i .同时用向量符号v, h, z = (v, h)表示可见态、隐藏态和组合态.在物理上其能量可以用 伊辛模型表示:

$$E_{\boldsymbol{z}} = -\sum_{a} b_{a} z_{a} - \sum_{a,b} w_{ab} z_{a} z_{b}, \qquad (6)$$

其中*ba*,*wab*是在训练的过程中进行需要优化的参数. 定义负对数相似度损失函数:

$$\mathcal{L} = -\sum_{\boldsymbol{v}} P_{\boldsymbol{v}}^{\text{data}} \log P_{\boldsymbol{v}} = -\sum_{\boldsymbol{v}} P_{\boldsymbol{v}}^{\text{data}} \log \frac{\sum_{\boldsymbol{h}} e^{-E_{\boldsymbol{z}}}}{\sum_{\boldsymbol{z}'} e^{-E_{\boldsymbol{z}'}}},$$
(7)

其中观测到可见态*h*的概率为边际分布*Pv* = $Z^{-1}\sum_{h} e^{-E_z}$, $Z = \sum_{z} e^{-E_z}$. 那么玻尔兹曼机的 任务就是优化哈密顿量的参量 $\theta \in \{b_a, w_{ab}\}$, 使边 际分布尽可能接近真实的数据分布*P*_v^{data}. 这样就 可以用梯度的方法来优化玻尔兹曼机:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{\boldsymbol{v}} P_{\boldsymbol{v}}^{\text{data}} \frac{\sum_{\boldsymbol{h}} \partial_{\theta} E_{\boldsymbol{z}} e^{-E_{\boldsymbol{z}}}}{\sum_{\boldsymbol{h}} e^{-E_{\boldsymbol{z}}}} - \frac{\sum_{\boldsymbol{z}} \partial_{\theta} E_{\boldsymbol{z}} e^{-E_{\boldsymbol{z}}}}{\sum_{\boldsymbol{z}} e^{-E_{\boldsymbol{z}}}}, \quad (8)$$
$$\frac{\partial \mathcal{L}}{\partial \theta} = \overline{\langle \partial_{\theta} E_{\boldsymbol{z}} \rangle_{\boldsymbol{v}}} - \langle \partial_{\theta} E_{\boldsymbol{z}} \rangle, \quad (9)$$

其中 $\langle \cdots \rangle$ 和 $\langle \cdots \rangle_v$ 分别为对自由变量的玻尔兹曼 均值以及对可见变量的玻尔兹曼均值. $\overline{\langle \cdots \rangle_v} \equiv \sum_v P_v^{\text{data}} \langle \cdots \rangle_v$ 是双重均值. 这样我们就得到更新 法则:

$$\delta b_a = \eta(\overline{\langle z_a \rangle_{\boldsymbol{v}}} - \langle z_a \rangle), \tag{10}$$

$$\delta w_{ab} = \eta (\overline{\langle z_a z_b \rangle_{\boldsymbol{v}}} - \langle z_a z_b \rangle). \tag{11}$$

关于如何估算等式右边的两项均值,在机器学习领域已经有较多的方案,其中一个简化的玻尔兹曼机模型——称之为受限玻尔兹曼机 (restricted Boltzmann machine, RBM)——已经成功的应用于实例^[40].

量子玻尔兹曼机的出发点是使用作为物理实体的量子比特来替代经典的比特或节点. 然而不同的是,量子力学体系的哈密顿量是 2^N × 2^N 维的矩阵. 而且量子力学算符大多是不对易的,这对于求解梯度十分不便. 横场伊辛模型的哈密顿量形式为:

$$H = -\sum_{a} \Gamma_{a} \boldsymbol{\sigma}_{a}^{x} - \sum_{a} b_{a} \boldsymbol{\sigma}_{a}^{z} - \sum_{a,b} w_{ab} \boldsymbol{\sigma}_{a}^{z} \boldsymbol{\sigma}_{b}^{z}, \quad (12)$$

其中 *I*, σ^z , σ^x 为泡利矩阵, 下标 *a* 表示作用于第 *a* 个量子比特. 与经典玻尔兹曼机类似, 定义配

分函数和密度矩阵分别为: $Z = \text{Tr}[e^{-H}], \rho = Z^{-1}e^{-H}.$ 对于给定的可见态 $|v\rangle$,可定义边际分布 $P_v = \text{Tr}[\Lambda_v \rho], \Lambda_v = |v\rangle \langle v| \otimes \mathcal{I}_h.$ 其中 \mathcal{I}_h 为作用 在隐藏变量的单位阵.

那么类似地,可以得到量子损失函数和对应梯 度为:

$$\mathcal{L} = -\sum_{\boldsymbol{v}} P_{\boldsymbol{v}}^{\text{data}} \log \frac{\text{Tr}[\Lambda_{\boldsymbol{v}} e^{-H}]}{\text{Tr}[e^{-H}]}, \quad (13)$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{\boldsymbol{v}} P_{\boldsymbol{v}}^{\text{data}} \left(\frac{\text{Tr}[\Lambda_{\boldsymbol{v}} \partial_{\theta} e^{-H}]}{\text{Tr}[\Lambda_{\boldsymbol{v}} e^{-H}]} - \frac{Tr[\partial_{\theta} e^{-H}]}{Tr[e^{-H}]} \right).$$
(14)

然而 H和 $\partial_{\theta} H$ 并不对易,也就是 $\partial_{\theta} e^{-H} \neq -e^{-H} \partial_{\theta} H$.这样在求解期望值的时候就会十分困难.尽管 Amin^[9] 通过分析的方法估算除了损失函数的界,在一些特定的条件下可以估算损失函数的梯度,但与前述量子生成模型相比,参数更新的成本已然增加了.

尽管如此,量子玻尔兹曼机仍然可以用来完成 监督学习任务,可以生成概率分布,可以进行鉴别 学习. Amin 等^[9]用量子玻尔兹曼机实现了伯努利 分布的概率生成. Kieferová^[12]用量子玻尔兹曼机 设计了一种生成模型用以实现部分的量子态层析, 并且除非 *BQP* = *BPP*,经典计算机无法有效地模 拟其训练过程. Dorband^[41]在 D-wave 量子退火机 上演示了量子玻尔兹曼机的构建过程. 但是,相较 于其他量子生成模型,量子玻尔兹曼机带来的近似 误差是难以避免的,同时也难以说明相较于经典版 本的优势所在.

2.4 量子自编码器

在经典的数据处理中, 通过机器学习中的自编 码器可以将数据从一个高维空间缩减到低维空间. 经典的自编码模型可以概括为: 通过参数化的神经 网络来训练一个数据集, 其数据维度为(*n*+*k*), 得 到擦除 *k*个特征或维度的新数据, 如果这些数据可 以通过解码器 (神经网络)重新生成接近于原数据 的一组新数据, 那么留下的 *n* 维数据被称为压缩数 据. 这个过程被称为数据压缩.数据压缩对于处理 稀疏特征数据集或高维数据集是十分直观和有效 的手段. 受此启发, 量子自编码器^[13–15](quantum autoencoder) 被设计为量子态压缩的工具.

Romero 等^[13] 提出的量子自编码器与经典类 似: 记 $\{p_i, |\psi_i\rangle_{AB}$ 是 (n+k) 个量子比特上的纯态 集合,子系统 A 和子系统 B 分别占有 n 和 k 个比特. 记 U^{θ} 为作用在 (n + k) 个比特上的幺正算符, 其中 $\theta = \{\theta_1, \theta_2, \cdots\}$ 是参数集合,以及 $|a\rangle_{B'}$ 为 k 比特寄存器的参考纯态. 通过经典的优化方式,我们希望平均保真度函数取得最大值:

$$C_1(\boldsymbol{\theta}) = \sum_i p_i \cdot F(|\psi_i\rangle, \rho_{i,\boldsymbol{\theta}}^{\text{out}}), \qquad (15)$$

其中保真度定义为 $F(|\psi_i\rangle, \rho_i^{\text{out}}) = \langle \psi_i | \rho_i^{\text{out}} | \psi_i \rangle, \rho_{i,\theta}^{\text{out}}$ 是线路经过解码后的密度矩阵:

量子自编码器的核心在于,如果想要有效地压 缩量子态,必须保证态 $|\psi'_i\rangle = U^{\theta}|\psi_i\rangle_{AB}$ 是A,B两 个子系统的直积态.即当下式成立时,才有 $C_1 = 1$:

$$U^{\vec{\theta}}|\psi_i\rangle_{AB} = |\psi_i^c\rangle_A \otimes |a\rangle_B, \qquad (17)$$

其中 $|\psi_i^c\rangle_A$ 是 $|\psi_i\rangle$ 的压缩后的态.

对输出态进行量子态层析以确定其密度矩阵 是非常困难的.可以证明测量 B 和 B' 寄存器的态 保真度等价于 AB 寄存器初态和末态的保真度 (我 们称 B 寄存器输出态为冗态).这样,可以定义另 一种平均保真度 C₂(**θ**),

$$C_{2}(\boldsymbol{\theta}) = \sum_{i} p_{i} \cdot F(\operatorname{Tr}_{A}[U^{\boldsymbol{\theta}}|\psi_{i}\rangle_{AB} \langle\psi_{i}|_{AB} (U^{\boldsymbol{\theta}})^{\dagger}], |a\rangle_{B}).$$
(18)

实际上求两个态的保真度,有一种 SWAP 测试线路可以完成^[42].

通过参数量子线路来构建作用于 AB 系统的 幺正变换,并通过 SWAP 测试线路得到保真度信 息.取损失函数为 $\mathcal{L} = -\log(1 - C_2)$,采用经典方 式进行优化,可以得到保真度最高的参数线路,这 样就实现了数据压缩.

如图 2 所示,量子态压缩模型可以由以下步骤 来构建:第一,制备初态|ψ_i〉和参考态|a〉,并假设制 备是有效的;第二,构建参数线路,初态在幺正算 符 U下演化;第三,通过 SWAP 测试得到冗态和 参考态之间的保真度.

量子自编码器在氢分子态的应用是一个成功 的例证^[13].其方法是用 4 qubits 来模拟氢分子,通 过变分量子求解器 (VQE)求出对于固定原子间距 的量子态,以此作为量子自编码器的输入态.氢分 子的态并非占满了 4 bit 的希尔伯特空间,这就为 量子态压缩提供了可能.直观上看,量子态压缩的 过程就像在整个希尔伯特空间中将若干子空间碎 片像拼图一样拼接成紧密的空间.量子自编码器在 量子加法器方面也有一定的应用^[14],这种加法器 用以实现将两个相同比特数的量子子系统编码到 一个新的子空间.Ding 等^[43]借助超导量子体系在 实验上验证了量子加法器的可行性,实验保真度与 理论值一致.



图 2 量子自编码器的训练示意图,其中p表示变量,摘自^[13] Fig. 2. Schematic representation of the hybrid scheme for training a quantum autoencoder where p represents variables, image from^[13].

Pepper 等^[44] 采用了 Romero 等^[13] 提出的量 子自编码器框架,不同的是,其数据压缩任务是 将多维量子比特 (qutrit, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle + \gamma|2\rangle$ (已归一化))压缩为二维量子比特 (qubit)态.通 过光量子实验体系,Pepper 等在实验上实现了 qutrits 的压缩,并且证明了这样的实验优化系统 对扰动是鲁棒的.Bondarenko 等^[45] 注意到经典数 据通过经典自编码器可以有效地去除噪声,因此量 子自编码器同样有类似的作用.通过构建以自旋翻 转噪声和任意幺正门噪声为主的 GHZ 含噪声量子 态,Bondarenko 等验证了量子自编码器在量子数 据去噪方面的作用.

如何判断量子态能否通过量子自编码器有效 压缩一直受到关注, Huang 等^[46] 在理论上论证了 实现有效量子自编码态压缩的条件, 并证明了若输 入数据的最大线性无关向量的个数不超过隐变量 空间 (压缩后的低维空间)的维度, 那么量子自编 码器就可以无损地将高维数据压缩至低维. Huang 等利用量子光学体系在实验上验证了其理论的结 果. 此外 Cao 等^[47] 提出可在量子退火机上运行、 通过噪声辅助的绝热模型, 同样可以完成量子自编 码任务. 同时, 对量子态压缩的探索也使得量子梯 度消失的现象——也即贫瘠高原的现象——得到 进一步的研究^[48].

2.5 其他生成模型

从前述量子生成模型可以看出, 经典机器学习 给我们带来了很多启发. 类似地, 经典领域中因子 图 (factor graph) 可以作为概率生成模型, Gao 等^[49] 提出了一种量子生成模型 (quantum generative model, QGM), 并证明此模型可以高效地表示任意 因子图. 并且通过理论证明至少存在一些特定的概 率分布, 可以被 QGM 有效表示, 但不能被任意具 有多项式量级变量的因子图有效表示. 如果多项式 层次作为 P vs. NP 问题的泛化是不塌缩的话, QGM 相对于经典的因子图具有指数优势.

同时,除了 QGM 的表示能力和泛化能力, Gao 等^[49]还提出了一种通用的量子学习算法,该 算法利用张量网络、对多体纠缠量子态的父哈密顿 量 (parent Hamiltonian)构建、以及递归的量子相 位估计进行学习.显然希望这种算法对任意问题都 能在多项式时间内解决是不现实的,但作者证明了 至少存在一些问题,QGM 相对于任意经典算法均 有指数加速的效果,前提是假设量子计算机不能被 经典计算机有效地模拟.

直观地理解 QGM 中的指数加速效果:机器学 习中的生成模型旨在通过寻找潜在的概率分布,对 自然界任意的数据生成过程进行建模.由于自然界 受量子力学定律支配,用经典生成模型中的概率分 布对真实世界中的数据进行建模具有很大的局限 性.Gao 等^[49]提出的 QGM 模型,用一个多体纠缠 量子态的概率幅对数据的相关性进行参数化.因为 量子概率幅的相干性产生的现象比经典的复杂得 多,QGM 才有可能在特定情况下实现性能的显著 提升.QGM 使用图态 |G>表示,由 m个量子比特 与图 G 构成.如图 3 所示.

Gao^[50]有另一篇工作同样旨在发掘量子相关 性在生成模型中的作用.其中比较了经典贝叶斯网 络顺序模型与对应的一维张量网络,展示了矩阵直 积态 (matrix product states, MPSs)比经典机器 学习模型有更强的表示能力.因为一维模型可以在 经典计算机上被有效模拟,Gao等采用数值模拟的 方式测试了真实数据集的生成,发现 MPS 生成能 力比经典模型有提升.此结果为理解基于 MPS 的 机器学习算法提供了新的视角,因为有一类张量网 络子类不能有效地被经典计算机模拟,但是可以在 量子计算机上构建,这就使得在机器学习中的生成 模型框架下量子优势是可能存在的.



图 3 经典和量子生成模型 (a) 因子图表示; (b) 张量网络态表示; (c) 量子生成模型定义^[49] Fig. 3. Classical and quantum generative models: (a) Illustration of a factor graph; (b) illustration of a tensor network state; (c) QGM definition^[49].

除以上的量子生成模型外,还有基于贝叶斯理 论提出量子贝叶斯线路^[16].这种量子线路引入辅 助比特表示先验概率分布,通过受控操作传递给数 据比特寄存器.这样可以极大地规避参数量子线路 中出现的生成概率不均匀的问题.

3 量子生成模型展望

随着量子计算机硬件的不断发展,量子生成模型也不只在量子虚拟机上得到实现,基于真实量子芯片的生成模型受到了广泛的关注,但是量子生成模型存在的问题仍待解决.

首先,量子生成模型是否可以生成任意概率分 布的数据?对于连续变量的数据,量子生成模型还 无法像经典神经网络一样有效地生成.同时,数据 是量子还是经典的决定了生成模型的选取.对于经 典数据,利用 MPQCs 或 MPSs 生成量子态,对量 子态的测量得到了量子信息的经典统计结果.如何 更有效地生成量子态以及如何从量子态读取更多 信息一直受到关注,不同的量子生成模型给出了不 同的策略,然而对于任意连续变量数据的生成还需 要进一步的探索.

其次,量子生成模型作为量子机器学习的一部 分,如何体现量子计算的优势,仍然是一个开放性 问题.对量子线路的表示能力的研究为我们提供了 一个思路,但是对于不同的生成模型很难给出统一 的表示能力定义.而 Gao 等^[49,50]从量子相关性出 发,直观上将量子内在相关性的参数化来实现经典 生成模型无法模拟的任务,说明了量子优势在机器 学习领域中的存在性.

对于量子生成模型,还需要进一步探索.在量 子领域,量子生成模型提供了 NISQ 体系下可执行 的算法,在机器学习领域,它引入了量子计算和量 子信息的理念,使机器学习算法突破经典计算的 瓶颈成为可能.

感谢中国科学技术大学薛程博士和赵健博士的讨论.

附录 A 参数量子线路

量子计算机遵循量子力学定律进行逻辑运算和信息存 取. 而量子力学体系区别于经典体系之处在于,由于叠加态 和纠缠态的存在,相同位数的量子数据比经典数据有着指 数强的表示能力. 将量子信息转化为经典信息需要对量子 态进行测量.

量子计算的基本计算单位是量子比特 (qubit), 单比特 量子叠加态记为 $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$ (已归一化).根据玻恩的 几率解释, 对单比特量子态进行测量,得到 0 态的概率为 $|\alpha|^2$,得到 1 态的概率为 $|1\rangle$.对于多量子比特系统,根据量 子力学体系的完备性,量子态可在计算基上展开,由 $|\Phi\rangle = \sum_i a_i |i\rangle$ 测得任意态 $|i\rangle$ 的概率为 $|a_i|^2$.对于泡利算符 σ_i , $i \in \{0, x, y, z\}$,其期望值可以通过算符的谱来计算,即 $\langle \Phi | \sigma_i | \Phi \rangle = \left\langle \Phi | \sum_j \lambda_j | \psi_j \rangle \langle \psi_j | | \Phi \right\rangle$,其中 λ_j 为本征值, $|\psi_j \rangle$ 为本征态.对于一个更复杂的哈密顿量 *H*,我们总能将其分 解为若干泡利算符的直积,再分别求出各个部分的期望值, 即 $H = \sum_{i,j,k,\dots,m} \alpha_{i,j,k,\dots,m} \cdot \sigma_i \otimes \sigma_j \otimes \sigma_k \dots \otimes \sigma_m$, $i, j, k \dots$

 $m \in \{0, x, y, z\}, \alpha_{i, j, k, \dots m}$ 为泡利基的实系数.

通过测量量子态在计算基上的投影或哈密顿量的期望 值得到的经典数据,其处理策略也根据量子生成模型的不 同而有各式的差异.如对于量子玻恩机,量子测量得到的数 据被映射为数据集的概率分布,显然这样对于小型数据集 或稀疏数据集有表示优势.而对于量子生成对抗网络,测量 得到的数据可直接作为生成数据以混淆源数据.

量子生成模型的核心在于利用量子计算机生成量子态, 对量子态进行测量得到经典的有效信息.如何有效地生成 量子态,是我们面临的首要问题.参数量子线路 (PQC) 为 这一问题提供了一个思路.

参数量子线路一般由若干单量子门和多量子纠缠门组 合得到,单量子门可以是任意单比特幺正门 $U_S(\theta)$,常用的 有泡利旋转门 $R_X(\theta) = e^{-i\frac{\theta}{2}\sigma_x}, R_Y(\theta) = e^{-i\frac{\theta}{2}\sigma_y}, R_Z(\theta) =$ $e^{-i\frac{\theta}{2}\sigma_z}$, 多量子门一般采用 $U_{CZ} = |0\rangle\langle 0|\otimes I + |1\rangle\langle 1|\otimes\sigma_z$, $U_{CNOT} = |0\rangle\langle 0|\otimes I + |1\rangle\langle 1|\otimes\sigma_x$ 等两比特门. 单量子门和多 量子门交替的结构极大地提高了态的纠缠能力, 扩展了态的表示空间.

记量子线路的参数为 θ ,则参数量子线路可以用 $U(\theta)$ 表示.若要保证参数量子线路是有效的,其参数总量和线路 深度都有相应的限制.一般设定参数的总量是多项式级的,即 $n \sim O(\text{poly}(N))$,其中n为参数个数,N为量子比特个数.同时线路的深度也设为多项式级的,即 $L \sim O(\text{poly}(N))$.

如何使参数量子线路有更强的表示能力一直受到关注. 最直观的方式是增加参数量子线路的深度,构成 MPQCs (multiple parameterized quantum circuits). 然而 MPQCs 的线路深度至多为多项式级的才能满足有效性要求,也即 图 A1 中的 PQC 模块数是常数级的,不随量子比特数目增 长.而且由于量子噪声的存在,增加 PQC 的模块数会极大 地影响量子态的保真度,所以在真实的任务中往往会限 制 PQC 的模块数.除此之外还可以通过非计算基补充测量 的方式,来获取更多的数据信息,这种方法在 QCAAN 线 路^[31] 中被提到.



Fig. A1. Illustration of MPQCs.

对经典神经网络的表示能力的研究是一项非常重要的 课题,参数量子线路的表示能力直接关系到其能否体现量 子优越性. Du 等^[22]证明了在多项式层次结构保持的前提 下,具有 O(poly(N))个门的参数量子线路,其表示能力比具 有 O(poly(N))个参数的经典玻尔兹曼机要强. 尽管如此,参 数量子线路和经典神经网络的直接比较还有一些困难,如 何定义对量子和经典体系均适用的表示能力指标,以及参 数量子线路能否体现量子的优越性,仍是开放性的问题.

参考文献

- Zhu J Y, Krähenbühl P, Shechtman E, Efros A A 2016 European Conference on Computer Vision, Berlin, September 16, 2016 p597
- [2] Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K 2016 arXiv: 1609.03499 [cs.SD]
- [3] Gómez-Bombarelli R, Wei J N, Duvenaud D, Hernández-

Lobato J M, Sánchez-Lengeling B, Sheberla D, Aguilera-Iparraguirre J, Hirzel T D, Adams R P, Aspuru-Guzik A 2018 *ACS Cent. Sci.* **4** 268

- [4] Isola P, Zhu J Y, Zhou T, Efros A A 2016 arXiv: 1611.07004[cs.CV]
- [5] Dallaire-Demers P L, Killoran N 2018 Phys. Rev. A 98 012324
- [6] Lloyd S, Weedbrook C 2018 Phys. Rev. Lett. **121** 040502
- [7] Benedetti M, Garcia-Pintos D, Perdomo O, Leyton-Ortega V, Nam Y, Perdomo-Ortiz A 2019 npj Quantum Inf. 5 1
- [8]~ Liu J G, Wang L 2018 Phys. Rev. A $\mathbf{98}$ 062324
- [9] Amin M H, Andriyash E, Rolfe J, Kulchytskyy B, Melko R 2018 Phys. Rev. X 8 021050
- [10] Khoshaman A, Vinci W, Denis B, Andriyash E, Amin M H 2019 Quantum Sci. Technol. 4 014001
- [11] Benedetti M, Realpe-Gómez J, Biswas R, PerdomoOrtiz A 2017 Phys. Rev. X 7 041052
- [12]~ Kieferová M, Wiebe N 2017 Phys. Rev. A $\mathbf{96}$ 062327
- [13] Romero J, Olson J P, Aspuru-Guzik A 2017 Quantum Sci. Technol. 2 045001
- [14] Lamata L, Alvarez-Rodriguez U, Martn-Guerrero J, Sanz M, Solano E 2018 Quantum Sci. Technol. 4 014007
- [15] Li R, Alvarez-Rodriguez U, Lamata L, Solano E 2017 Quantum Meas. Quantum Metrol. 4 1
- [16]~ Du Y, Liu T, Tao D 2018 arXiv: 1805.11089 [quant-ph]
- [17] Peruzzo A, McClean J, Shadbolt P, Yung M H, Zhou Z Q, Love P J, Aspuru-Guzik A, O'Brien J L 2014 Nat. Commun. 5 4213
- [18] Nielsen M A, Chuang I L 2002 Quantum computation and quantum information (Cambridge: Cambridge University Press) pp221–225
- [19] Preskill J 2018 Quantum 2 79
- [20] Harrow A W, Hassidim A, Lloyd S 2009 Phys. Rev. Lett. 103 150502
- [21] Huang H L, Du Y, Gong M, Zhao Y, Wu Y, Wang C, Li S, Liang F, Lin J, Xu Y, Yang R, Liu T, Hsieh M H, Deng H, Rong H, Peng C Z, Lu C Y, Chen Y A, Tao D, Zhu X, Pan J W 2020 arXiv: 2010.06201 [quant-ph]
- [22] Du Y, Hsieh M-H, Liu T, Tao D 2020 Phys. Rev. Research 2 033125
- [23] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, WardeFarley D, Ozair S, Courville A, Bengio Y 2014 Proceedings of the 27th International Conference on Neural Information Processing Systems 2 pp2672–2680
- [24] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A C 2017 arXiv: 1704.00028 [cs.LG]
- [25] Zoufal C, Lucchi A, Woerner S 2019 npj Quantum Inf. 5 103
- [26] Zeng J, Wu Y, Liu J G, Wang L, Hu J 2019 Phys. Rev. A 99 052306
- [27] Schuld M, Bergholm V, Gogolin C, Izaac J, Killoran N 2019 Phys. Rev. A 99 032331
- [28] MacKay D J C 2002 Information Theory, Inference & Learning Algorithms (Cambridge: Cambridge University)
- [29] Situ H, He Z, Wang Y, Li L, Zheng S 2020 Information Sciences 538 193
- [30] Hu L, Wu S H, Cai W, Ma Y, Mu X, Xu Y, Wang H, Song Y, Deng D L, Zou C L, Sun L 2019 Sci. Adv. 5 eaav2761
- [31] Rudolph M S, Toussaint N B, Katabarwa A, Johri S, Peropadre B, Perdomo-Ortiz A 2020 arXiv: 2012.03924 v2 [quant-ph]
- [32] Cheng S, Chen J, Wang L 2018 Entropy 20 583
- [33] Hinton G E, Sejnowski T J 1986 In Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1 282

- [34] Hinton G E 2012 In Neural Networks: Tricks of the Trade Berlin Heidelberg, Germany, 2012 p599
- [35] Coyle B, Mills D, Danos V, Kashefi E 2020 npj Quantum Inf.6 60
- [36] Hofmann T, Schölkopf B, Smola A J 2008 Ann. Statist. 36 1171
- [37] Hinton G E, Sejnowski T J 1983 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Washington D. C., USA, 1983 p448
- [38] Hinton G E, Osindero S, Teh Y-W 2006 Neural Comput. 18 1527
- [39] Salakhutdinov R, Hinton G 2009 Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics Florida, USA 2009 p448
- [40] Smolensky P 1986 Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol.1) (Cambridge: MIT press) pp194–281
- [41] Dorband J E 2015 12th International Conference on

Information Technology-New Generations Las Vegas, USA, April 13–15, 2015 p703

- [42] Buhrman H, Cleve R, Watrous J, Wolf R D 2001 Phys. Rev. Lett. 87 167902
- [43] Ding Y, Lamata L, Sanz M, Chen X, Solano E 2019 Adv. Quantum Technol. 2 1800065
- [44] Pepper A, Tischler N, Pryde G J 2019 Phys. Rev. Lett. 122 060501
- [45] Bondarenko D, Feldmann P 2020 Phys. Rev. Lett. 124 130502
- [46] Huang C J, Ma H, Yin Q, Tang J F, Dong D, Chen C, Xiang G Y, Li C F, Guo G C 2020 *Phys. Rev. A* 102 032412
- [47] Cao C, Wang X 2021 Phys. Rev. Applied. 15 054012
- [48] Cerezo M, Sone A, Volkoff T, Patrick L C, Coles J 2021 Nat. Commun. 12 1791
- [49] Gao X, Zhang Z Y, Duan L-M 2018 Sci. Adv. 4 eaat9004
- [50] Gao X, Anschuetz E R, Wang S T, Cirac J I, Lukin M D 2021 arXiv: 2101.08354 v1 [quant-ph]

SPECIAL TOPIC—Machine learning and physics

Quantum generative models for data generation^{*}

Sun Tai-Ping¹⁾ Wu Yu-Chun^{1)2)†} Guo Guo-Ping¹⁾²⁾³⁾

1) (CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China)

2) (Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China)

3) (Origin Quantum Computing Company Limited, Hefei 230026, China)

(Received 17 May 2021; revised manuscript received 30 June 2021)

Abstract

In recent years, many generation-based machine learning algorithms such as generative adversarial networks, Boltzmann machine, auto-encoder, etc. are widely used in data generation and probability distribution simulation. On the other hand, the combined algorithms of quantum computation and classical machine learning algorithms are proposed in various styles. Especially, there exist many relevant researches about quantum generative models, which are regarded as the branch of quantum machine learning. Quantum generative models are hybrid quantum-classical algorithms, in which parameterized quantum circuits are introduced to obtain the cost function of the task as well as its gradient, and then classical optimization algorithms are used to find the optima. Compared with its classical counterpart, quantum generative models map the data stream to high-dimensional Hilbert space with parameterized quantum circuits. In the mapping space, data features are easier to learn, which can surpass classical generative models in some tasks. Besides, quantum generative models are potential to realize the quantum advantage in noisy intermediate-scale quantum devices.

Keywords: quantum machine learning, quantum generative models, quantum advantage

PACS: 03.67.-a, 03.67.Ac

DOI: 10.7498/aps.70.20210930

^{*} Project supported by the National Key Research and Development Program of China (Grant No. 2016YFA0301700), the National Natural Science Foundation of China (Grant No. 11625419), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB24030600), and the Anhui Initiative in Quantum Information Technologies, China (Grant No. AHY080000).

[†] Corresponding author. E-mail: wuyuchun@ustc.edu.cn





Institute of Physics, CAS

基于自旋体系的量子机器学习实验进展

田宇 林子栋 王翔宇 车良宇 鲁大为

Experimental progress of quantum machine learning based on spin systems Tian Yu Lin Zi-Dong Wang Xiang-Yu Che Liang-Yu Lu Da-Wei 引用信息 Citation: Acta Physica Sinica, 70, 140305 (2021) DOI: 10.7498/aps.70.20210684 在线阅读 View online: https://doi.org/10.7498/aps.70.20210684 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

多量子比特核磁共振体系的实验操控技术

Experimental technique for multi-qubit nuclear magnetic resonance system 物理学报. 2017, 66(15): 150302 https://doi.org/10.7498/aps.66.150302

核磁共振量子信息处理研究的新进展

New research progress of nuclear magnetic resonance quantum information processing 物理学报. 2018, 67(22): 220301 https://doi.org/10.7498/aps.67.20180754

固态金刚石氮空位色心光学调控优化

Optimization of optical control of nitrogen vacancy centers in solid diamond 物理学报. 2020, 69(14): 147601 https://doi.org/10.7498/aps.69.20200072

基于金刚石氮-空位色心自旋系综与超导量子电路混合系统的量子节点纠缠

Entanglement of quantum node based on hybrid system of diamond nitrogen-vacancy center spin ensembles and superconducting quantum circuits

物理学报. 2018, 67(7): 070302 https://doi.org/10.7498/aps.67.20172634

基于金刚石体系中氮-空位色心的固态量子传感

Solid quantum sensor based on nitrogen-vacancy center in diamond 物理学报. 2018, 67(16): 160301 https://doi.org/10.7498/aps.67.20180788

基于金刚石氮-空位色心的精密磁测量

High-resolution magnetometry based on nitrogen-vacancy centers in diamond 物理学报. 2018, 67(16): 167601 https://doi.org/10.7498/aps.67.20181084

专题: 机器学习与物理

基于自旋体系的量子机器学习实验进展*

田宇 林子栋 王翔宇 车良宇 鲁大为†

(南方科技大学物理系, 深圳 518055)

(2021年4月12日收到; 2021年5月25日收到修改稿)

机器学习因其在模式识别等问题上的优势已经被广泛应用到各个研究领域,然而其运算能力在一定程度上受到经典计算机算力的制约.近年来,随着量子技术的高速发展,量子计算加速的机器学习在诸多量子体系中进行了初步实验验证,并在某些特定问题上展示出了超越经典算法的优势.本文主要介绍两类典型的自旋体系——核磁共振体系和金刚石氮空位色心体系,并回顾近年来量子机器学习在这两类体系上的一些代表性实验工作.

关键词:量子机器学习,自旋体系,核磁共振,氮空位色心 **PACS**: 03.65.-w, 03.67.Ac, 03.67.Lx

DOI: 10.7498/aps.70.20210684

1 引 言

机器学习是人工智能的重要研究方向之一,该 领域主要采用的方法是对大量数据进行分析,以识 别出数据中包含的信息,进而提取出数据特征^[1]. 它起源于计算机科学,在近十年间备受瞩目,在科 学技术等诸多领域产生了广泛的影响^[2-4].机器学 习从基于学习方式上大致可以分为三大类:监督学 习,无监督学习,强化学习.监督学习是通过已有 的带标签的数据调整数学模型参数,从而让模型能 够解决分类或者回归问题,监督学习算法包含支持 向量机 (SVM)、线性回归、决策树及神经网络.无 监督学习则用来对没有标签信息的数据进行聚类 或者对未标准化的数据进行预处理,常用的算法有 主成分分析 (PCA)、词嵌入及 K 最近邻 (KNN). 强化学习则是通过决策行动得到的环境反馈的奖 惩信息作为模型参数的训练依据,最终使模型能够 解决规划决策和模型优化问题,属于强化学习的算 法有 Q-learning、策略梯度算法、SARSA(stateaction-reward-state-action)算法.上述算法尽管早 在 20世纪就已经被提出,但是由于计算机算力的 限制以及数据的缺乏,该领域一直发展缓慢.直到 2012 年 Hinton 教授和他的学生 Krizhevsky^[5] 设计的 AlexNet 在 ImageNet 竞赛上表现优异,让 学界及工业界开始关注这种利用大数据和高性能 图形处理器 (GPU)训练的复杂神经网络模型的价 值,随后许多优秀的深度学习模型,如 VGGNet^[6], ResNet^[7], DenseNet^[8] 的出现标志着深度学习时代 的来临.

尽管现在以深度学习为代表的机器学习算法 在计算机视觉、自然语言处理、大数据分析等领域 已经有了重大的成就,但是随着模型的扩大,计算 机硬件的算力不足及经费开销已经成为该领域发

© 2021 中国物理学会 Chinese Physical Society

^{*} 国家重点研究发展计划 (批准号: 2019YFA0308100)、国家自然科学基金 (批准号: 12075110, 11975117, 11905099, 11875159, U1801661)、广东基础和应用基础研究基金会 (批准号: 2019A1515011383)、广东省国际合作计划 (批准号: 2020A0505100001)、 深圳市科学技术和创新委员会 (批准号: ZDSYS20170303165926217, KQTD20190929173815000, JCYJ20200109140803865, JCYJ20170412152620376, JCYJ20180302174036418)、鹏城学者、广东省创新研究与计划中心 (批准号: 2019ZT08C044) 和广东 省重点实验室 (批准号: 2019B121203002) 资助的课题.

[†] 通信作者. E-mail: ludw@sustc.edu.cn

展的瓶颈. 例如 2020 年 OpenAI 公布的目前参数 最多的自然语言模型 GPT-3^[9],其训练大致需要 512 块 V100 显卡训练 4 个月,训练成本已经高达 1200 万美元. 现如今,经典计算机再次无法满足机 器学习对算力日益增长的需求,人们迫切地需要发 起一场"计算革命".量子计算机便是这场革命运动 中的排头兵.

量子计算机中的量子态具备纠缠、叠加等特 性,这些特性是经典计算机所不具备的资源,可以 用来实现对于算法的加速,称为量子加速 (quantum speedup)^[10]. 量子计算机的基础运算单元为 量子比特,其状态不似经典比特只有0和1,而是 处在0和1的叠加态上.因此,量子比特所包含的 信息容量远远大于经典比特,而且量子比特之间可 以被纠缠起来,这些特点使得量子计算机在实现对 数据的并行运算处理上具有天然的优势. 换句话 说,在经典计算机上很难解决或无法解决的问题可 能在量子计算机上很容易得到解决 [11]. 目前, 量子 计算已经能够实现对某些特定问题的加速运算,并 在实验上实现了量子计算机发展规划的第二个里程 碑——量子优越性 (quantum supremacy). 2019 年, Google团队基于超导体系构建的量子处理器 "Sycamore"只用了 200 s 就完成了经典计算机需 要 10000 年的采样任务 [12]. 2020 年, 潘建伟等 [13] 构建的光量子计算原型机"九章"在处理高斯玻色 取样问题上的计算速率比经典计算机提高了100万 亿倍.

量子加速实现的关键在于能否找到合适的量 子算法^[14]. 1985年, Deutsch^[15]提出了世界上第 一个量子算法, 该算法充分地利用了量子态的叠加 特性, 只需要通过一次测量就能够判断出目标函数 的性质. 1994年, 用于大数分解的 Shor 算法实现 了质因数分解的指数加速^[16]. 1995年, 无序搜索 的 Grover 算法实现了对于目标态查找的平方加 速^[17,18]. 这些重要算法的提出标志着量子计算的 研究进入了一个全新的阶段. 2009年, Harrow等^[19] 提出了著名的 HHL 算法, 该量子算法对于求解线 性方程组具有指数加速的作用. 因为机器学习中的 数据多以高维矩阵的形式存在, 学习的过程中常伴 随着线性方程组的求解问题, 该算法的提出标志着 量子机器学习作为一个重要研究领域的正式建立.

经历了十多年的发展,量子机器学习已经在各 种实验体系中得到广泛应用.目前,通用型量子计 算机的实现存在着多种技术路线,例如:超导电路体系、囚禁离子体系、光学体系,以及以核磁共振(nuclear magnetic resonance, NMR)和金刚石氮空位(nitrogen vacancy, NV)色心为代表的自旋体系.自旋体系具有控制精度高、相干时间长等优势^[20],在量子计算初期阶段发展出了许多技术手段,包括脉冲编译、最优化控制、动力学解耦等技术^[20],目前也是量子机器学习实验实现的重要平台.

本文将简要介绍核磁共振和金刚石 NV 色心 两种自旋体系的基本原理,并梳理近年来国内外各 研究组在这两类平台上所实现的具有代表性的量 子机器学习的工作.

2 核磁共振体系

核磁共振体系是人们研究最早的量子计算体 系之一.早在1938年, Rabi等^[21]发现了著名的 Rabi 振荡现象:位于磁场中的原子核会沿着磁场方向呈 正向或反向平行排列,在施加射频场之后,这些原 子核的自旋方向则会发生翻转.之后,Bloch^[22]于 1946年发现处于外磁场中的特定核自旋会吸收特 定频率的射频场能量,这是人类对于核磁共振现象 最早的认识.经过五十多年的发展,核磁共振已经 在化学、医疗等领域有了诸多应用^[23-26],并且成 熟的操控技术使人们可以精确操控核磁共振中 耦合起来的两能级量子系统.在量子计算概念被 提出后,核磁共振也作为各个量子计算潜在方案 中操控比特数最多、操控精度最高的方案而被广泛 研究^[20].

核磁共振系统可以用系统哈密顿量和控制哈 密顿量联合进行的动力学演化来描述.系统哈密顿 量给出在静磁场中单个的,或者耦合起来的核自旋 的能量形式;而控制哈密顿量来自核自旋共振的控 制射频场.

在核磁共振谱仪中,样品被放置在沿着 2 方向 的静磁场 B₀中.对于单个自旋为 1/2 的粒子来说, 其在沿着 2 方向的静磁场 B₀中的动力学演化被哈 密顿量主导:

$$\boldsymbol{H}_{0} = -\hbar\gamma B_{0}\boldsymbol{I}_{z} = -\hbar\omega_{0}\boldsymbol{I}_{z} = \begin{pmatrix} -\hbar\omega_{0}/2 & 0\\ 0 & \hbar\omega_{0}/2 \end{pmatrix},$$
(1)

其中, \hbar 是约化普朗克常数, γ 是原子核的旋磁比, $\omega_0/(2\pi)$ 是拉莫尔频率, I_z 是沿着 \hat{z} 方向的核自旋 算符. 在核磁共振领域中, I_x , I_y , I_z 和泡利算符 有着如下对应:

$$I_x = \frac{1}{2}\boldsymbol{\sigma}_x, \ I_y = \frac{1}{2}\boldsymbol{\sigma}_y, \ I_z = \frac{1}{2}\boldsymbol{\sigma}_z,$$
 (2)

并且

$$\boldsymbol{\sigma}_{x} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{\sigma}_{y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$
$$\boldsymbol{\sigma}_{z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$
(3)

这意味着自旋为 1/2 的粒子在外磁场中会发生所 谓的塞曼效应,从而产生两个能级差为 $\hbar\omega_0$ 的本征 态 $|0\rangle = \binom{1}{0}$, $|1\rangle = \binom{0}{1}$.核磁共振量子计算把这两 个本征态作为量子比特的 0 和 1. 当射频场的能量 和能级差匹配时,核自旋将在两个本征态之间进行 跃迁.不同种类的核自旋拥有不同的旋磁比 γ ,从 而拥有不同的拉莫尔进动频率 $\omega_0/(2\pi)$.而相同种 类的原子核一般因为周遭不同的电子云排布,受到 不同程度的磁场屏蔽作用 (化学位移),从而也有不 同的拉莫尔频率.不同的拉莫尔频率意味着不同的 能级差,于是这些核自旋构成的量子比特可以通过 不同的共振频率来加以区分.

当样品的分子中有多个核自旋 (多量子比 特)时,核自旋之间会产生相互作用.相互作用的 种类有两种:直接相互作用 (偶极-偶极耦合)和间 接相互作用 (标量耦合).在常用的液体核磁共振样 品中,由核磁矩产生的直接相互作用会因为液体分 子的快速滚动而被平均掉.只剩下由化学键产生的 间接相互作用,在弱耦合的情况下 ($|\omega_0^i - \omega_0^j| \gg 2\pi J_{ii}$)其哈密顿量的形式为

$$\boldsymbol{H}_J = 2\pi\hbar \sum_{i < j} J_{ij} \boldsymbol{I}_z^i \boldsymbol{I}_z^j, \qquad (4)$$

其中J_{ij}为自旋i和自旋j之间的相互作用强度.

所以大多数核磁共振实验中的系统哈密顿量 可表示为核自旋在静磁场中的哈密顿量和核自旋 间两两相互作用的哈密顿量之和:

$$\boldsymbol{H}_{\text{sys}} = -\hbar \sum_{i} \omega_0^i \boldsymbol{I}_z^i + 2\pi\hbar \sum_{i < j} J_{ij} \boldsymbol{I}_z^i \boldsymbol{I}_z^j.$$
(5)

在核磁谱仪中,控制核自旋跃迁的射频场在多 旋转坐标系下的哈密顿量可以写成

$$\boldsymbol{H}_{\text{control}} = \sum_{i,r} -\hbar\omega_{1}^{\text{r}} (\cos\left[\left(\omega_{\text{rf}}^{\text{r}} - \omega_{0}^{i}\right)t + \phi^{\text{r}}\right] \boldsymbol{I}_{x}^{i} \\ -\sin\left[\left(\omega_{\text{rf}}^{\text{r}} - \omega_{0}^{i}\right)t + \phi^{\text{r}}\right] \boldsymbol{I}_{y}^{i}), \tag{6}$$

其中ω_{ff}为旋转坐标系的频率 (与射频场旋转频率 一致), φ^r为射频场的相位.

所以核磁共振体系的哈密顿量可以表示为

$$\boldsymbol{H} = \boldsymbol{H}_{\text{sys}} + \boldsymbol{H}_{\text{control}}.$$
 (7)

基于对哈密顿量的控制,核磁共振系统可以通过赝 纯态^[27,28]的制备实现系统的初始化;通过调整射 频场参数和核自旋间相互作用时间构建通用逻辑 门;从核磁样品系综的自由感应衰减信号中重构出 量子态.

从目前的实验进展来看,核磁共振量子计算已 经非常成熟,这为那些较为复杂的量子算法提供了 一个很好的演示平台.目前量子算法在解决线性代 数计算问题上已经展现了初步的加速能力,而例如 "傅里叶变换"、"求本征值本征态"、"解线性方程" 的问题都是经典机器学习算法中常见的子程序.接 下来,本文将回顾国内外量子机器学习算法基于核 磁共振体系的实验实现.

2.1 解线性方程组

解线性方程组几乎是所有科研、工程领域都会 面临的问题,而经典算法解线性方程组的复杂度限 制着计算问题的尺度.2009年,Harrow等^[19]提出 了一个能够在解线性方程组上实现指数加速的 HHL量子算法.它不仅为解线性方程组提供了加 速方案,也是许多量子机器学习算法的基础.HHL 算法的核心是用量子计算机进行矩阵求逆,从而求 解线性方程组.一个线性方程组可以写成

$$\boldsymbol{A} \left| \boldsymbol{x} \right\rangle = \left| \boldsymbol{b} \right\rangle,$$
 (8)

其中 $N \times N$ 的参数矩阵可以用它自己的本征值 λ_j 和本征态 $|u_j\rangle$ 展开 $\mathbf{A} = \sum_j \lambda_j |u_j\rangle \langle u_j|$,向量也可以展开为 $|\mathbf{b}\rangle = \sum_j \beta_j |u_j\rangle$.参数矩阵的逆矩阵可以表示为 $\mathbf{A}^{-1} = \sum_j \lambda_j^{-1} |u_j\rangle \langle u_j|$,于是线性方程组的解

$$|\boldsymbol{x}\rangle \propto \boldsymbol{A}^{-1} |\boldsymbol{b}\rangle \propto \sum_{j} \left(\beta_{j}/\lambda_{j}\right) |\boldsymbol{u}_{j}\rangle.$$
 (9)

在量子线路中可以通过相位估计算法^[29]得到A的本征值 λ_j ,然后对辅助比特旋转2arcsin(C/λ_j)角度,再做一个相位估计的逆过程即可得到

$$\sum_{j} \left(\sqrt{1 - \frac{C^2}{\lambda_j^2}} \left| 0 \right\rangle + \frac{C}{\lambda_j} \left| 1 \right\rangle \right) \beta_j \left| \boldsymbol{\lambda}_j \right\rangle \left| \boldsymbol{u}_j \right\rangle, \quad (10)$$

其中C是一个常数. 当辅助比特投影到|1)态上时,

量子态坍缩为 $\sum_{j} C(\beta_{j}/\lambda_{j}) |u_{j}\rangle \propto |x\rangle$. 在已知最 好的经典算法中需要 $O(N\log N)$ 的时间复杂度来 求得解向量 $|x\rangle$. 上述量子算法只需要 $O[(\log N)^{2}]$ 的时间复杂度就可以做到同样的事. 尽管在当前技 术下,量子态制备和量子态读出在时间复杂度上仍 然存在需要解决的难题, HHL 算法在解决特定问 题上仍有指数加速的潜力.

实验方面,光学、核磁共振、超导等量子计算 平台都相继实现了 HHL 算法^[30-32],如图 1 所示. 其中由中国科学技术大学杜江峰团队完成的工 作^[31] 是在 4 bit的核磁共振系统实现了基于量子逻 辑门的 HHL 算法.算法完成了解 2 × 2线性方程组 的任务,并重复对 3 个不同的向量 |b)进行验证 (相 同的矩阵 *A*).实验用氘代氯仿作为样品,其中的 ¹³C和 3 个¹⁹F核自旋作为 4 个量子比特.由 2 bit 组成的寄存器用来实现相位估计算法,由单比特充 当的寄存器用来储存向量 |b),剩余的 1 bit 作为辅 助比特进行读出前的投影测量.对于不同输入的 |b), 4 bit 的末态保真度均高于 96%.实验以优良的准 确性验证了 HHL 算法的可行性.

HHL 算法及其实验实现均基于逻辑门模型. 受限于当前技术下量子比特的数量,基于逻辑门的 复杂量子算法通常只能在一个很小的维度上进行 演示.而绝热量子计算 (adiabatic quantum computing, AQC) 为操控量子系统提供了一种有别于基 于逻辑门的方法.因为量子机器学习通常涉及的多 元优化可以直接被 AQC 实现,所以对于量子机器 学习而言,AQC 可能是最有希望取得实际应用的 量子模型.AQC 的核心是让系统初态处于一个实 验上容易实现的简单哈密顿量 **H**₀(*t*)的基态,这个 哈密顿量会随着时间缓慢地朝着目标哈密顿量 H_p 变化,这个变化过程可以由瞬时哈密顿量表示:

$$\boldsymbol{H}(t) = [1 - s(t)] \boldsymbol{H}_{0} + s(t) \boldsymbol{H}_{p}, \qquad (11)$$

其中s(t)随着时间从0变化到1,导致H(t)由 H_0 变化至 H_p . 绝热理论告诉我们:当变化的过程足够慢,量子系统将始终待在瞬时哈密顿量H(t)的基态上.这样就通过设计目标哈密顿量来实现目标量子态的获取.

近年,基于 AQC 求解线性方程组的量子算法 被提出^[33].清华大学龙桂鲁课题组^[34]也在 2018 年 在核磁共振系统上实现了两种基于 AQC 解8×8 线性方程组的算法.实验样品为¹³C被标记的巴豆 酸分子,分子上的4个C核自旋作为4个量子比特. 实验中第一种算法的绝热演化的瞬时哈密顿量为

$$\boldsymbol{H}(s) = \boldsymbol{A}^{2}(s) - \boldsymbol{A}(s) \left| \bar{\boldsymbol{b}} \right\rangle \left\langle \bar{\boldsymbol{b}} \right| \boldsymbol{A}(s), \qquad (12)$$

其中 $A(s) = (1-s)\sigma_z \otimes I + s\sigma_x \otimes A$, 是一个包含 了线性方程组系数矩阵的变化参数. $|\bar{b}\rangle = |+,b\rangle$, $|\pm\rangle \ge \sigma_x$ 在计算基矢表示下的本征态, $|b\rangle$ 则是待解 线性方程组右侧的向量. 通过对s的构建和缓慢改 变, 最终系统的本征态将从 $|-,b\rangle$ 演化至 $|+,x\rangle$. 只 要将辅助比特舍弃, 即可得到解向量 $|x\rangle$. 第二种算 法的瞬时哈密顿量为

 $H'(s) = \sigma^+ \otimes A(s) P_b^\perp + \sigma^- P_b^\perp A(s),$ (13) 其中 $\sigma^\pm = (\sigma_x \pm i\sigma_z)/2, P_b^\perp = I - |\bar{b}\rangle \langle \bar{b} | = - \langle \bar{b} \rangle$ 投影算符. 经过绝热演化,系统从初态 $|0\rangle \otimes |-,b\rangle$ 变为末态 $|0\rangle \otimes |+,x\rangle$. 最终得到解向量 $|x\rangle$. 这个实 验演示了在不使用相位估计算法的情况下,AQC 算法同样可以在较低的时间复杂度下解线性方程 组,并且能够在相同的比特资源下实现更多维度的



计算. 对于当下的量子计算发展, 如何拓展量子体 系的维度是一个重大的挑战, 因此量子比特数对于 任何实验系统而言都是稀缺资源. 该实验所展示 的 AQC 算法对辅助比特数的要求并不依赖于量 子系统的维度, 这对于量子机器学习算法实现应用 具有重大意义.

2.2 解线性微分方程

在应用科学面临的许多动力学问题中,线性微分方程组 (LDEs) 扮演了一个很重要的角色. LEDs 问题可以概括为:对于给出的 $N \times N$ 矩阵M, N 维的向量 $|b\rangle$ 以及初始态向量 $|x(0)\rangle$,希望从含时的微分方程组

$$d\boldsymbol{x}(t) / dt = \boldsymbol{M}\boldsymbol{x}(t) + |\boldsymbol{b}\rangle$$
(14)

求出解向量|x⟩.

和线性方程组问题类似, 经典计算机解线性微 分方程要耗费大量计算资源. 当面临例如量子系统 或者流体力学的问题时,计算的维度大大增加,这 对于经典计算机来说难以解决. 随着量子计算的发 展,解线性方程组的量子算法已经被提出^[19,33],也 相继在量子计算平台上进行验证. 而解线性微分方 程组的量子算法虽然也被提出[35-37],但是这些算 法对于目前的量子计算机很难实现. 清华大学龙桂 鲁团队^[38]提出了一种容易在量子线路中实现的 基于逻辑门的解 LDEs 算法,并且在 4 bit 的核 磁共振体系上实现了解4×4的线性微分方程组, 如图 2 所示. 线性微分方程组的解向量通式 $x(t) \approx \sum_{m=0}^{k} \frac{(Mt)^{m}}{m!} x(0) + \sum_{n=1}^{k} \frac{M^{n-1}t^{n}}{n!} b$ f #述是非幺正演化,而传统的量子计算是基于封闭系 统和幺正演化. 该算法利用辅助系统使含有非幺正 演化子系统的大系统进行幺正演化,最终得以用便 于实验实现的逻辑门线路完成解微分方程组. 解向 量通式可以由计算基矢|j>展开,

$$|\boldsymbol{x}(t)\rangle = \sum_{m=0}^{k} \frac{||\boldsymbol{x}(0)||(||\boldsymbol{M}||\boldsymbol{A}t)^{m}}{m!} |\boldsymbol{x}(0)\rangle + \sum_{n=1}^{k} \frac{||\boldsymbol{b}||(||\boldsymbol{M}||\boldsymbol{A})^{n-1}t^{n}}{n!} |\boldsymbol{b}\rangle, \quad (15)$$

其中 $\boldsymbol{x}(0)$ 和 \boldsymbol{b} 可以由量子态 $|\boldsymbol{x}(0)\rangle = \sum_{j} \frac{x_{j}(0)}{||\boldsymbol{x}(0)||} |\boldsymbol{j}\rangle,$ $|\boldsymbol{b}\rangle = \sum_{j} \frac{b_{j}}{||\boldsymbol{b}||} |\boldsymbol{j}\rangle. \boldsymbol{M}$ 可由算子 $\boldsymbol{A} = \sum_{i,j \frac{M_{ij}}{||\boldsymbol{M}||}} |\boldsymbol{i}\rangle\langle \boldsymbol{j}|$ 来描述.



图 2 解线性微分方程的量子线路图.线路中第一个辅助 寄存器是单比特,第二个辅助寄存器为 $T = \log_2(k+1)$ 比特,然后是一个工作系统.所有的辅助寄存器被初始化为 $|0\rangle|0\rangle^{T}$,控制操作 $U_x 和 U_b$ 分别被用来生成 $|\boldsymbol{x}(0)\rangle$ 和 $|\boldsymbol{b}\rangle$. 在编码和解码期间的演化算子为 $\sum_{\tau=0}^{k} |\tau\rangle\langle\tau| \otimes U_{\tau}$.在 线路的结尾,在所有辅助比特为 $|0\rangle$ 的子空间中测量工作系统的态矢^[38]

Fig. 2. Quantum circuit for solving linear differential equations. The first auxiliary register in the circuit is a single bit, and the second auxiliary register is $T = \log_2 (k + 1)$ bits, then is a working system $|\phi\rangle$. All auxiliary registers are initialized to $|0\rangle |0\rangle^{\mathrm{T}}$, and then the operation U_x and U_b are used to generate $|\boldsymbol{x}(0)\rangle$ and $|\boldsymbol{b}\rangle$. The evolution operator during encoding and decoding is $\sum_{\tau=0}^{k} |\tau\rangle \langle \tau | \otimes U_{\tau}$. At the end of the circuit, the state vector of the working system is measured in the subspace where all auxiliary bits are $|0\rangle^{[38]}$.

有别于其他以相位估计为核心的量子算法,该 算法的量子加速来自于通过辅助比特将非幺正演 化变为幺正演化,这种做法叫作 LCU(linear combination of unitaries)^[39].对于未来非幺正演化的量 子算法研究,这个工作具有很好的启发性.

2.3 量子支持向量机

支持向量机 (support vector machine, SVM) 是一类广泛使用的通过监督学习实现二元分类的 广义线性分类器,其在人脸识别、文本分类、手写 字符识别等场景中有重要应用^[40]. SVM 是通过构 造一个线性映射将数据映射到高维特征空间,在这 个高维空间找到一个超平面能使两类数据的间隔 最大,并通过这个超平面对新的数据进行分类.

已知训练数据集共有 *M* 个训练样本,每个样本都是一个在 *N* 维特征空间的特征向量:

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$$
$$\mathbf{x}_i = [d_1, d_2, \cdots, d_N], \qquad (16)$$

且对应的标签集为

$$Y = \{y_1, y_2, \cdots, y_M\},\$$

$$y \in \{-1, 1\}.$$
 (17)

在训练样本所在的特征空间中, 通过训练超平面的 法向量*w*和截距*b*寻找到一个超平面能够让任意数 据到这个超平面的距离要大于等于 1, 即能够让两 类训练数据分离:

$$y_i \left(\boldsymbol{w} \cdot \boldsymbol{x_i} + b \right) \ge 1. \tag{18}$$

因此判别新数据x类别时, 当 $w \cdot x + b \ge 1$ 时, 认为 该数据为+1类, 反之当 $w \cdot x + b \le -1$, 则认为数 据为-1类. 但通常很多数据是无法线性可分的, 因 此需要通过一个非线性的函数将数据从低维特征 空间映射到更高维度的空间, 同时需要定义该映射 函数的内积, 称为核函数, 以回避复杂的高维显式 内积计算.

SVM 在识别阶段时间复杂度是 *O*[poly(*NM*)], *N*是特征空间的维度, *M*是训练集的样本数. 而在 训练阶段通常采用二次规划求解超平面, 而求解二 次规划将涉及 *M* 阶矩阵的求解, 时间复杂度约为 *O*(*S*³ + *SN* + *SNM*)和 *O*(*NM*²)之间, *S*为支持 向量个数, 当 *M* 数目很大时该矩阵的存储和计算 将耗费大量的计算机内存和运算时间.

为了解决经典 SVM 无法使用大规模数据进 行训练的难题, 2014 年 Rebentrost 等^[41]提出了量 子支持向量机 (quantum support vector machine, QSVM). QSVM 在训练阶段,将法向量w表示为 训练样本 x_i 的线性组合,

$$\boldsymbol{w} = \sum_{i=1}^{M} \alpha_i \boldsymbol{x}_i, \qquad (19)$$

通过最小二乘法,权重参数α以及截距b可以通过 求解线性方程得出:

$$(b, \alpha_1, \alpha_2, \cdots, \alpha_M)^{\mathsf{T}} = \tilde{\boldsymbol{F}}^{-1}(0, y_1, y_2, \cdots, y_M)^{\mathsf{T}},$$
(20)

其中 \tilde{F} 是包含核函数K的 $(M+1) \times (M+1)$ 的 矩阵

$$\tilde{\boldsymbol{F}} \equiv \left(\boldsymbol{K} + \gamma^{-1} \boldsymbol{I}_M\right), \quad K_{i,j} = x_i \cdot x_j. \quad (21)$$

 $I_M = M \times M$ 的单位矩阵, $\gamma = 2$ 设定的训练误差 与 SVM 目标的相对权重. 其将 SVM 训练转化为 最小二乘法问题再通过 HHL 算法求解矩阵 \tilde{F} 的 逆, 能使 QSVM 的时间复杂度在训练和识别阶段 都能达到 $O[\log(NM)]$. 文献 [42] 中杜江峰团队在 NMR 系统实现 QSVM 在手写数字识别的实验, 采用碘三氟乙烯作为样品, 其中有一个¹³C 核自旋 和三个¹⁹F 核自旋. 两个比特作为训练数据寄存器, 一个比特作为标签寄存器, 一个比特作为辅助比 特. 实验巧妙地将字符"6"和"9"编码成二维线性可 分的特征向量用于训练和测试, 实验保真度接近 99%, 手写数字识别错误率低于 4%, 如图 3 所示.

2.4 量子主成分分析

主成分分析 (principal component analysis, PCA) 是机器学习中一种常用且费时的无监督学 习算法.这一方法利用正交变换把由线性相关变量 表示的观测数据转换为少数几个由线性无关变量 表示的数据,线性无关的变量称为主成分.这个算 法主要用于发现数据中的基本结构,即数据中变量 之间的关系^[43].

已知数据集一共*M*个样本,每个样本都是一个在*N*维特征空间的特征向量:

$$X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M\},\$$
$$\boldsymbol{x}_i = [d_1, d_2, \cdots, d_N].$$
 (22)

将数据归一化为*x*后构造协方差矩阵



图 3 手写字符"6"和"9"的识别结果, 第1—4行分别代表手写字符, 实验指示符, 相干项的幅度和识别结果⁴²

Fig. 3. Recognition results of handwritten characters of "6" and "9". Lines 1 to 4 represent handwritten characters, experimental indicators, amplitude, and recognition results, respectively^[42].

$$\boldsymbol{C} = \frac{1}{M} \tilde{\boldsymbol{X}} \tilde{\boldsymbol{X}}^{\mathrm{T}}, \qquad (23)$$

对协方差矩阵 C 进行特征值分解,得到所有特征值和对应特征向量,并对特征向量按其对应特征值大小降序排列,取前 K 个特征向量构成投影矩阵 P,最终利用投影矩阵 P 将数据集从 N 维特征空间投影到 K 维特征空间.由此可见特征值分解直接确定了 PCA 算法的时间复杂度为O (N²).

2014年, Lloyd 等^[44]提出了量子主成分分析 (quantum principal component analysis, qPCA). qPCA 用密度矩阵 ρ 表示协方差矩阵C, 通过相位 估计法将密度矩阵 ρ 编码在辅助比特从而解出 特征值, 而相位估计法所需的控制门 $U = e^{-i\rho t}$ 可 以用目标比特和编码比特之间用时间非常短的 SWAP操作构建:

$$\lim_{\Delta t \to 0} e^{-i\boldsymbol{\rho}\Delta t} \boldsymbol{\sigma} e^{i\boldsymbol{\rho}\Delta t} = \lim_{\Delta t \to 0} \operatorname{Tr}_P \left(e^{-i\boldsymbol{S}\Delta t} \boldsymbol{\sigma} \otimes \boldsymbol{\rho} e^{i\boldsymbol{S}\Delta t} \right),$$
(24)

因此 qPCA 能够将时间复杂度降低到 $O(\log(N))$.

然而上述方法需要的辅助比特数量随密度矩 阵的大小指数增加,大量的资源消耗以至于在实验 平台上很难该算法进行验证,所以文献 [45] 中实验 团队通过参数量子电路实现 qPCA, 如图 4 所示, 该方法构建出一个厄密算符*U*,能够将密度矩阵*ρ* 被对角化同时特征值按大小降序排列:

$$\boldsymbol{\rho}_{\mathrm{f}} = \mathcal{U}\boldsymbol{\rho}\mathcal{U}^{\dagger} = \sum_{j=0}^{N-1} \lambda_j \left| \boldsymbol{j} \right\rangle \left\langle \boldsymbol{j} \right|.$$
 (25)

这需要构建一个参数量子电路U(θ)以及一个算 符**P**:

$$\boldsymbol{P} = \frac{1}{N(N-1)} \sum_{j=1}^{n} 2^{j-1} \left(\boldsymbol{\sigma}_{z}^{j} + 1 \right), \qquad (26)$$

构建出损失函数 $L(\boldsymbol{\theta})$:

$$L(\boldsymbol{\theta}) = \operatorname{Tr}\left(\mathcal{U}(\boldsymbol{\theta})\,\rho\mathcal{U}^{\dagger}(\boldsymbol{\theta})\cdot\boldsymbol{P}\right). \tag{27}$$

通过梯度下降优化参数θ最小化损失函数L(θ), 就能训练出目标厄密算符U,从而实现 qPCA算 法. qPCA 求解出人脸图片训练集协方差矩阵的特 征值和对应的特征向量,通过特征向量构造投影矩 阵将高维的图片数据投影到低维空间,再通过欧拉 距离公式将测试集图片分类实现人脸识别.实验使 用四比特核磁共振量子系统,将量子比特编码在核 自旋,通过选择跃迁法初始化量子系统,并用梯度



图 4 通过 qPCA 实现人脸识别的流程图.通过混合经典量子控制方法对 PQC U(θ)进行迭代优化,其中在量子处理器上测量 目标函数 L(θ)和梯度 g(θ).参数 θ 的存储和更新在经典计算机上实现.用优化后的 Ug 来计算特征脸矩阵 D 和协方差矩阵 C 的 特征向量^[45]

Fig. 4. Workflow for human face recognition via qPCA. The PQC $\mathcal{U}(\theta)$ is iteratively optimized via the hybrid classical quantum control approach, where the objective function $L(\theta)$ and the gradient $g(\theta)$ are measured on the quantum processor. The storage and update of the parameters θ are implemented on a classical computer. The optimized PQC with the operator U_g is applied to compute the eigenvectors of the eigenface matrix D and the covariance matrix $C = A A^{T}$ ^[45]. 下降优化控制脉冲方法来载入经典数据,最后利用 量子计算机和经典计算机混合系统优化参数量子 电路.实验最终求出的特征值保真度达到 99%,测 试集人脸识别正确率为 100%.

3 NV 色心体系

金刚石中的 NV 色心由一个替代 C 原子的 N 原子以及相邻位置 C 原子的缺失产生的空位组 成^[46],如图 5(a)所示. NV 色心是一种很重要的量 子体系,得益于其固态及室温下可操控等特点,在 量子计算、量子信息、量子精密测量等多个领域都 有很好的应用前景^[47-49].目前,国内外都有科研团 队在基于 NV色心体系的量子技术上展开研究.自 然状态下,NV色心具有两种电荷态-电中性和带负 电荷.带负电的 NV-由于其便于初始化、操控、读 出,对其本身及其应用的研究也最为广泛和深入, 本章节中我们将要回顾的相关研究也均是基于 NV⁻.

对于 NV⁻, 色心在俘获一个电子后带负电. N 原子与相邻的 C 原子形成共价键, 与空位之间 不成键, 所以贡献两个电子, 同样, 空位周围的 3 个 C 原子也分别贡献 1 个电子, 本身的 5 个电 子, 加上俘获的电子, 共有 6 个电子. 在实验中可 以认为其等价于 1 个自旋为 1 的电子自旋. NV⁻的 基态和激发态的电子态分别为 ${}^{3}A_{2}$ 和 ${}^{3}E$, 均为自 旋三重态, 基态因自旋之间的相互作用分裂成 $|m_{s} = 0\rangle$ 和 $|m_{s} = \pm 1\rangle$, 而 $|m_{s} = \pm 1\rangle$ 是简并的 [50], 如 图 5(b) 所示. 选取其中的两个能级, 就可以构建一 个量子比特. 除电子自旋外, 形成色心的¹⁴N(¹⁵N) 原子也可以作为一个自旋为 1(1/2) 的核自旋量子 比特, 若在色心周围存在¹³C原子, 同样可以作为 自旋为 1/2 的量子比特, 这是目前基于 NV 色心实 验体系来扩展量子比特的主要手段.

在不外加静磁场的情况下,由于自旋之间的 相互作用,劈裂成 $|m_s = 0$)和 $|m_s = \pm 1$),其零场劈 裂为 2.87 GHz. 施加静磁场后,简并的能级 $|m_s = \pm 1$)将会分开,并且劈裂会随着轴向磁场的 增强而增大.

在外场下 NV 色心的系统哈密顿量 (以电子自旋与¹⁴N 核自旋的两比特系统为例):

$$\boldsymbol{H}_{\rm NV} = D_{\rm gs} \boldsymbol{S}_z^2 + P_{\rm N} (\boldsymbol{I}_z^{\rm N})^2 - B_0 (\gamma_{\rm e} \boldsymbol{S} + \gamma_{\rm N} \boldsymbol{I}^{\rm N}) + \boldsymbol{S} \cdot \boldsymbol{A}^{\rm N} \cdot \boldsymbol{I}^{\rm N},$$
(28)

其中, $S_z^2 = I_z^N \beta N V$ 电子自旋与¹⁴N核自旋 算符, $D_{gs} = 2.87 \text{ GHz}$, 为 NV 的电子自旋零场劈 裂, $P_N = -4.95 \text{ MHz}$ ^[51] 为 NV 中心¹⁴N核自旋的 四偶极矩劈裂. (28) 式中的第三项为 NV 在外磁场 下的塞曼劈裂, 其中 $\gamma_e = 2.082 \text{ MHz}/\text{G} = \gamma_N =$ -0.308 kHz/G 分别为 NV 电子自旋与¹⁴N核自旋 的旋磁比. 最后一项是 NV 电子自旋与¹⁴N核自旋 的超精细相互作用 (hyperfine interaction), A^N 为 超精细结构张量.

要进行量子计算,量子体系不仅要具备可操控的量子比特,还要能够初始化以及读出其状态.对于 NV 色心,其电子自旋是通过激光实现初始化和



图 5 (a) 金刚石 NV 色心结构图; (b) NV 色心电子能级跃迁过程示意图, ³A₂和 ³E分别代表基态和激发态, ¹A₁和 ¹E为中间亚 稳态, 从激发态直接跃迁回基态会发出荧光, 而经中间态回基态不会发出荧光

Fig. 5. (a) NV color center structure; (b) schematic diagram of the transition process of NV color center electron energy level, ${}^{3}A_{2}$ and ${}^{3}E$ represent the ground state and excited state, respectively, ${}^{1}A_{1}$ and ${}^{1}E$ are the intermediate metastable states, which from the excited state directly transitions back to the ground state and emit fluorescence. But the path througt metastable state returns to the ground state without emitting fluorescence.

读出的. NV-的零声子线位于 637 nm, 在室温下, 由于声子的参与, 一般使用 532 nm 的激光激发其 至激发态. 当处于激发态时, 有两种路径可以回到 基态. 第一种是辐射 637—750 nm 的荧光由激发 态直接跃迁回基态, 第二种则是经过中间态回到基 态. 对于第二种情况, 自旋是不守恒的, 且不会辐 射 637—750 nm 的荧光 [52]. 值得注意的是, 如果 NV 色心的电子在激发前处于 $|m_s = \pm 1\rangle$ 自旋态,被 激发到第一激发态之后,其将更倾向于通过中间态 回到基态,这就意味着将会有更大的概率得到 |m_s = 0⟩的自旋态.同时,如果 NV 色心的电子在基 态时处于|m_s = 0)的自旋态, 它会有更大的概率沿 着辐射跃迁的路径,释放荧光后直接回到基态^[53]. 所以我们还是有更大的概率得到|ms=0)的自旋 态. 经过这个过程, 电子在 $|m_s = \pm 1\rangle$ 上的布居度将 不断减少, |m_s = 0)上的布居度将不断的增大^[46]. 由此,可以实现对色心电子自旋的初始化.同样读 出时也会用到上面的性质. 经过中间态回到基态的 跃迁过程并不会辐射 637—750 nm的荧光, 所以通 过荧光强度可以判断出电子自旋跃迁回基态时经 历的路径.同时结合所知的,处于不同自旋态的电 子自旋跃迁时选择路径的倾向也不同,可以得出结 论: NV 色心在不同自旋态时, 对应的辐射荧光强 度的不同,以此区分色心的自旋态.

NV 色心体系由于其比较成熟的操控技术、较长的退相干时间 (NV 基态自旋具有固体中任何电

子自旋中最长的室温单自旋退相干时间 (T₂), 在某 些样品中大于 1.8 ms), 成为了很多科研组在实验 体系下实现量子算法的选择. 将 NV 色心体系作为 量子模拟器进行模拟以帮助解决物理问题也是近 年来的一个重要研究方向.

3.1 卷积神经网络

拓扑相的发现改变了人们对量子相的理解.不 像传统的相位,拓扑相位不满足对称破裂的范式. 相反,每个对称类中都存在不同的拓扑相.在实验 中,拓扑相的识别通常依赖于测量具有拓扑起源的 某些属性,而在实验上实现一个明确可以辨别拓扑 相的方法,是十分具有挑战性的.机器学习很有希 望帮助解决这一问题.在邓东灵与段路明课题组合 作的工作^[54]中,利用机器学习对三维手性拓扑绝 缘体的外来拓扑相进行了识别.

卷积神经网络 (convolutional neural networks, CNN) 作为如今发展最迅速的一种神经网络模型, 在计算机视觉、医学影像处理、自然语言处理等领 域广泛使用^[55,56],该模型的共享权值的特点,大大 减少了模型训练的复杂度,同时能够进行表征学 习,并且提取出数据的全局特征和局部特征.

考虑一种三维手性拓扑绝缘体的最小的3带 哈密顿量 $\mathcal{H} = \sum_{k \in BZ} \Psi_k^{\dagger} H_k \Psi_k$,单粒子的动量分 辨哈密顿为:

$$\mathcal{H}_{\boldsymbol{k}} = \begin{pmatrix} 0 & q_1(\boldsymbol{k}) - iq_2(\boldsymbol{k}) & 0 \\ q_1(\boldsymbol{k}) + iq_2(\boldsymbol{k}) & 0 & q_3(\boldsymbol{k}) + iq_0(\boldsymbol{k}) \\ 0 & q_3(\boldsymbol{k}) - iq_0(\boldsymbol{k}) & 0 \end{pmatrix},$$
(29)

其中 $q_1(\mathbf{k}) = t \sin k_x$, $q_2(\mathbf{k}) = t \sin k_y$, $q_3(\mathbf{k}) = t \sin k_z$, $q_0(\mathbf{k}) = t (\cos k_x + \cos k_y + \cos k_z + h)$. t 被设定为 1, 而 h 是一个与维度无关的可调参量. 通过调控 NV 色心电子自旋基态的三个能级, 在实验上成功 模拟出了所需的哈密顿量, 如图 6(a) 所示. 通过 将 NV 色心基态的三个能级编码为 $|1\rangle$, $|0\rangle$, $|-1\rangle$ 态, 观察 H_k 可以发现其并没有对角项,并且 $|1\rangle$ 与 $|-1\rangle$ 之间并没有直接的耦合. 因此实验中可以施加能级 之间的共振微波, 通过绝热过程, 可以实现对应不 同 k 值点的哈密顿量.

利用 NV 色心中的电子自旋模拟的三维手

性拓扑绝缘体哈密顿量生成的原始数据,训练卷 积神经网络以识别拓扑相,这样的监督学习类似 于处理经典的图像识别.如图 6(b)所示,将在 10 × 10 × 10规则网格的密度矩阵的实验数据作为输 入,通过一个用来预测拓扑不变量的预训练卷积 神经网络迁移到分类不同拓扑相的问题中,输出 各个可能的拓扑相得经典概率.实验发现即使 只用输入 10%实验数据,模型也能够达到 90% 的识别率,表明了即使使用最少数量的数据样本, 也可以训练 CNN 来成功识别拓扑阶段,如图 7 所示.



图 6 (a) 利用共振微波操控 NV 色心基态能级; (b) 可对拓扑相进行分类的 3D 卷积神经网络的体系结构, 输入是在 10 × 10 × 10 规则网格上的密度矩阵的实验数据. 每个密度矩阵由八个实数表示. 输出是每个可能相的分类概率^[54]

Fig. 6. (a) Using resonance microwave to control the ground state energy level of NV color center; (b) architecture of the 3D CNN to classify the topological phases. The input is experimental data of density matrices on a $10 \times 10 \times 10$ regular grid. Each density matrix is represented by eight real numbers^[54].



图 7 迭代次数增加时的训练和验证准确性.训练和验证 准确性在训练过程开始时迅速增加,然后达到了很高的饱 和值 (≈ 98%)^[54]

Fig. 7. The training and verification accuracy when the number of iterations increases. The training and validation accuracy increased rapidly at the beginning of the training process, and then reached a high saturation value ($\approx 98\%$)^[54].

训练卷积神经网络的方式将比传统的方法更 加有效. 首先, 对于未知序参量和拓扑不变量的拓 扑相, 由于不需要知道各个拓扑不变量的精确解 而只需要必要的训练数据, 所以可以被更广泛地应 用在未知拓扑相之中. 再者, 没有在常规网格上提 供数据的情况下, 离散积分的常规方法无法使用. CNN 在这方面具有更好的应用范围. 因此在实验 中所需要的样本也就更少. 并且, 机器学习是直接 从混态之中提取数据, 而不需要依赖于纯态定义 的拓扑不变量. 这在量子实验体系中是具有极大优 势的.

3.2 量子主成分分析

由于主成分分析 (PCA) 在对数据降维的同时 能够尽可能保留有效信息的特性,其在机器学习 领域有着十分广泛的应用前景.量子主成分分析 (qPCA)由于量子算法的量子加速特性,可以比经 典算法更加高效地解决问题. 对未知的低秩密度矩阵的量子主成分分析以量子形式快速揭示了与大特征值相对应的特征值,并为机器学习和数据分析提供了潜在的量子加速器. 对于 qPCA, 如何萃取出其主成分一直是一个问题. 前文提到, qPCA 算法解决了时间复杂度的问题, 但实验上仍然需要大量的运算资源, 对量子比特数和操控的精确度都有着很高的要求, 所以完成实验一直都是一件很困难的事情.

杜江峰课题组^[57] 基于金刚石 NV 色心体系利 用量子机器学习完成了主成分分析算法 (PCA), 如图 8 所示. 在实验中使用了基于共振的量子主成 分分析 (RqPCA), 通过引入一个辅助比特, 实现了 算法的量子指数加速.

实验体系的哈密顿量为 $H = |1\rangle\langle 1| \otimes \rho + \frac{\omega}{2}\sigma_z \otimes I_n$,其中第一项为辅助比特与量子寄存器的耦合项,第二项为所添加的可调能量偏执. I_n 为与 ρ 维度相同的单位阵.当施加一个以强度 c 驱动辅助比特的外场,若 $\omega \approx \lambda_i$,则会发生从 $|0\rangle|\lambda_i\rangle$ 到 $|1\rangle|\lambda_i\rangle$ 的跃迁.当初态制备到 $\rho_{mi} = |0\rangle\langle 0| \otimes \rho$ 时,系统将在下面的哈密顿量下演化:

$$\mathcal{H}_{\mathrm{Rq}}(\omega) = \frac{\omega}{2} \boldsymbol{\sigma}_{z} \otimes \boldsymbol{I}_{n} + c \boldsymbol{\sigma}_{x} \otimes \boldsymbol{I}_{n} + |1\rangle \langle 1| \otimes \boldsymbol{\rho}$$
$$= \sum_{i} \frac{\omega - \lambda_{i}}{2} \boldsymbol{\sigma}_{z} \otimes |\lambda_{i}\rangle \langle \lambda_{i}| + c \boldsymbol{\sigma}_{x} \otimes \boldsymbol{I}_{n}$$
$$+ \boldsymbol{I}_{2} \otimes \sum_{i} \frac{\lambda_{i}}{2} |\lambda_{i}\rangle \langle \lambda_{i}|.$$
(30)

通过对 ω 进行扫频,当满足某个特定 λ_i ,探测比特会以


图 8 共振量子主成分分析算法原理图 (a) 探针-寄存器耦合系统的能级结构, $|\lambda_i\rangle \ge \rho$ 的第*i*个本征态, 而 $\lambda_i \in [0, 1]$ 是对应的本征值, 如果扫描频率 $\omega \approx \lambda_i$, 就会引起探针量子位的拉比振荡; (b) 使用 Suzuki-Trotter 分解的 RqPCA 的量子电路, 对探针量子位进行投影测量得到 [1)表明该算法成功^[57]

Fig. 8. Algorithm schematic of RqPCA: (a) The energy structure of the coupled probe-register system. $|\lambda_i\rangle$ is the i-th eigenstate of ρ and $\lambda_i \in [0, 1]$ is the corresponding eigenvalue. Once the scanning frequency $\omega \approx \lambda_i$, the Rabi oscillations of the probe qubit is induced; (b) the quantum circuit of RqPCA. The projective measurement of the probe qubit in the state $|1\rangle$ indicates success of the algorithm, with principal component being distilled in the register^[57].

$$P_i(\omega) = \lambda_i D_i^2 \sin^2\left(\frac{c\tau}{D_i}\right), \quad i = 1, 2, \cdots, n \quad (31)$$

的概率跃迁,其中 $D_i = \sqrt{\frac{(2c)^2}{(2c)^2 + (\omega - \lambda_i)^2}}$.在实验中,通过测量探测比特在 $|1\rangle$ 上的布居度,可以得到相应的谱线,而每一个峰就对应一个相应的特征值.对一个4×4的矩阵进行主成分分析,其效率可达 86%,保真度可达 0.90.

3.3 量子自编码器

自动编码器的想法在神经网络领域已经流行 了数十年,通常是为了降低数据的维度.它以一种 无监督的方式学习有效的数据编码,包括一个编码 器,用于学习一组数据的表示形式;以及一个解码 器,用于从简化的编码中生成尽可能接近其原始输 入的表示形式.自编码器通过简单地学习将输入复 制到输出来工作.这一任务(就是输入训练数据, 再输出训练数据的任务)听起来似乎微不足道,然 而这过程有一个很大的难点,那就是总要根据自己 的需求去找到合适的方式来约束训练的过程.我们 最常用的降维其实就是一个很好的例子,限制数据 的维度并不是一个简单的过程.我们所加的这些限 制条件,一方面要能够满足我们的要求(比如对噪 声的处理或者还原),另一方面还需要能够有效地 防止程序机械的将数据复制输出,使其具有高效表达的能力.

受到经典自编码器的启发,其在量子领域也可 以用来解决一些传统方法难以轻松解决的问题.接 下来我们介绍其中一种很有效的应用.文献 [58] 基于 NV 体系构建了量子自编码器,如图 9 所示, 并通过提高量子纠缠的寿命进行了验证.量子纠缠 作为极为重要的量子资源,却又非常脆弱,容易受 到环境的干扰而发生退相干,因此保护纠缠,抑制 退相干是量子领域里的核心话题.有一种方法是将 纠缠的有效信息通过机器学习的方式编码到相干 时间长的子空间,而在需要取出信息时再进行反向 解码操作,还原为原有信息,这样就变相的提高了 量子纠缠的寿命.

文中基于参数化量子线路的思想构建编码器, 编码器设定为共包含四个参数的操控算符,四个参 数分别对应两束选择性微波脉冲 MW1和 MW2 的振幅和相位.之后采用混合量子经典方法(HQCA) 进行训练,由量子系统和经典系统共同组成训练体 系,基于梯度算法对参数进行迭代,成本函数和梯 度的计算在量子系统中执行,其余部分则由经典系 统完成.每次迭代都根据由量子系统测量得到的成 本函数的值以及设定的步长来进行梯度运算,以合 适的学习率对参数进行调整更新,直至其满足实验 需求而完成迭代.这种方式能够缓解量子资源不足 的问题,同时经过机器学习过程也能抑制量子系统 本身造成的实验误差.本文基于金刚石 NV 色心体 系对这种方法进行了验证,把电子自旋和核自旋的 贝尔态中的纠缠信息编码到相干时间较长的核自 旋上进行保护,使得纠缠的寿命由 2 μs 提升到 3 ms,提升超过了 1000 倍.



图 9 (a) 量子自编码器线路图, 通过编码操作 $\mathcal{U}_{\mathcal{E}}$ 将 $|\Psi_i\rangle$ 中的信息压缩到 $|\phi\rangle_{code}$ 中, 在需要时通过解码操作 $\mathcal{U}_{\mathcal{D}}$ 将 $|\phi\rangle_{code}$ 还原为 $|\Psi_f\rangle$; (b) 优化编码器的基于梯度算法的 HQCA 的训练过程, ρ_{in} 是编码器的输入状态, ρ_{out} 是辅助 量子位的输出状态, $f\left(\mathcal{U}_{\mathcal{E}}^{(q)}\right)$ 是成本函数, q是迭代次数^[58] Fig. 9. (a) Quantum autoencoder circuit. The target information of $|\Psi_i\rangle$ can be encoded to the code state $|\phi\rangle_{code}$ via the encoder $\mathcal{U}_{\mathcal{E}}$. $|\phi\rangle_{code}$ can be reconstructed to $|\Psi_f\rangle$ when needed by the decoder $\mathcal{U}_{\mathcal{D}}$. (b) Training process of the gradient-based HQCA to optimize encoder. Here, ρ_{in} is the input state of the encoder, and ρ_{out} is the output state on the ancilla qubits. $f\left(\mathcal{U}_{\mathcal{E}}^{(q)}\right)$ is the cost function, where q is the current iterative number^[58].

4 展 望

值得注意的是,本文主要介绍了量子机器学习 基于 NMR 和 NV 色心这两种自旋量子体系在近 年来的实验进展,并不是关于量子机器学习的全面 综述,只介绍了这两种实验体系涉及到的部分量子 机器学习算法.除了这两种体系,量子机器学习在 诸如超导^[32,59]、光学^[30,60]等体系中都有着良好的 发展.

量子机器学习是近几年新兴的交叉研究领域, 有着巨大的发展潜力.虽然已经取得了一定的成 果,但是在未来的发展中依然存在着很多问题^[14]. 首先,在解决实际问题时,需要将经典的信息导入 到量子体系当中,这个经典到量子的信息转换过程 本身就很复杂,会耗费大量的资源,目前仍没有行 之有效的方法.其次,常见的各类量子体系还都不 够成熟,量子机器学习的发展在今后相当长的一段 时期内仍将受限于量子比特的数目、相干性、可纠 错性等问题,构建通用量子计算机或专用量子处理 器仍十分困难.最后,对于已有的量子算法,难以 断言其处理问题的过程中,是否存在着超越经典算 法的优势,同时也缺少评估量子算法有效性的系统 性方法.

虽然还存在种种问题,现阶段量子机器学习在 解决各学科领域的问题已经显示出巨大优势潜力, 尤其是为一些常见的量子物理问题的分析和处理 提供了一种有效的手段,例如量子拓扑相的鉴定 等^[54,61-66],未来必将能够继续助力各领域的突破 和发展.

参考文献

- Mitchell T M 1997 Machine Learning (Boston, MA, USA: McGraw-Hill)
- [2] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, Zdeborová L 2019 *Rev. Mod. Phys.* 91 045002
- [3] Athey S 2018 The Impact of Machine Learning on Economics, in The Economics of Artificial Intelligence: An Agenda (Chicago: University of Chicago Press) pp507-547
- [4] Liakos K G, Busato P, Moshou D, Pearson S, Bochtis D 2018 Sensors 18 2674
- Krizhevsky A, Sutskever I, Hinton G E 2012 Advances in Neural Information Processing Systems 25 pp1097-1105.
- [6] Simonyan K, Zisserman A 2014 arXiv: 1409.1556 [cs.CV]
- [7] He K, Zhang X, Ren S, Sun J 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp770-778
- [8] Huang G, Liu Z, Van Der Maaten L, Weinberger K Q 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp4700-4708
- [9] Brown T B, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S 2020 arXiv: 2005.14165 [cs.CL]
- [10] Rønnow T F, Wang Z, Job J, Boixo S, Isakov S V, Wecker D, Martinis J M, Lidar D A, Troyer M 2014 Science 345 420
- [11] Feynman R P 1982 Int. J. Theor. Phys. 21 467
- [12] Arute F, Arya K, Babbush R, et al. 2019 Nature 574 505
- [13]~ Zhong H S, Wang H, Deng Y H, et al. 2020 Science $\mathbf{370}~1460$
- [14] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S 2017 Nature 549 195
- [15] Deutsch D 1985 A. Math. Phys. Sci. 400 97
- [16] Shor P W 1994 Proceedings 35th Annual Symposium on Foundations of Computer Science Santa Fe, NM, USA, Nov. 20–22, 1994 pp124–134
- [17] Grover L K 1996 Proceedings of the Twenty-Eighth Annual

ACM Symposium on Theory of Computing Philadelphia PA, USA, 1996 pp212–219

- [18] Grover L K 1997 Phys. Rev. Lett. 79 325
- [19] Harrow A W, Hassidim A, Lloyd S 2009 Phys. Rev. Lett. 103 150502
- [20] Vandersypen L M K, Chuang I L 2004 Rev. Mod. Phys. 76 1037
- [21] Rabi I I, Zacharias J R, Millman S, Kusch P 1938 Phys. Rev. 53 318
- [22] Bloch F 1946 Phys. Rev. 70 460
- [23] Stewart W E, Siddall T H 1970 Chem. Rev. 70 517
- [24] Hore P J 2015 Nuclear Magnetic Resonance (United States: Oxford University Press)
- [25] Harris R K 1986 Nuclear Magnetic Resonance Spectroscopy (United States: OSTI)
- [26] Freeman R 1987 Handbook of Nuclear Magnetic Resonance (United States: OSTI)
- [27] Gershenfeld N A, Chuang I L 1997 Science 275 350
- [28] Cory D G, Fahmy A F, Havel T F 1997 Proc. Natl. Acad. Sci. 94 1634
- [29] Nielsen M A, Chuang I 2001 Quantum Computation and Quantum Information (10th Anniversary Edition) (United States: Cambridge University Press)
- [30] Barz S, Kassal I, Ringbauer M, Lipp Y O, Dakić B, Aspuru-Guzik A, Walther P 2014 Sci. Rep. 4 6115
- [31] Pan J, Cao Y, Yao X, Li Z, Ju C, Chen H, Peng X, Kais S, Du J 2014 Phys. Rev. A 89 022313
- [32] Cai X D, Weedbrook C, Su Z E, Chen M C, Gu M, Zhu M J, Li L, Liu N L, Lu C Y, Pan J W 2013 *Phys. Rev. Lett.* **110** 230501
- [33] Subaşı Y, Somma R D, Orsucci D 2019 Phys. Rev. Lett. 122 60504
- [34] Wen J, Kong X, Wei S, Wang B, Xin T, Long G 2019 Phys. Rev. A 99 012320
- [35] Leyton S K, Osborne T J 2008 arXiv: 0812.4423 [quant-ph]
- [36] Berry D W 2014 J. Phys. A: Math. Theor. 47 105301
- [37] Berry D W, Childs A M, Ostrander A, Wang G 2017 Commun. Math. Phys. 356 1057
- [38] Xin T, Wei S, Cui J, Xiao J, Arrazola I, Lamata L, Kong X, Lu D, Solano E, Long G 2020 Phys. Rev. A 101 032307
- [39] Shao C, Li Y, Li H 2019 J. Syst. Sci. Complex. 32 375
- [40] Platt J C 1998 Technical Report MSR-TR-98-14, Redmond, WA, USA
- [41] Rebentrost P, Mohseni M, Lloyd S 2014 Phys. Rev. Lett. 113 130503
- [42] Li Z, Liu X, Xu N, Du J 2015 Phys. Rev. Lett. 114 140504
- [43] Jolliffe I T 1986 Principal Component Analysis (Berlin: Springer)
- [44] Lloyd S, Mohseni M, Rebentrost P 2014 Nat. Phys. 10 631

- [45] Xin T, Che L, Xi C, Singh A, Nie X, Li J, Dong Y, Lu D 2021 Phys. Rev. Lett. 126 110502
- [46] Loubser J H N, Wyk J A 1978 Rep. Prog. Phys. 41 1201
- [47] Barry J F, Schloss J M, Bauch E, Turner M J, Hart C A, Pham L M, Walsworth R L 2020 Rev. Mod. Phys. 92 015004
- [48] Shi F, Zhang Q, Wang P, Sun H, Wang J, Rong X, Chen M, Ju C, Reinhard F, Chen H, Wrachtrup J, Wang J, Du J 2015 *Science* 347 1135
- [49] Pompili M, Hermans S L N, Baier S, Beukers H K C, Humphreys P C, Schouten R N, Vermeulen R F L, Tiggelman M J, Martins L S, Dirkse B, Wehner S, Hanson R 2021 *Science* 372 259
- [50] Doherty M W, Manson N B, Delaney P, Hollenberg L C L 2011 New J. Phys. 13 025019
- [51] Smeltzer B, McIntyre J, Childress L 2009 Phys. Rev. A 80 050302(R)
- [52] Manson N B, Harrison J P, Sellars M J 2006 Phys. Rev. B 74 104303
- [53] Batalov A, Zierl C, Gaebel T, Neumann P, Chan I Y, Balasubramanian G, Hemmer P R, Jelezko F, Wrachtrup J 2008 Phys. Rev. Lett. 100 077401
- [54] Lian W, Wang S T, Lu S, Huang Y, Wang F, Yuan X, Zhang W, Ouyang X, Wang X, Huang X, He L, Chang X, Deng D L, Duan L 2019 *Phys. Rev. Lett.* **122** 210503
- [55] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T 2018 Pattern Recognit. 77 354
- [56] Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman D J 2021 Mech. Syst. Signal Process. 151 107398
- [57] Li Z, Chai Z, Guo Y, Ji W, Wang M, Shi F, Wang Y, Lloyd S, Du J 2021 arXiv: 2104.02476 [quant-ph]
- [58] Zhou F, Tian Y, Song Y, Qiu C, Wang X, Chen B, Xu N, Lu D 2021 Preserving Entanglement in a Solid-Spin System Using Quantum Autoencoders Prepr.
- [59] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M, Gambetta J M 2019 Nature 567 209
- [60] Arrazola J M, Bromley T R, Izaac J, Myers C R, Brádler K, Killoran N 2019 Quantum Sci. Technol. 4 24004
- [61] Xiao L, Zhan X, Bian Z H, Wang K K, Zhang X, Wang X P, Li J, Mochizuki K, Kim D, Kawakami N, Yi W, Obuse H, Sanders B C, Xue P 2017 *Nature Physics* 13 1117
- [62] Che Y, Gneiting C, Liu T, Nori F 2020 Phys. Rev. B 102 134213
- [63] Holanda N L, Griffith M A R 2020 Phys. Rev. B 102 054107
- [64] Zhang Y, Ginsparg P, Kim E 2020 Phys. Rev. Research 2 023283
- [65] Ming Y, Lin C, Bartlett S D, Zhang W 2019 npj Computational Materials 5 88
- [66] Zhang P, Shen H, Zhai H 2018 Phys. Rev. Lett. 120 066401

SPECIAL TOPIC—Machine learning and physics

Experimental progress of quantum machine learning based on spin systems^{*}

Tian Yu Lin Zi-Dong Wang Xiang-Yu Che Liang-Yu Lu Da-Wei[†]

(Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China)
 (Received 12 April 2021; revised manuscript received 25 May 2021)

Abstract

Machine learning is widely applied in various areas due to its advantages in pattern recognition, but it is severely restricted by the computing power of classic computers. In recent years, with the rapid development of quantum technology, quantum machine learning has been verified experimentally verified in many quantum systems, and exhibited great advantages over classical algorithms for certain specific problems. In the present review, we mainly introduce two typical spin systems, nuclear magnetic resonance and nitrogen-vacancy centers in diamond, and review some representative experiments in the field of quantum machine learning, which were carried out in recent years.

Keywords: quantum machine learning, spin systems, nuclear magnetic resonance, nitrogen-vacancy centers in diamond

PACS: 03.65.-w, 03.67.Ac, 03.67.Lx

DOI: 10.7498/aps.70.20210684

^{*} Project supported by the National Key Research and Development Program of China (Grant No. 2019YFA0308100), the National Natural Science Foundation of China (Grant Nos. 12075110, 11975117, 11905099, 11875159, U1801661), the Guangdong Basic and Applied Basic Research Foundation, China (Grant No. 2019A1515011383), the Guangdong International Collaboration Program, China (Grant No. 2020A0505100001), the Science, Technology, and Innovation Commission of Shenzhen Municipality, China (Grant Nos. ZDSYS20170303165926217, KQTD20190929173815000, JCYJ20200109140803865, JCYJ20170412152620376, JCYJ20180302174036418), the Pengcheng Scholars, the Guangdong Innovative and Entrepreneurial Research Team Program, China (Grant No. 2019ZT08C044), and the Guangdong Provincial Key Laboratory, China (Grant No. 2019B121203002).

[†] Corresponding author. E-mail: ludw@sustc.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

机器学习辅助绝热量子算法设计

林键 叶梦 朱家纬 李晓鹏

Machine learning assisted quantum adiabatic algorithm design Lin Jian Ye Meng Zhu Jia-Wei Li Xiao-Peng 引用信息 Citation: Acta Physica Sinica, 70, 140306 (2021) DOI: 10.7498/aps.70.20210831 在线阅读 View online: https://doi.org/10.7498/aps.70.20210831 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

量子计算与量子模拟 Quantum computation and quantum simulation 物理学报. 2018, 67(12): 120301 https://doi.org/10.7498/aps.67.20180710

基于冗余图态的多人协作量子计算

Collaborative quantum computation with redundant graph state 物理学报. 2019, 68(11): 110302 https://doi.org/10.7498/aps.68.20190142

基于量子算法的量子态层析新方案

A novel scheme of quantum state tomography based on quantum algorithms 物理学报. 2019, 68(14): 140301 https://doi.org/10.7498/aps.68.20190157

铅基钙钛矿铁电晶体高临界转变温度的机器学习研究

High critical transition temperature of lead-based perovskite ferroelectric crystals: A machine learning study 物理学报. 2019, 68(21): 210502 https://doi.org/10.7498/aps.68.20190942

新型超导量子比特及量子物理问题的研究 Novel superconducting qubits and quantum physics 物理学报. 2018, 67(22): 228501 https://doi.org/10.7498/aps.67.20180845

利用超导量子电路模拟拓扑量子材料

Topological quantum material simulated with superconducting quantum circuits 物理学报. 2018, 67(22): 220302 https://doi.org/10.7498/aps.67.20181857

专题: 机器学习与物理

机器学习辅助绝热量子算法设计*

林键 叶梦 朱家纬 李晓鹏*

(复旦大学物理学系,上海 200433)

(2021年5月1日收到; 2021年6月13日收到修改稿)

量子计算在近十年取得了长足的进展.随着量子调控技术达到前所未有的高度,包括超导量子比特、光量子器件、原子系综等在内的量子实验平台都进入到了崭新的时代.目前在特定计算任务上超越经典的量子 计算优势也已经被报道.其中一种可以有效运用可控量子器件的计算方案是采用绝热量子计算.绝热量子计 算中算法的选择与研究至关重要,其将直接决定量子计算优势是否能够最大限度地被挖掘.本综述主要介绍 近期机器学习在绝热量子算法设计方面的应用,并讲述该计算架构在 3-SAT 和 Grover 搜索等问题上的应用. 通过与未经机器学习优化设计的绝热量子算法对比,研究表明机器学习方法的应用可以极大提高绝热量子 算法的计算效率.

关键词: 绝热量子计算, 量子算法, 量子模拟, 机器学习 PACS: 03.67.Ac, 03.67.Lx, 89.70.Eg, 07.05.Mh

DOI: 10.7498/aps.70.20210831

1 引 言

量子计算概念最早可以追溯到 20 世纪 80 年 代,当时 Beniof^[1]提出了量子图灵机概念,Feynman^[2]有了量子模拟的想法.而后 Deutsch^[3]提出 量子线路模型来实现普适量子计算,Yao^[4]证明了 量子线路模型与量子图灵机的等价性——两者可 以在多项式时间内互相模拟.在量子算法方面, 1994 年 Shor^[5]基于量子线路模型提出了可在多项 式时间内求解质因数分解问题的量子算法.由于质 因数分解问题的困难性是 Rivest-Shamir-Adleman (RSA) 公钥加密体系安全性的保障,这一由 Shor 提出的多项式量子算法引起了密码学和相关 实验领域的高度关注^[6-12].量子计算有了从理论研 究走向实际应用的趋势,量子计算机的研发开始引 起多方投入^[13].同时,人们也从量子信息理论和量 子计算复杂度的角度展开研究,期待设计出更多具 有显著量子计算优势的量子算法[14-17].

目前已知的具有超越经典计算优势的量子算 法主要可以归为三大类[16] 第一类是利用量子傅 里叶变换寻找周期的量子算法,包括 Shor 算法^[5]、 Simon 算法^[18]、以及 Hallgren 算法^[19]等. 第二类 是以有量子加速的 Grover 搜索算法^[20] 为基础的 搜索及优化算法——可以在 $O(\sqrt{N})$ 时间内完成对 N个项目的搜索. 第三类是在量子计算机上对复 杂的量子多体体系进行高效模拟. 这是基于 Fevnman"利用量子计算机来进行量子模拟"的想法^[2]. 量子系统的希尔伯特空间会随量子自由度(比如原 子数等)的数目指数增长,所以如果用经典计算 机来模拟量子多体系统需要大量的内存资源和 计算资源. 而量子计算机可以直接用量子比特进行 计算,具有对复杂量子多体系统模拟的天然优势. 这一方面对量子化学计算、材料科学的复杂微观 物理机制 (如高温超导)的揭示等具有重要科学 价值.

* 国家自然科学基金 (批准号: 11934002)、国家重点基础研究发展计划 (973 计划)(批准号: 2017YFA0304204) 和上海量子信息技 术市级科技重大专项 (批准号: 2019SHZDZX01) 资助的课题.

© 2021 中国物理学会 Chinese Physical Society

[†] 通信作者. E-mail: xiaopeng_li@fudan.edu.cn

量子算法与经典算法的很大一点不同是量子 力学允许的量子操作、量子叠加和量子纠缠很难直 接与直观经验建立联系,甚至是反直觉的.这极大 地增加了量子算法设计的难度,同时使得人们在经 典计算算法设计上积累的经验很难被直接借用^[16].

在开发设计量子算法时,我们期待设计出能够 相比于经典计算更高效的量子算法.对于可以在量 子计算机上以多项式时间解决的问题,人们把它们 归为 BQP (bounded-error quantum polynomial time)复杂类.目前人们还没有一个普适的理论来 确定这类问题的边界,在后文中我们将对此做更具 体的介绍.虽然通常人们不认为量子计算能够指数 加速 NP-complete 问题,但在具体算法设计中,任 何能够提升量子计算能力的设计方法和技巧都值 得尝试^[21].在量子计算的发展进程中,我们也可以 适当借鉴经典计算领域的发展.设计出具有量子优 势的量子算法还有待于量子研究领域与经典计算 研究领域的深度交融.

Farhi 等^[22,23] 在 2001 年提出了与量子线路 模型相应的绝热量子计算模型—— AQC (adiabatic quantum computation). 在绝热量子计 算中,我们首先会构造一个非平庸的问题哈密顿 量,其基态编码了我们关心问题的答案.然后我们 让系统从一个与问题哈密顿量非对易的平庸哈密 顿量基态开始做演化,一直演化到这个编码的问题 哈密顿量. 如果整个演化过程是完全绝热的并且基 态与激发态之间始终存在能级差^[24],我们最后也 就能获得问题哈密顿量的基态,即这个计算问题的 正确解. 这也就将一个计算问题变成了一个哈密顿 量求基态的问题. 绝热量子计算模型可以看作连续 时间的量子计算,其与离散时间的量子线路模型的 等价性也在理论上得到了证明[25,26]."演化体系哈 密顿量"这一想法与量子模拟非常接近. 而人们在 量子模拟及量子控制方面也具备了很多理论知识 和实验经验[27-30]. 所以绝热量子计算概念的提出 不但给我们提供了一种全新的普适量子计算框架, 而且有利于将我们对量子模拟的物理直觉与量子 算法设计结合起来以打开新思路. 后文将详细介绍 绝热量子计算的算法设计.

经典计算领域的算法设计复杂程度也在愈趋 加大,如何实现经典算法的自动化设计也变得越 来越重要.对于一个问题,如果存在多种可以解 决的算法,那么如何高效地挑选一个最优的算法 就非常关键.这就涉及算法选择问题 (algorithm selection)^[31];而对于一个算法,在不同的问题例子中如何去优化算法构型 (algorithm configuration)^[32]也相当重要.我们将在后文中详细介绍机器学习在经典算法设计领域的应用.另一方面,近些年机器学习也在处理量子多体问题上,特别是在物态的相分类^[33-37]、多体波函数的表示及基态制备^[38-41]、优化量子操控^[42-45]等方向上有了一系列的应用.

虽然经典与量子算法的设计领域差别很大,但 两者的复杂性使得它们都面临着巨大挑战.通过借 鉴机器学习在经典算法设计与量子多体物理中的 成功应用,我们也希望机器学习方法能辅助量子算 法设计.这不仅会帮助我们设计出具有量子优越性 的算法,同时设计获得的量子算法也有望实现机器 学习的量子加速^[46].我们期待这两个领域的交融 会碰撞出灵感的火花.

2 绝热量子计算与哈密顿量编码

本节将对绝热量子计算的概念以及绝热量子 计算中问题哈密顿量的编码方案与自旋玻璃物理 的联系进行介绍.而后,给出三个具体计算问题的 哈密顿量构造方案.最后讨论运用绝热量子计算模 型探索 BQP 复杂类的研究途径.

2.1 绝热量子计算

绝热量子计算作为一种普适的量子计算框 架^[22,23,25,26,47],其原理是将一个计算问题变成量子 体系求基态的问题.设想有两个非对易的哈密顿量 *H*_b和*H*_p,其中初始哈密顿量*H*_b的基态非常容易制 备,而问题哈密顿量*H*_p的基态则编码了我们所关 心问题的解.让量子系统从初始哈密顿量*H*_b的基 态开始演化,直到系统到达问题哈密顿量*H*_p.一般 可以用以下含时哈密顿量来刻画该过程:

 $H(s(t/T)) = [1 - s(t/T)]H_b + s(t/T)H_p,$ (1) 其中s(t/T)为哈密顿量的演化路径(从0变化到1), T为总的演化时间.如果演化时间T足够长并且 系统的基态与激发态之间始终存在能级差^[24],那 么由绝热定理可以保证,这个量子系统将时刻处于 瞬时哈密顿量的基态.量化绝热条件是绝热定理成 立的必要条件^[48].由此,可以估计出绝热演化时间 $T \gg O(\Delta_{\min}^{-2})^{[22,49,47]},其中 \Delta_{\min} 为基态与第一激发$ 态之间的最小能隙.通过测量演化 T时间后的系统状态,将得到问题哈密顿量的基态,即问题的解.

绝热量子计算在被提出之时就将目标指向解 决 NP-complete 和 NP-hard 问题^[22,23]. 而后, 对其 是否能够超越经典计算的质疑也接踵而来——因 为对于这类 NP-complete 和 NP-hard 问题, 其最 小能隙呈指数减小, 所以绝热量子计算需要的时间 是指数增长的. 其并不能做到相比于经典算法的指 数加速, 但可能在系数因子上会比经典的算法更 优^[50-55].

2.2 哈密顿量编码

本节将介绍绝热量子计算哈密顿量编码与自 旋玻璃问题的关联.也将介绍三个典型计算问题的 哈密顿量构造方法.

2.2.1 哈密顿量编码与自旋玻璃问题

在理论和实验中,总可以将绝热量子计算的哈 密顿量编码为伊辛自旋模型的量子形式^[56,57].一个 经典的伊辛自旋模型可以写作:

$$H(s_1, s_2, \cdots, s_n) = \sum_{i < j} J_{ij} s_i s_j + \sum_{i=1}^n h_i s_i.$$
 (2)

在绝热量子计算中,通过将自旋*s*_i写成泡利算符形 式来得到问题哈密顿量*H*_p:

$$H_{\rm p} = H(\sigma_1^z, \cdots, \sigma_n^z), \tag{3}$$

其中 σ_i^z 为作用在第i个自旋上的泡利 Z算符, J_{ij} 和 h_i 为实数. 可以将初始哈密顿量 H_b 设置为:

$$H_{\mathsf{b}} = \frac{1}{2} \sum_{i} [\mathbb{1} - \sigma_i^x], \qquad (4)$$

其中 σ_i^x 为作用在第i个自旋上的泡利 X 算符.

这样的问题哈密顿量构造与物理中的自旋玻 璃模型可以一一对应.自旋玻璃问题是一个在凝 聚态物理和统计及计算物理领域中悠久且丰富的 物理问题^[58-60].自旋玻璃问题和 NP(nondeterministic polynomial)问题的联系也备受关注.这 种关联给了我们从物理角度来理解计算中的困难 的机会^[61-63].Karp^[64]在1972年研究发现21个组 合及图论计算问题都可以在多项式时间归约到一 个 NP-complete问题上,也就证明了它们都是 NP-complete问题上,也就证明了它们都是 NP-complete问题.Lucas^[65]研究了这些典型NPcomplete问题.MP-complete问题.MP-complete问题如何编码为自旋玻璃形式的问题哈 密顿量.一般人们也会把这类编码后的问题叫作 QUBO (quadratic unconstrained binary optimization) 问题.

2.2.2 基于绝热量子计算的 Grover 搜索算法

基于线路模型的 Grover 搜索算法被证明具有 超越经典搜索算法的平方加速^[20].这个搜索问题 是指在 $N = 2^n$ 个项目中寻找到标记的项.对于一 个函数 $f: \{0,1\}^n \mapsto \{0,1\}, 只有被标记的项 f(m) =$ 1,对于任意的 $x \neq m, f(x) = 0$.我们的目标是用 最少的询问神谕 (oracle) 次数来找到这个标记的 项目 m. 对应到绝热量子计算,可以将初始哈密顿 量 写成 $H_b = \mathbf{1} - |\psi_0\rangle\langle\psi_0|$,以及问题哈密顿量 $H_p = \mathbf{1} - |m\rangle\langle m|, 其中|\psi_0\rangle = 1/\sqrt{2^n} \sum_{i=0}^{2^n-1} |i\rangle, |m\rangle$ 为某一个被标记的态.在这样的哈密顿量构造下, 如果将演化路径简单地选择为s = t/T,其中的最 小能隙出现在s = 1/2:

$$\Delta_{\min} = \Delta(s = 1/2) = \frac{1}{\sqrt{N}} = \frac{1}{\sqrt{2^n}}.$$
 (5)

由绝热定理条件可知, $T \gg 1/\Delta_{\min}^2 \sim O(2^n))^{[47]}$, 这 与经典搜索算法的复杂度一致. 所以在绝热量子计 算中简单地选择线性哈密顿量演化路径不会使得 Grover 搜索问题具有量子加速. 而在如此编码下, 研究表明: 可以解析地优化哈密顿量演化路径以使 得上述 Grover 搜索问题在绝热量子计算中依旧具 有平方的量子加速^[66].

2.2.3 基于绝热量子计算的 3-SAT 算法

布尔可满足性问题 (Boolean satisfiability problem) 中含有 n 个布尔变量 z_i ,由其组成了一 系列子句 (clause) C_{α} ,其中每一个子句 C_{α} 内都 含有 k 个变量并以"或"(\lor)连接,如: $C_{\alpha} =$ $(b_1 \lor b_2 \lor \cdots \lor b_k)$,其中 $b_i \in \{z_1, z_2, \cdots, z_n, \neg z_1, \neg z_2, \cdots \neg z_n\}$.最终我们希望找到一串布尔变量 z_i , 使得所有用"与"(\land)连接的子句都得到满足,即:

$$\Phi \equiv C_1 \wedge C_2 \wedge \dots \wedge C_r = 1. \tag{6}$$

如果每个子句中有相同变量个数k = 2,这类问题称作 2-SAT.这类问题可以在经典计算中有效地被解决,归属于复杂类 P. 而对于 $k \ge 3$ 的情况,这类问题都是 NP-complete 问题,它们可以在多项式时间内互相转化. Farhi 等^[22]提出绝热量子计算时就尝试对 3-SAT 问题进行测试.为了构造 3-SAT 问题哈密顿量,我们举例一个涉及三个布尔变量 z_{ic}, z_{ic}, z_{kc} 的子句 C,并且对此定义一个经典

的能量函数:

$$h_{C}(z_{i_{C}}, z_{j_{C}}, z_{k_{C}}) = \begin{cases} 0, & \text{m} \mathbb{R} z_{i_{C}}, z_{j_{C}}, z_{k_{C}} \notin \mathbb{P} \mathcal{P} \mathcal{P} \mathcal{O} \mathcal{C} \mathbb{R} \mathcal{L}; \\ 1, & \text{m} \mathbb{R} z_{i_{C}}, z_{j_{C}}, z_{k_{C}} \# \mathbb{P} \mathcal{P} \mathcal{P} \mathcal{O} \mathcal{C} \mathcal{T} \mathbb{R} \mathcal{L}. \end{cases}$$
(7)

定义 $h(z_1, z_2, \dots, z_n) = \sum_C h_C$ 并且将其在以 泡利 Z 算符本征矢为计算基下对角化成量子算符 形式的问题哈密顿量 H_p , 也就是对应于:

$$H_{\mathbf{p}}|z_1\rangle|z_2\rangle\dots|z_n\rangle = h(z_1, z_2, \cdots, z_n)|z_1\rangle|z_2\rangle\cdots|z_n\rangle.$$
(8)

于是, 这就将 3-SAT 问题的解编码到了 H_p的基态上.

2.2.4 基于绝热量子计算的质因数分解 算法

质因数分解问题是希望将一个大数 N 分解为 两个质因数 p 和 q,也就是实现 $N \rightarrow p \times q$ 的分解. RSA 公钥加密体系的安全性正是基于当前经典算 法无法在多项式时间内求解质因数分解问题. 在经 典计算领域,大家尝试了求解该问题的不同方法, 如基于启发式算法的设计[67],机器学习方法[68,69], 以及仿生算法[70,71] 和随机架构算法[72] 等. 而在量 子计算领域, Shor 算法可在多项式时间内解决质 因数分解问题^[5]. 但 Shor 算法对量子比特数量和 门的保真度要求很高,在目前的实验条件下^[73],还 只能在比较小的数字上做分解^[9,74]. 另一方面, 近 些年大家对在绝热量子计算中实现质因数分解也 做了大量工作[75,76,47,77,78,55]. 在绝热量子计算质因 数分解问题中,构造问题哈密顿量一般有两种主要 方式. 一种是直接将问题写成损失函数 fcost = (N-p×q)^{2[79]},为了避免其中耦合强度出现指数 增长,人们提出了另一种基于乘法表的损失函数构 造方法[77]. 其中, 通过将二进制数映射为泡利算符, 继而引入额外的量子比特将高阶耦合项降到二阶, 人们可以将这一问题约化到前文提到的 QUBO 问 题,并得到相应的问题哈密顿量.

2.3 绝热量子计算与 BQP 复杂类

自从 Shor⁶ 提出质因数分解的多项式算法以 来,量子计算领域获得了更广泛的关注,也涌现出 许多不同新的研究方向.然而不同于量子计算复杂 度理论、量子密码学以及量子纠错等领域的快速发 展,量子算法设计作为量子计算研究的核心问题之 一,特别是设计出相比经典算法具有指数加速的量子算法,并没有如人们想象中那样顺利推进.对于这一现象,Shor^[16]指出,一个可能的原因,是由于人们没有像设计经典算法一样好的直觉设计量子算法.而找到能充分展现量子计算机超越经典计算机能力的 BQP 问题具有十分重要的现实意义.

在计算复杂度理论中, BQP 问题是指在量子 计算机上存在多项式规模的量子线路并且出错 概率小于 1/2 求解的一类判定问题 (decision problem)^[80], 简言之, 就是能在量子计算机上在多项 式时间内求解的问题. 与其类似的经典计算问题 是 BPP(bounded-error probabilistic polynomial time)问题,它被定义为能在多项式时间内被概率 图灵机以有界的错误率求解的判定问题. 虽然在具 体问题中,如质因数分解⑤、二次符号权重计数问 题 (quadratically signed weight enumerator problem)^[81]、琼斯多项式估计 (approximation of Jones polynomials)^[82,83], local Hamiltonian 本征 值采样问题 (LHES)、相位估计采样问题 (PES)、 酉矩阵平均本征值估计 (LUAE)^[84] 等问题上量子 计算可以做到指数加速,但 BQP 计算复杂类的边 界仍然是未解决的理论问题. 在量子计算机上可以 高效解决的问题仍有待进一步探索.

前文提到,由于人们缺少量子世界观以及量子 线路模型难以提供好的算法设计直觉,那么绝热量 子算法会是寻找 BQP 问题的一条途径. 量子线路 模型被证明能以多项式量级的步骤转换为一个绝 热量子算法^[25].因此利用量子线路模型定义的 BQP 问题, 也可以等价地在绝热计算机上定义. 若 对于要求解的问题有已知的量子线路算法,那么可 以根据已知的线路模型中的一系列门操作构造出 初始哈密顿量和问题哈密顿量,使得问题哈密顿量 的基态为历史态 $\frac{1}{\sqrt{L+1}}\sum_{l=0}^{L} |\alpha(l)\rangle \otimes |1^{l}0^{L-l}\rangle$,其 中|α(l))表示原来的线路模型中每个时刻对应的逻 辑态, |1^l0^{L-l} >是标记演化时刻的时间态.于是原本 线路模型中的求解过程转换成为寻找问题哈密顿 量的基态.这个转换过程花费的步骤呈多项式增 长. 在线路模型中, 质因数分解问题 题就是一个典 型的 BQP 问题. 通过绝热量子算法判定的 BQP 问题也有典型的例子,如胶合树问题[85].如果机器 学习方法可以辅助设计绝热算法,就有望通过这种 方式找到更多的 BQP 问题, 进而探索 BQP 复杂 类的边界.

3 机器学习与量子经典组合算法

本节将首先对机器学习的几个方向及其在经 典算法设计中的应用做简要介绍.而后介绍量子与 经典组合算法以及机器学习在量子控制中的应用.

3.1 机器学习分类

1956年的达特茅斯会议中,"用机器来模仿 人类学习以及其他方面的智能"的观点被首次提 出[86]. 机器学习往往面对的是大量的数据, 通过学 习来拟合出其中的复杂关系. 我们期待机器能自行 学会数据中的关联,并能给出符合人类逻辑认知甚 至超越人类能力的预判. 近些年来, 机器学习在图 像识别[87]、自然语言处理[88]以及策略游戏[89]等方 面表现出令人称叹的能力,其中非常值得一提的就 是误差逆传播算法 (error back propagation)^[90] 在 多层神经网络中的应用. 一个多层神经网络可被分 为输入层、隐藏层和输出层,其中每一个隐藏层和 输出层的神经元中都含有激活函数 (可被激活或抑 制来模仿生物的神经元机能). 在训练时我们将信 号逐层向前传递直到输出层,而后将误差逆传递来 更新权重. 我们期待训练好的网络会有很强的表示 能力与泛化能力.也即是,对于一个完全陌生的输 入数据,网络也能给出符合预期甚至超越人类认知 的判断. 机器学习的方法主要有三大类^[91]: 监督学 习、无监督学习和强化学习.监督学习中具有代表 性的是处理"分类"和"回归"问题. 需要给机器大量 的带标签数据. 机器通过学习数据特征和标签的关 联,获得对新数据进行预测的能力.如果预测的结 果是离散的, 就属于"分类"; 如果预测的结果是连 续的, 就属于"回归". 对于无监督学习, 给机器的 是不带标签的数据,也就是希望机器能够自己发现 数据之间的共同特征,将相关的部分归为一类进行 "聚类". 强化学习 [89] 则是让智能体与环境进行有 探索地交互来训练获得最大奖励. 智能体在某一个 状态 st 下根据策略做出动作 a, 并且获得环境的奖 励 r, 到达下一个状态 st+1. 其中的会用到 Q 值来 表示在某个状态下做不同动作的未来奖励的估计. 对于像围棋这样的游戏,状态空间会随着格点个数 指数增长.考虑到要存下这么多状态空间需要大量 内存,在近期的强化学习中都采用了神经网络的强 大表示能力来表示 Q值. 在人类专家知识输入不断减少的情况下,强化学习智能体在策略游戏中依旧表现得非常出色^[92-95].

3.2 机器学习在经典算法设计中的应用

随着经典算法设计变得越来越复杂,机器学习 也被用在设计经典算法上.1976 年 Rice^[31] 就提出 了"算法选择问题",他将"算法选择问题"与"没有 免费午餐定理"^[96] 相提并论——对于任何算法, 想要其表现好于其他算法就必须付出代价.换 句话说,即没有一个普适的最好算法来解决一 大类问题.在面对拥有多种求解算法的一类问题 (特别是 NP-hard 问题)中,不同问题实例的求解 效果不尽相同.如何挑选出其中最好的算法就 显得非常关键^[97].下面通过回顾在经典领域的 自动化算法设计,期待能对量子算法的设计有一些 启发.

在早期工作中,人们通过将算法选择问题映射 为马尔科夫决策过程,利用强化学习选择算法来使 得算法运行时间最短^[98]以及并行不同算法加速求 解组合搜索问题^[99].为了预测不同算法在具体问 题求解中的所需时间,需要根据人类专家知识预先 选择出可能影响问题计算时间的特征,将一系列问 题的特征和和真实算法所需运行时间作为数据集, 通过学习利用回归方式预测每个算法在具有某些 特征的问题上求解所需时间[100,101]. 值得一提的是, 连续多年蝉联 SAT 比赛冠军的 SATzilla^[102]在处 理 3-SAT 问题时会利用预设的求解算法在短时间 内求解那些简单的问题实例. 而对于那些没有在短 时间内被求解的问题实例,其将根据问题特征来挑 选出预测的最好算法进行求解. 曾经用于分类的元 学习 (meta-learning) 也被运用到算法选择中^[103], 不同的机器学习方法在算法选择问题上的表现也 得到了评估和对比[104]. 机器学习在推荐系统 (特别 是在购物网站)中的成功应用也推动了自动化算法 推荐系统的出现[105].

与算法选择问题相应的,算法本身就具有许多 可被调整的参数.手动对大量的参数进行"调参"不 仅费时也非常依赖于专业知识.算法构型的设计^[32] 就是在高维参数空间中选择出最佳的算法构型参 数.目前已开发的一系列算法构型设计工具包^[106-109] 都可以给出优化的固定参数算法构型.但人工智 能算法在计算过程中需要不断迭代,最佳的算法 参数一般会随着整个程序运行的时间而发生变化. 为此,利用强化学习^[110]以及基于启发式算法^[111] 的动态算法构型设计框架也被提出.

算法选择问题是希望获得一个选择机制以在 面对新的问题实例时挑选出最佳的算法, 而算法构 型设计是对算法本身的参数做优化. 在经典计算算 法设计上人们也有将两者进行融合^[112] 以获得对困 难问题的高效计算.

3.3 量子与经典组合算法

量子算法的设计与研究并不是一蹴而就的. 在 研究与设计量子算法的过程中,人们也会将经典算 法中的一些思想与手段加以利用,进而设计出量子 与经典的组合算法. 一般地,这些组合算法按照形 式可以分为以下两类.

其一是利用量子系统具有的优越性来实现一 些经典算法,其中具有代表性的是量子机器学习 (quantum machine learning, QML). 量子机器学 习领域主要研究如何借助量子系统中的叠加与纠 缠等性质来实现经典机器学习算法的加速[113].在 机器学习算法中,有很多算法本质上都可以分解为 基于矩阵的一些线性代数运算.在这些线性代数 运算中,对于傅里叶变换题、寻找矩阵特征值与特 征向量[15]以及求解线性方程组[114,115]等运算,都 有着相比经典算法有指数或者多项式级别加速 的量子算法. 这一系列具有量子加速的线性代数 运算 (quantum basic linear algebra subroutines, qBLAS)^[113]可以加速许多机器学习领域中的算法, 例如最小二乘法[116]、梯度下降法与牛顿法[117]、半 正定规划^[118]、主成分分析^[119]、拓扑分析^[120]、支持 向量机[121]等.在这些机器学习算法的实现中,为 了避免经典数据的输入与读取成为限制算法效 率的瓶颈,量子随机读取内存 (quantum random access memory, gRAM)^[122] 技术被提出, 并旨在极 大地提升数据读取的效率. 量子机器学习的另一大 研究领域是利用量子模拟器或可编程量子线路以 建立量子深度学习网络 (deep quantum learning network)^[123,124]. 基于玻尔兹曼分布的量子玻尔兹 曼机 (quantum Boltzmann machine) 将神经网络 表示成为伊辛模型下量子自旋及其间的相互作用, 通过训练和优化过程使得量子系统可以学习到数 据的概率分布[125]. 相较于经典版本, 量子玻尔兹曼 机可以更有效地加速训练过程[126],同时在自旋相 互作用模型的选取上也更具灵活性^[125].除此之外, 量子机器学习不仅可以和经典机器学习一样接收 经典信息并进行处理,还可以直接处理量子系统与 量子过程产生的量子信息^[113].

其二则是将量子态的制备、演化与测量过程与 经典的优化算法相结合,利用经典计算机调节并优 化量子计算过程中的相应参数. 其中具有代表性的 算法有量子近似优化算法 (quantum approximate optimization algorithm, QAOA) 与变分量子本征 求解 (variational quantum eigensolver, VQE) 算 法. 量子近似优化算法, 最初由 Farhi 与 Goldstone^[127] 在 2014 年提出, 主要被用于解决一些 NP-hard 的组合优化问题. 一般地, 量子计算机的演化过程 可以用2p个幺正算符来描述,其中 p为预先设定 的正整数决定量子线路的深度[127]. 量子近似优化 算法利用经典的优化算法调节这些算符,进而影响 对应的量子计算过程,并通过迭代最终使演化结果 能够很好地近似对应组合优化问题的最优解.量子 近似优化算法不仅被证明具有通用计算的能力[128]. 同时还在例如连续优化[129]、线性代数[130]等领域 中的一些问题中有着良好的表现. 除此之外, 它也 被认为具有实现量子优越性的潜力[131],并且在谷 歌"悬铃木"^[132]、D-Wave 2000 Q^[133]等量子计算硬 件上表现出了良好的适配性,但是该算法的量子计 算优势还需要更准确地刻画[134].

Peruzzo 等^[135] 在 2014 年提出的变分量子本 征求解算法,则是为了解决量子化学领域的相关问 题.变分量子本征求解算法借助变分原理,通过预 先拟设 (ansatz) 来选择量子初态与量子线路,并在 量子演化后利用哈密顿量平均 (Hamiltonian averaging) 的手段估计能量期望值,最终利用经典 的非线性优化过程优化参数直至寻找到符合要求 的近似解^[136]. 尽管理论上传统求解特征值的量子 相位估计算法有着很好的性能,但它对于量子系统 的相干性有着很高的要求.相对地,变分量子本征 求解算法对于相干性的要求大大降低^[135].目前,在 不同的量子计算硬件上,变分量子本征求解算法可以 很好地求解H₂^[137]、HeH^{+[135,138]}、LiH^[139,140]、BeH₂^[139] 等分子系统的基态能量问题以及H₂^[141]等分子系 统的激发态能量问题.

3.4 机器学习在量子控制中的应用

经典最优控制理论通常需要对物理系统建立 一个数学模型,其基本目的是控制系统来根据参考 轨迹运动或者根据目标函数优化系统的动力学^[142]. 但如果这个数学模型过于复杂以至于无法解析得 到参考路径之时,那么机器学习就是一个可供选择 的方式^[143,144]. 与经典控制类似的量子控制在量子 计算与量子信息的应用中起到至关重要的作用,其 核心是控制量子动力学过程向既定的方向(比如特 殊的量子态)去演化,简单来说就是对量子系统的 控制^[27].

对于传统的贝叶斯优化,需要知道系统动力学的知识^[145].而在机器学习方法下,可以将量子系统视为一个黑箱——此时量子控制的策略会根据系统结果的输出,来近似知道对应的动力学过程^[146,147].人们可以利用机器学习在量子计算及量子测量中进行量子调控^[148-153],实现在高维量子多体系统中的非凸优化^[154,42],以及利用神经网络对控制脉冲进行设计^[155]等.

近些年,强化学习在量子系统优化控制中的应 用也备受关注.如在量子线路模型中,通过在强化 学习的环境中加入不同的控制误差来训练优化智 能体以实现普适的量子控制^[43].另外,强化学习在 实现高保真度目标态的快速制备^[45,156,157]、量子线 路优化^[158]、控制非平衡量子热力学过程^[159]以及 在量子开放系统中进行最优化控制并与传统优化 方法进行对比^[160-162],结合强化学习与量子绝热捷 径技术实现对单个量子比特进行更快更鲁棒地控 制^[163,164]等领域得到广泛应用.

机器学习 (特别是强化学习) 在有噪声的中等 规模量子 (NISQ)^[73] 控制中与传统量子优化方法, 如 GRAPE^[165]、CRAB^[166]一并成为了新的一种量 子最优化控制方法,并且能够帮助人们对自旋玻璃 物理以及对量子相变物理进行控制,辅助建立更直 观的物理图像^[42,45,167].

4 强化学习在绝热量子算法设计中的应用

本文第2节谈到绝热量子计算的定义,了解到 为了避免出现从基态向激发态的跃迁 (Landau-Zener transition)^[168,169],原则上需要给系统很长的 演化时间.在绝热量子计算中,人们通过解析局部 优化哈密顿量演化路径,使系统在最小能隙处降低 演化速率来保证不发生跃迁,并实现了 Grover 搜索问题的平方加速^[66]. 必须要指出的是,想要在复杂的量子多体系统 中做到对整个能谱的全局认知本身就非常困难.所 以对于复杂量子多体体系,很难解析地知道这些 最小能隙的位置来局域地优化哈密顿量演化路 径^[170-172].而在经典及量子最优化控制部分的介绍 中,我们已经谈到可以尝试将复杂的物理系统看作 黑箱,利用机器学习来获取最优化的控制.

本节将具体介绍我们利用强化学习辅助设计 绝热量子算法的一个工作[173]. 从前文的介绍中了 解到绝热量子计算的表现与演化路径密切相关.在 接下来的内容中,所说的绝热量子算法的设计就对 应于绝热演化的路径设计. 我们在第2节中介绍了 几个计算问题的哈密顿量编码方式.而对于给定一 个计算问题, 总有不同的问题实例. 如在 Grover 搜索问题中对不同的目标态的搜索以及在 3-SAT 问题中不同子句的选择,这都会使不同问题 实例具有不同的答案. 绝热算法设计或者说哈密顿 量演化路径设计不能依赖于具体的某一个问题实 例. 这也就有别于对具体目标量子态制备[45] 以及 实现快速的量子门操作[174,155,43],如何学习并自动 化设计绝热量子计算中哈密顿量演化路径以使得 计算过程体现出量子优势就是一个量子算法设计 问题[173,175,176]. 对此, 我们构造了自动化绝热量子 算法设计框架,如图1.这一框架特别适合对那些 很难被求解但容易被验证的问题进行绝热算法设 计,如 Grover 搜索问题、质因数分解问题、3-SAT 问题等等.在该框架中,我们参数化哈密顿量演化 路径为:

$$s\left(\frac{t}{T}\right) = \frac{t}{T} + \sum_{m=1}^{C} b_m \sin\left(\frac{m\pi t}{T}\right), \qquad (9)$$

其中 C 为截断阶数. 当 C 趋于无穷时, 这样的参数 化形式就是完备的. 强化学习中智能体 (agent) 的 状态 s 为需要设计的哈密顿量演化路径中的全部 参数 b_m,称作"路径态"(path state). 智能体的动 作 a 是对路径态中参数 b_m的操作, 其根据绝热量 子计算机的输出结果对错来获得不同奖励 r, 即答 案正确奖励为 1,答案错误奖励为 0. 强化学习的 目标是最大化奖励,所以通过让智能体从线性路径 开始对路径参数进行调整,也就能优化设计出好的 绝热量子算法. 这样的框架就非常适合在 D-wave 机器^[177]中应用. 值得一提的是, 在训练智能体的 时候,将同一系统规模的大量问题实例一起输入并 对最后的表现进行平均,这样的处理能够让算法设 计更为鲁棒.



图 1 强化学习辅助绝热量子算法设计的示意图^[173].其中 强化学习中的智能体 (agent) 根据绝热量子计算 (AQC) 输 出的结果来获取奖励,并根据深度神经网络近似表示的 *Q*值表格来选择动作更新绝热量子算法

Fig. 1. Schematic diagram of the reinforcement learning approach for quantum adiabatic algorithm design^[173]. The learning agent collects the reward according to the result obtained from adiabatic quantum computing (AQC) and produces an action to update the quantum adiabatic algorithm based on its Q table represented by a deep neural network.

智能体中深度神经网络近似地表示 Q值表格, 并用其来估计当前状态下选择不同动作的未来累 积奖励. 在例如围棋游戏中, 智能体的动作是离散 的. 而这里通过类似模拟退火的方式连续化了强化 学习智能体的动作,实现了自动化设计具有量子加 速的绝热量子 Grover 搜索算法. 其中固定系统总 的演化时间 T 与系统规模 n 的关系为T = $\sqrt{2^n}$. 人们解析地得到了基于第2节中介绍的 Grover 搜 索算法哈密顿量编码方式下的具有量子加速的绝 热算法[66]. 在这一演化时间内, 线性的演化路径会 大概率以失败告终. 而利用强化学习自动化绝热量 子算法设计框架获得的算法,其可以在这一演化时 间内到达与解析获得的算法[66]相当的结果(成功 概率 99.9% 以上), 在过程中甚至有超越解析算 法[66] 的表现, 如图 2. 通过对系统的能谱以及强化 学习得到的路径进行观察,发现演化路径在能隙最 小处变化得最缓慢,出现了平台[173].这个重要特征 与解析结果[66] 是一致的.



图 2 强化学习辅助设计的绝热量子算法在 Grover 搜索问题上的表现^[173].其中成功概率 (success probability) 是演化终态与目标态交叠的平方,总的演化时间 T与系统规模 n的关系为 $T = \sqrt{2^n}$.图中蓝色虚线表示的线性演化路径成功概率会随着系统尺寸增大不断降低.红色实线和黑色虚线分别表示强化学习设计得到的演化路径和解析获得的非线性路径^[66]的表现.在选择的演化时间下,两者的成功概率都能接近于 1,说明两者都具有平方的量子加速

Fig. 2. Performance of reinforcement learning designed quantum adiabatic algorithm in success probability for Grover search problem^[173]. The success probability is obtained by taking the square of wave-function overlap of the final evolved quantum state with the target state. The total adiabatic evolution time is chosen as $T = \sqrt{2^n}$ where *n* is the system size. The blue dashed line denotes the success probability of linear path which decreases as increasing the system size. The red solid line and black dashed line denote the performance of the reinforcement learning designed path and the nonlinear path^[66], respectively. Given the choice of total evolution time, the success probability close to 1 by both implies that they both exhibit quadratic quantum speed up. 我们测试了强化学习在量子比特数量拓展过 程中的表现,如图 3. 其中线性演化路径的结果非 保真度 (infidelity) 增长得很快,说明其计算表现 能力不佳 (前文中提到其被证明没有量子加速). 我 们测试了将在 10 qubits 系统强化学习得到的路径 直接用到 11—16 qubits 上,发现虽然保真度会有 下降但这也会比直接用线性路径更好. 而如果将 在 *i* qubits 系统中强化学习得到的路径用到*i*+1 qubits 系统中强化学习得到的路径用到*i*+1 qubits 系统并计算其非保真度. 这样拓展具有远超 线性路径的表现能力,其结果的非保真度都接近 于 1%. 对于另一种实验友好的编码方式,即如果 将初始哈密顿量写成 $H_b = \frac{1}{2} \sum_i [\mathbf{1} - \sigma_i^x], 解析的$ 方法^[66] 无法得到最优的演化路径,而基于强化学习的方式依旧可以获得这一具有平方加速的量子算法^[173].



图 3 强化学习在 Grover 搜索问题的绝热量子算法设计中的拓展性^[173].其中绿线是线性路径的表现,蓝线是将10 qubits 系统中强化学习学到的路径推广到更大系统,橘线是将在 n qubits 系统强化学习获得的路径推广到 n + 1 qubits 系统

Fig. 3. Transferability of reinforcement learning based quantum adiabatic algorithm design for Grover search problem^[173]. The green line denotes the infidelity of linear path. The blue line denotes the infidelity of the path obtained by training the 10 qubits system. The orange line denotes the performance of applying the path learned from the *n* qubits system to the n + 1 qubits system.

在对 3-SAT 这个 NP-complete 问题研究中, 我们对 10 qubits 系统且仅对包含 3 个子句的问题 进行强化学习来获得绝热量子算法.将这样设计得 到的算法直接推广到其他不同子句数量的问题上, 其表现能力与一般的线性演化路径相比具有明 显的提升.这样获得的绝热算法具备一定的可 迁移性^[173].

在这个工作中[173],我们利用强化学习优化设

计了参数 s(t).也可以将其推广,将初始哈密顿 量和问题哈密顿量前的参数在保证边界条件下 分别优化.有研究表明,这样的分别优化会有更 好的表现^[178,179].人们也考虑在演化过程中设计加 入额外的哈密顿量,并且让这些额外的哈密顿量 在初始和结束演化时关闭来提升绝热量子计算 的能力^[180,172,47].此外,强化学习在自动化设计优化 量子线路^[181–183]、完成在量子模拟中的哈密顿量 构造^[184]、优化量子纠错码^[185]、优化数字量子模 拟^[186]以及容错量子计算^[187]等量子计算方面也有 广泛应用.

5 结束语

量子计算因其具有超越经典计算的优势而受 到高度关注.其中量子计算算法的设计开发与量子 计算的硬件实现都至关重要.本文对绝热量子计 算、机器学习及其在经典算法设计中的应用做了回 顾,介绍了机器学习,特别是强化学习在量子最优 化控制中以及绝热量子计算算法设计中的具体应 用.我们看到了机器学习在设计经典算法和求解量 子多体物理上的成功应用,也期待机器学习能够对 复杂且违反经典直觉的量子算法设计提供更多帮 助.这不仅能够更好地将量子计算优势挖掘出来, 量子计算的计算优势也能更有力地加速机器学习 对大量数据的处理.我们预期量子计算与机器学习 的交融会给这两个领域带来新的契机和突破.

参考文献

- [1] Benioff P 1980 J. Statistical Phys. 22 563
- [2] Feynman R P 1982 Int. J. Theor. Phys. 21 133
- [3] Deutsch D E 1989 Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences 425 73
- Yao A C C 1993 Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science (IEEE) pp352-361
- [5] Shor P W 1994 Proceedings 35th annual symposium on foundations of computer science (IEEE) pp124-134
- [6] Ekert A, Jozsa R 1996 Rev. Mod. Phys. 68 733
- [7] Gerjuoy E 2005 Am. J. Phys. 73 521
- [8] Aumasson J P 2017 Computer Fraud & Security 2017 8
- [9] Vandersypen L M, Steffen M, Breyta G, Yannoni C S, Sherwood M H, Chuang I L 2001 Nature 414 883
- [10] Lu C Y, Browne D E, Yang T, Pan J W 2007 Phys. Rev. Lett. 99 250504
- [11] Politi A, Matthews J C, O'brien J L 2009 Science **325** 1221
- $[12]\quad {\rm Monz}\ {\rm T},\, {\rm Nigg}\ {\rm D},\, {\rm Martinez}\ {\rm E}\ {\rm A},\, {\rm et}\ {\rm al}.\ 2016\ Science\ 351\ 1068$
- [13] Ladd T D, Jelezko F, Laflamme R, Nakamura Y, Monroe C, O'Brien J L 2010 Nature 464 45
- [14] Bernstein E, Vazirani U 1997 SIAM J. Comp. 26 1411

- [15] Nielsen M A, Chuang I 2002 Quantum Computation and Quantum Information (Cambridge: Cambridge University Press)
- [16] Shor P W 2003 J. ACM (JACM) 50 87
- [17] Watrous J 2008 arXiv preprint arXiv: 0804.3401
- [18] Simon D R 1997 SIAM J. Comp. 26 1474
- [19] Hallgren S 2002 Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing pp653-658
- [20] Grover L K 1997 Phys. Rev. Lett. **79** 325
- [21] Shao C, Li Y, Li H 2019 J. Syst. Sci. 32 375
- [22] Farhi E, Goldstone J, Gutmann S, Sipser M 2000 arXiv preprint quant-ph/0001106
- [23] Farhi E, Goldstone J, Gutmann S, Lapan J, Lundgren A, Preda D 2001 Science 292 472
- [24] Zhang D J, Yu X D, Tong D 2014 Phys. Rev. A 90 042321
- [25] Aharonov D, Van Dam W, Kempe J, Landau Z, Lloyd S, Regev O 2008 SIAM Rev. 50 755
- [26] Yu H, Huang Y, Wu B 2018 Chin. Phys. Lett. 35 110303
- [27] Dong D, Petersen I R 2010 IET Control Theory & Applications 4 2651
- [28] Cirac J I, Zoller P 2012 Nat. Phys. 8 264
- [29] Georgescu I M, Ashhab S, Nori F 2014 Rev. Mod. Phys. 86 153
- [30] Gross C, Bloch I 2017 Science 357 995
- [31] Rice J R 1976 In Advances in Computers (Vol. 15) (Elsevier) pp65–118
- [32] Hutter F, Hoos H H, Leyton-Brown K, Stützle T 2009 Journal of Artificial Intelligence Research 36 267
- [33] Wang L 2016 Phys. Rev. B 94 195105
- [34] Carrasquilla J, Melko R G 2017 Nat. Phys. 13 431
- [35] Van Nieuwenburg E P, Liu Y H, Huber S D 2017 Nat. Phys. 13 435
- [36] Deng D L, Li X, Sarma S D 2017 Phys. Rev. X 7 021021
- [37] Zhang P, Shen H, Zhai H 2018 Phys. Rev. Lett. 120 066401
- [38] Gao X, Duan L M 2017 Nat. Commun. 8 1
- [39] Huang Y, Moore J E 2017 arXiv preprint arXiv: 1701.06246
- [40] Cai Z, Liu J 2018 *Phys. Rev. B* 97 035116
- [41] Carleo G, Troyer M 2017 Science 355 602
- [42] Day A G, Bukov M, Weinberg P, Mehta P, Sels D 2019 *Phys. Rev. Lett.* **122** 020601
- [43] Niu M Y, Boixo S, Smelyanskiy V N, Neven H 2019 npj Quantum Information 5 33
- [44] Zhang X M, Wei Z, Asad R, Yang X C, Wang X 2019 npj Quantum Information 5 85
- [45] Bukov M, Day A G, Sels D, Weinberg P, Polkovnikov A, Mehta P 2018 Phys. Rev. X 8 031086
- [46] Aaronson S 2015 Nat. Phys. 11 291
- [47] Albash T, Lidar D A 2018 Rev. Mod. Phys. 90 015002
- [48] Tong D 2010 Phys. Rev. Lett. **104** 120401
- [49] Amin M H 2009 Phys. Rev. Lett. 102 220401
- [50] Altshuler B, Krovi H, Roland J 2010 Proceedings of the National Academy of Sciences 107 12446
- [51] Jörg T, Krzakala F, Semerjian G, Zamponi F 2010 Phys. Rev. Lett. 104 207206
- [52] Dickson N G, Amin M 2011 Phys. Rev. Lett. 106 050502
- [53] Hen I, Young A 2011 Phys. Rev. E 84 061152
- [54] Bapst V, Foini L, Krzakala F, Semerjian G, Zamponi F 2013 *Phys. Rep.* **523** 127
- [55] Hauke P, Katzgraber H G, Lechner W, Nishimori H, Oliver W D 2020 Rep. Prog. Phys. 83 054401
- [56] Santoro G E, Martoňák R, Tosatti E, Car R 2002 Science 295 2427
- [57] Boixo S, Albash T, Spedalieri F M, Chancellor N, Lidar D A 2013 Nat. Commun. 4 1
- [58] Sherrington D, Kirkpatrick S 1975 Phys. Rev. Lett. 35 1792

- [59] Barahona F 1982 Journal of Physics A: Mathematical and General 15 3241
- [60] Kirkpatrick T R, Thirumalai D 1987 Phys. Rev. B 36 5388
- [61] Mézard M, Parisi G, Virasoro M A 1987 Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications (Vol. 9) (Singapore: World Scientific Publishing Company)
- [62] Hartmann A K, Weigt M 2005 Phase Transitions in Combinatorial Optimization Problems (Vol. 67) (Wiley Online Library)
- [63] Mezard M, Montanari A 2009 Information, Physics, and Computation (Oxford University Press)
- [64] Karp R M 1972 In Complexity of Computer Computations (Berlin: Springer) pp85–103
- [65] Lucas A 2014 Frontiers in Physics 2 5
- [66] Roland J, Cerf N J 2002 Phys. Rev. A 65 042308
- [67] Gendreau M, Potvin J Y, et al. 2010 Handbook of Metaheuristics (Vol. 2) (Berlin: Springer)
- [68] Meletiou G, Tasoulis D, Vrahatis M N, et al. 2002 In IASTED 2002 Conference on Artificial Intelligence pp483-488
- [69] Stekel A, Chkroun M, Azaria A 2018 arXiv: 1803.09237
- [70] Yampolskiy R V 2010 International Journal of Bio-Inspired Computation 2 115
- [71] Monaco J V, Vindiola M M 2017 In 2017 IEEE International Symposium on Circuits and Systems (ISCAS) pp1-4
- [72] Borders W A, Pervaiz A Z, Fukami S, Camsari K Y, Ohno H, Datta S 2019 *Nature* 573 390
- [73] Preskill J 2018 Quantum 2 79
- [74] Dash A, Sarmah D, Behera B K, Panigrahi P K 2018 arXiv: 1805.10478
- [75] Dridi R, Alghassi H 2017 Sci. Rep. 7 1
- [76] Xu K, Xie T, Li Z, et al. 2017 Phys. Rev. Lett. 118 130504
- [77] Jiang S, Britt K A, McCaskey A, Humble T, Kais S 2018 Sci. Rep. 8 17667
- [78] Wang B, Hu F, Yao H, Wang C 2020 Sci. Rep. 10 1
- [79] Peng X, Liao Z, Xu N, Qin G, Zhou X, Suter D, Du J 2008 *Phys. Rev. Lett.* **101** 220405
- [80] Bernstein E, Vazirani U 1993 ACM, New York 11
- [81] Knill E, Laflamme R 2001 Inform. Proc. Lett. 79 173
- [82] Freedman M H, Larsen M, Wang Z 2002 Commun. Math. Phys. 227 605
- [83] Freedman M H, Kitaev A, Wang Z 2002 Commun. Math. Phys. 227 587
- [84] Wocjan P, Zhang S 2006 arXiv preprint quant-ph/0606179
- [85] Somma R D, Nagaj D, Kieferová M 2012 Phys. Rev. Lett. 109 050501
- [86] Solomonoff R J 1985 Human Systems Management 5 149
- [87] Bishop C M 2006 Pattern Recognition and Machine Learning (Berlin: Springer)
- [88] Indurkhya N, Damerau F J 2010 Handbook of Natural Language Processing (Vol. 2) (Los Angeles: CRC Press)
- [89] Sutton R S, Barto A G 2018 Reinforcement Learning: An Introduction (Cambridge: MIT Press)
- [90] Rumelhart D E, Hinton G E, Williams R J 1986 Nature 323 533
- [91] Alpaydin E 2020 Introduction to Machine Learning (Cambridge: MIT Press)
- [92] Silver D, Huang A, Maddison C J, et al. 2016 Nature 529 484
- [93] Silver D, Schrittwieser J, Simonyan K, et al. 2017 Nature 550 354
- [94] Silver D, Hubert T, Schrittwieser J, et al. 2018 Science 362 1140
- [95] Schrittwieser J, Antonoglou I, Hubert T, et al. 2020 Nature 588 604

- [96] Wolpert D H, Macready W G 1997 IEEE Transactions on Evolutionary Computation 1 67
- [97] Kerschke P, Hoos H H, Neumann F, Trautmann H 2019 Evolutionary Computation 27 3
- [98] Lagoudakis M G, Littman M L 2000 ICML pp511–518
- [99] Gomes C P, Selman B 2001 Artificial Intelligence 126 43
- [100] Leyton-Brown K, Nudelman E, Shoham Y 2002 International Conference on Principles and Practice of Constraint Programming (Berlin: Springer) pp556-572
- [101] Leyton-Brown K, Nudelman E, Andrew G, McFadden J, Shoham Y 2003 In IJCAI (Vol. 3) pp1542–1543
- [102] Xu L, Hutter F, Hoos H H, Leyton-Brown K 2008 Journal of Artificial Intelligence Research 32 565
- [103] Smith-Miles K A 2009 ACM Computing Surveys (CSUR) 41 1
- [104] Kotthoff L, Gent I P, Miguel I 2012 Ai Communications 25 257
- [105] MISIT M, Sebag M 2017 Artificial Intelligence 244 291
- [106] Ansótegui C, Sellmann M, Tierney K 2009 International Conference on Principles and Practice of Constraint Programming (Springer) pp142–157
- [107] Hutter F, Hoos H H, Leyton-Brown K 2011 International Conference on Learning and Intelligent Optimization (Berlin: Springer) pp507-523
- [108] Fitzgerald T, Malitsky Y, O' Sullivan B, Tierney K 2014 Seventh Annual Symposium on Combinatorial Search
- [109] López-Ibánez M, Dubois-Lacoste J, Cáceres L P, Birattari M, Stützle T 2016 Operations Research Perspectives 3 43
- [110] Biedenkapp A, Bozkurt H F, Eimer T, Hutter F, Lindauer M 2020 Proceedings of the Twentyfourth European Conference on Artificial Intelligence (ECAI'20), Jun 2020
- [111] Speck D, Biedenkapp A, Hutter F, Mattmüller R, Lindauer M 2020 arXiv preprint arXiv: 2006.08246
- [112] Xu L, Hoos H, Leyton-Brown K 2010 Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 24)
- [113] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S 2017 Nature 549 195
- [114] Harrow A W, Hassidim A, Lloyd S 2009 Phys. Rev. Lett. 103 150502
- [115] Childs A M, Kothari R, Somma R D 2017 SIAM J. Comput. 46 1920
- [116] Wiebe N, Braun D, Lloyd S 2012 Phys. Rev. Lett. 109 050505
- [117] Rebentrost P, Schuld M, Wossnig L, Petruccione F, Lloyd S 2019 New J. Phys. 21 073023
- [118] Brandao F G, Svore K M 2017 In 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS) (IEEE) pp415-426
- [119] Lloyd S, Mohseni M, Rebentrost P 2014 Nat. Phys. 10 631
- [120] Lloyd S, Garnerone S, Zanardi P 2016 Nat. Commun. 7 1
- [121] Rebentrost P, Mohseni M, Lloyd S 2014 Phys. Rev. Lett. 113 130503
- [122] Giovannetti V, Lloyd S, Maccone L 2008 Phys. Rev. Lett. 100 160501
- [123] Adachi S H, Henderson M P 2015 arXiv preprint arXiv: 1510.06356
- [124] Denil M, De Freitas N 2011 In Neural Information Processing Systems (NIPS) Conf. on Deep Learning and Unsupervised Feature Learning Workshop (Vol. 5) (2011)
- [125] Amin M H, Andriyash E, Rolfe J, Kulchytskyy B, Melko R 2018 Phys. Rev. X 8 021050
- [126] Wiebe N, Kapoor A, Svore K M 2014 arXiv preprint arXiv: 1412.3489
- [127] Farhi E, Goldstone J, Gutmann S 2014 arXiv preprint arXiv: 1411.4028

- [128] Lloyd S 2018 $arXiv\ preprint\ arXiv:\ 1812.11075$
- [129] Verdon G, Arrazola J M, Brádler K, Killoran N 2019 arXiv preprint arXiv: 1902.00409
- [130] Borle A, Elfving V E, Lomonaco S J 2020 arXiv preprint arXiv: 2006.15438
- [131] Farhi E, Harrow A W 2016 arXiv preprint arXiv: 1602.07674
 [132] Willsch M, Willsch D, Jin F, De Raedt H, Michielsen K 2020 Quantum Information Processing 19 1
- [133] Harrigan M P, Sung K J, Neeley M, et al. 2021 Nat. Phys. 17 332
- [134] Hastings M B 2019 arXiv preprint arXiv: 1905.07047
- [135] Peruzzo A, McClean J, Shadbolt P, Yung M H, Zhou X Q, Love P J, Aspuru-Guzik A, O'brien J L 2014 Nat. Commun. 5 1
- [136] Cao Y, Romero J, Olson J P, et al. 2019 Chem. Rev. 119 10856
- [137] O'Malley P J, Babbush R, Kivlichan I D, et al. 2016 Phys. Rev. X 6 031007
- [138] Shen Y, Zhang X, Zhang S, Zhang J N, Yung M H, Kim K 2017 Phys. Rev. A 95 020501
- [139] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M, Gambetta J M 2017 Nature 549 242
- [140] Hempel C, Maier C, Romero J, et al. 2018 Phys. Rev. X 8 031022
- [141] Colless J I, Ramasesh V V, Dahlen D, et al. 2018 Phys. Rev. X 8 011021
- [142] Kirk D E 2004 Optimal Control Theory: An Introduction (Courier Corporation)
- [143] Sutton R S, Barto A G, Williams R J 1992 IEEE Control Systems Magazine 12 19
- [144] Kaelbling L P, Littman M L, Moore A W 1996 J. Artificial Intell. Res. 4 237
- [145] Wiseman H M, Mancini S, Wang J 2002 Phys. Rev. A 66 013807
- [146] Guță M, Kot lowski W 2010 New J. Phys. 12 123032
- [147] Magesan E, Gambetta J M, Córcoles A D, Chow J M 2015 *Phys. Rev. Lett.* **114** 200501
- [148] Hentschel A, Sanders B C 2010 Phys. Rev. Lett. 104 063603
- [149] Lovett N B, Crosnier C, Perarnau-Llobet M, Sanders B C 2013 Phys. Rev. Lett. 110 220501
- [150] Tiersch M, Ganahl E, Briegel H J 2015 Sci. Rep. 5 1
- [151] Banchi L, Pancotti N, Bose S 2015 arXiv preprint arXiv: 1509.04298
- [152] Wigley P B, Everitt P J, van den Hengel A, et al. 2016 Scientific Reports 6 1
- [153] August M, Ni X 2017 Phys. Rev. A 95 012335
- [154] Palittapongarnpim P, Wittek P, Zahedinejad E, Vedaie S, Sanders B C 2017 Neurocomputing 268 116
- [155] Yang X C, Yung M H, Wang X 2018 Phys. Rev. A 97 042324
- [156] He R H, Wang R, Wu J, Nie S S, Zhang J H, Wang Z M 2020 arXiv preprint arXiv: 2012.00326
- [157] Ma H, Dong D, Ding S X, Chen C 2020 arXiv preprint arXiv: 2012.15427
- [158] Fösel T, Niu M Y, Marquardt F, Li L 2021 arXiv preprint arXiv: 2103.07585
- [159] Sgroi P, Palma G M, Paternostro M 2021 Phys. Rev. Lett. 126 020601
- [160] An Z, Song H J, He Q K, Zhou D 2021 Phys. Rev. A 103 012404
- [161] Dong D 2021 arXiv preprint arXiv: 2101.07461
- [162] Xu H, Li J, Liu L, Wang Y, Yuan H, Wang X 2019 npj Quant. Inform. 5 82
- [163] Ding Y, Ban Y, Martín-Guerrero J D, Solano E, Casanova J, Chen X 2021 Phys. Rev. A 103 L040401

- [164] Ai M Z, Ding Y, Ban Y, Martín-Guerrero J D, Casanova J, Cui J M, Huang Y F, Chen X, Li C F, Guo G C 2021 arXiv preprint arXiv: 2101.09020
- [165] Khaneja N, Reiss T, Kehlet C, Schulte-Herbrüggen T, Glaser S J 2005 J. Magnetic Resonance 172 296
- [166] Caneva T, Calarco T, Montangero S 2011 Phys. Rev. A 84 022326
- [167] Guo S F, Chen F, Liu Q, Xue M, Chen J J, Cao J H, Mao T W, Tey M K, You L 2020 arXiv preprint arXiv: 2011.11987
- [168] Landau L D 1932 Phys. Z. Sowjetunion 2 19
- [169] Zener C 1932 Proceedings of the Royal Society of London (Series A) 137 696
- [170] Morita S 2007 J. Phys. Soc. Japn. 76 104001
- [171] Avron J, Fraas M, Graf G, Grech P 2010 Phys. Rev. A 82 040304
- [172] Zeng L, Zhang J, Sarovar M 2016 J.Phys. A: Mathematical and Theoretical 49 165305
- [173] Lin J, Lai Z Y, Li X 2020 Phys. Rev. A 101 052327
- [174] Chen X, Lizuain I, Ruschhaupt A, Guéry-Odelin D, Muga J 2010 Phys. Rev. Lett. 105 123003
- [175] Chen Y Q, Chen Y, Lee C K, Zhang S, Hsieh C Y 2020 arXiv preprint arXiv: 2004.02836

- [176] Yang X, Liu R, Li J, Peng X 2020 Phys. Rev. A 102 012614
- [177] Https://www.dwavesys.com [2021-5-1]
- [178] Rezakhani A, Kuo W J, Hamma A, Lidar D, Zanardi P 2009 Phys. Rev. Lett. 103 080502
- [179] Rezakhani A, Pimachev A, Lidar D 2010 Phys. Rev. A 82 052305
- [180] Farhi E, Goldstone J, Gosset D, Gutmann S, Meyer H B, Shor P 2009 arXiv preprint arXiv: 0909.4766
- [181] McKiernan K A, Davis E, Alam M S, Rigetti C 2019 arXiv preprint arXiv: 1908.08054
- [182] Zhang Y H, Zheng P L, Zhang Y, Deng D L 2020 Phys. Rev. Lett. 125 170501
- [183] Ostaszewski M, Trenkwalder L M, Masarczyk W, Scerri E, Dunjko V 2021 arXiv preprint arXiv: 2103.16089
- [184] Peng P, Huang X, Yin C, Joseph L, Ramanathan C, Cappellaro P 2021 arXiv preprint arXiv: 2102.13161
- [185] Nautrup H P, Delfosse N, Dunjko V, Briegel H J, Friis N 2019 Quantum 3 215
- [186] Bolens A, Heyl M 2020 arXiv preprint arXiv: 2006.16269
- [187] Sweke R, Kesselring M S, van Nieuwenburg E P, Eisert J 2020 Machine Learning: Science and Technology 2 025005

SPECIAL TOPIC—Machine learning and physics

Machine learning assisted quantum adiabatic algorithm design^{*}

Lin Jian Ye Meng Zhu Jia-Wei Li Xiao-Peng[†]

(Department of Physics, Fudan University, Shanghai 200433, China)

(Received 1 May 2021; revised manuscript received 13 June 2021)

Abstract

Quantum computing has made dramatic progress in the last decade. The quantum platforms including superconducting qubits, photonic devices, and atomic ensembles, have all reached a new era, with unprecedented quantum control capability developed. Quantum computation advantage over classical computers has been reported on certain computation tasks. A promising computing protocol of using the computation power in these controllable quantum devices is implemented through quantum adiabatic computing, where quantum algorithm design plays an essential role in fully using the quantum advantage. Here in this paper, we review recent developments in using machine learning approach to design the quantum adiabatic algorithm. Its applications to 3-SAT problems, and also the Grover search problems are discussed.

Keywords: adiabatic quantum computation, quantum algorithm, quantum simulation, machine learningPACS: 03.67.Ac, 03.67.Lx, 89.70.Eg, 07.05.MhDOI: 10.7498/aps.70.20210831

^{*} Project supported by the National Natural Science Foundation of China (Grant No. 11934002), the National Basic Research Program of China (Grant No. 2017YFA0304204), and the Shanghai Municipal Science and Technology Major Project, China (Grant No. 2019SHZDZX01).

 $[\]dagger$ Corresponding author. E-mail: xiaopeng_li@fudan.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

量子态制备及其在量子机器学习中的前景

赵健 陈昭昀 庄希宁 薛程 吴玉椿 郭国平

Quantum state preparation and its prospects in quantum machine learning Zhao Jian Chen Zhao-Yun Zhuang Xi-Ning Xue Cheng Wu Yu-Chun Guo Guo-Ping 引用信息 Citation: Acta Physica Sinica, 70, 140307 (2021) DOI: 10.7498/aps.70.20210958 在线阅读 View online: https://doi.org/10.7498/aps.70.20210958 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

基于混合光模式阵列的自由空间编码通信

Free-space optical communication based on hybrid optical mode array encoding 物理学报. 2017, 66(14): 144102 https://doi.org/10.7498/aps.66.144102

基于机器学习构建的环三亚甲基三硝胺晶体势

Energetic potential of hexogen constructed by machine learning 物理学报. 2020, 69(23): 238702 https://doi.org/10.7498/aps.69.20200690

铅基钙钛矿铁电晶体高临界转变温度的机器学习研究

High critical transition temperature of lead-based perovskite ferroelectric crystals: A machine learning study 物理学报. 2019, 68(21): 210502 https://doi.org/10.7498/aps.68.20190942

基于量子算法的量子态层析新方案

A novel scheme of quantum state tomography based on quantum algorithms 物理学报. 2019, 68(14): 140301 https://doi.org/10.7498/aps.68.20190157

一种编码式宽带多功能反射屏

A wideband coding reflective metasurface with multiple functionalities 物理学报. 2017, 66(6): 064203 https://doi.org/10.7498/aps.66.064203

基于鲁棒极端学习机的混沌时间序列建模预测

Chaotic time series prediction based on robust extreme learning machine 物理学报. 2018, 67(3): 030501 https://doi.org/10.7498/aps.67.20171887

专题: 机器学习与物理

量子态制备及其在量子机器学习中的前景*

赵健1) 陈昭昀1) 庄希宁1)3) 薛程1) 吴玉椿1)2)† 郭国平1)2)3)

(中国科学技术大学,中国科学院量子信息重点实验室,合肥 230026)
 2)(合肥综合性国家科学中心人工智能研究院,合肥 230088)
 3)(合肥本源量子计算科技有限责任公司,合肥 230026)

(2021 年 5 月 21 日收到)

经典计算机的运算能力依赖于芯片单位面积上晶体管的数量,其发展符合摩尔定律.未来随着晶体管的 间距接近工艺制造的物理极限,经典计算机的运算能力将面临发展瓶颈.另一方面,机器学习的发展对计算 机的运算能力的需求却快速增长,计算机的运算能力和需求之间的矛盾日益突出.量子计算作为一种新的计 算模式,比起经典计算,在一些特定算法上有着指数加速的能力,有望给机器学习提供足够的计算能力.用量 子计算来处理机器学习任务时,首要的一个基本问题就是如何将经典数据有效地在量子体系中表示出来.这 个问题称为态制备问题.本文回顾态制备的相关工作,介绍目前提出的多种态制备方案,描述这些方案的实 现过程,总结并分析了这些方案的复杂度.最后对态制备这个方向的研究工作做了一些展望.

关键词:态制备,量子机器学习,编码 PACS: 03.67.Ac, 03.67.Lx, 03.67.-a

DOI: 10.7498/aps.70.20210958

1 引 言

机器学习是一门人工智能领域的科学,其通过 计算机学习训练已知的数据,并利用训练好的数据 模式预测未知数据的信息.随着计算机性能的不断 增强,机器学习对数据的处理能力也不断提升,被 广泛应用到各个领域^[1].这包括图像识别^[2,3]、人脸 识别^[4–6]等分类问题,也包括最优决策问题,如 Alpha Go^[7], Alpha Zero^[8]的围棋对弈等.经典数 据有许多处理和训练方式,如神经网络、聚类等方 法.为了准确提取未知数据的特征,训练方式的选 择需要参考相应的数据类型.当处理大规模的数据 时,为了获取数据特征,往往采取深度学习的学习 方式,如包含数十亿权重的神经网络^[9],这充分展 示了深度学习在处理大数据时的效果.

当今的机器学习发展,特别是在大数据的处理 方面,对经典计算机的运算能力有很高的需求. 1965年戈登·摩尔提出摩尔定律,指集成电路上可 容纳的元器件数目约每两年增加一倍.一方面,在 不久的将来随着晶体管在芯片上的间距接近1 nm, 接近传统工艺制造的物理极限;另一方面数据的爆 炸式增长,对算力需求越来越高.于是为了应对大 数据的处理, 需要一个创新的计算体系结构. 量子 计算作为一种新的计算模型,比起经典计算,在一 些特定算法上有着指数加速的能力,有望为大数据 的处理提供足够的计算需求. 如果一个量子信息计 算处理器能够产生经典计算机难以模拟的统计模 式,那么量子计算与机器学习结合便可能识别经典 机器学习难以识别的特征.为此,人们将量子运算 和经典的机器学习相结合,提出了机器学习的量子 版本,称为量子机器学习,并将这种寄希望于量子

© 2021 中国物理学会 Chinese Physical Society

^{*} 国家重点研究发展计划 (批准号: 2016YFA0301700)、国家自然科学基金 (批准号: 11625419)、安徽省量子信息技术倡议 (批准 号: AHY080000) 和中国科学院战略重点研究计划 (批准号: XDB24030600) 资助的课题.

[†] 通信作者. E-mail: wuyuchun@ustc.edu.cn

机器学习的优势称为量子计算在经典机器学习中 潜在的加速能力^[10].量子机器学习包括用经典机 器学习的方法处理量子物理中的问题和用量子计 算的方式解决经典机器学习的问题.前者需要将量 子物理中的量子态转换为经典数据,再用经典机器 学习的方法来提取数据信息,如构造经典神经网络 训练这些经典数据后,得到某些量子态的特征.后 者在处理经典数据时,某些步骤中的计算过程可以 通过量子态的酉变换来辅助实现,这其中不可避免 地需要将经典数据对应成量子态.

量子机器学习中,需要运用量子计算机处理经 典数据,这涉及经典数据的在量子体系中的表示问 题.这种将经典数据映射到量子计算机中的过程, 称为态制备问题^[11,12].态制备的种类有很多,大部 分是将经典数据转换为了量子态,也存在一些将经 典数据映射到哈密顿量的方式.态制备种类的选择 直接影响了执行机器学习算法的选择,这意味着不 同的态制备方法决定了提取经典数据信息的差异, 影响了后续在量子系统里的操作,影响整个机器学 习算法的计算复杂度.同时,态制备作为量子机器 学习的其中一部分,其制备精度和成功率会影响整 个机器学习算法的有效性.

态制备问题不受限于机器学习的应用, 它同样 是一些算法的基础,如解线性方程组的 HHL 量子 算法[13]. 基于解线性方程组的量子算法, 有量子主 成分分析算法[14],可用于聚类和特征识别;也有支 持向量机算法^[15],用于对大规模的数据分类问题. 这类量子算法的共同点都是为了解决实际的经典 问题, 需要以经典数据为输入和输出. 这可以分为 三个步骤:首先运用态制备将经典数据转为量子 态,再用量子计算机对量子态进行酉变换,最后多 次地量子测量概率性得到一个经典结果. 整个算法 的复杂度受各个步骤的影响,本文仅列出不同态制 备方式的复杂度. 如果考虑量子算法的复杂度, 可 通过量子线路的语言,对所需的基本量子操作,即 基本量子门计数得到所有门的个数. 类比于经典算 法的分类方式,量子算法分为不含黑箱 (oracle) 的 显式算法和含黑箱的算法. 前者的复杂度指的是所 有基本量子门个数,后者往往勿略黑箱的执行时间 而考虑黑箱的执行次数,称为质询复杂度.一般地, 若数据规模是O(N)的,量子基本门的时间层数是 $\mathcal{O}(\operatorname{Poly}(\log N))$ 的,称量子算法的执行时间是有效 的. 在每个时间层, 允许多个量子比特同时执行一次量子基本门. 同样的数据规模, 若用到的量子比特是 *O*(Poly(log *N*))的, 称量子算法的比特数是有效的. 量子基本门的个数受量子比特数和时间层数的影响, 在一个时间层至多有量子比特数的量子门同时执行, 故显式算法的复杂度上界为量子基本门的时间层数与比特数的乘积.

从经典计算机到量子系统态制备的方式叫作 编码. 编码的种类大体上可以分为三种, 分别是基 底编码、振幅编码和量子抽样编码[12]. 基底编码用 于处理二值数据向量,将数据编码到量子态的基底 上;振幅编码是最为常见的态制备方式,将数据编 码到量子态的振幅上,数据向量可以是连续变量, 数据特征信息体现到量子态的振幅大小;量子抽样 编码可以看成前两种编码的结合,是对在整个计算 基基底的经典概率分布进行振幅编码,对于某个给 定的经典概率分布,量子抽样编码退化为了振幅编 码. 上述由经典数据编码到量子态的过程, 在量子 系统中也可以视为从初态到目标量子态的一种酉 变换. 广义上讲, 可以称从经典数据到酉变换的过 程为编码,如由经典数据决定量子系统演化哈密顿 量的方式也可以看成一种编码,这种编码称为哈密 顿量编码.

态制备中振幅编码的相关工作最为丰富,除了 平凡的编码方式,振幅编码可以从 2002 年 Grover 和 Rudolph^[16]的工作谈起,其将满足条件可积的 一种数据分布制备成了量子态,制备过程依赖于经 典函数的有效计算,且没有给出量子线路语言,编 码的有效性需进一步探讨.Kaye 等^[17]以类似的方 式得到了任意量子态的制备,给出了可称之为含黑 箱的量子线路.Soklakov和 Schack^[11]于 2005 年 用其他形式的黑箱给出了在一定限制条件下的有 效的概率性算法.振幅编码中不得不提到量子随机 存取存储器 (quantum random access memory, QRAM)的方法,这是一种从已知量子态出发,由 经典数据直接得到新的量子态的过程.

本综述简要叙述各种态制备的编码方式,并给 出一些简单的例子.根据各个编码方式的适用情 景,对不同编码进行比较,列出态制备的复杂度, 表明应谨慎乐观对待量子态制备问题.

符号说明 文中希尔伯特空间用 \mathcal{H} 表示,任意 的单位化向量 $|\psi\rangle \in \mathcal{H}$ 表示量子态,其中 $|0\rangle = (1,0)^{T}$,

2 编码方式

这里给出态制备的问题模型. 对给定的经典数 据, 不妨假定数据集 $\mathcal{D} \subset \mathcal{R}^n$ 是有限集, |D| = l, 每 个数据 $x = (x_1, \dots, x_n) \in \mathcal{D}$, 用一个单射 f将 D 的 所有子集构成的集合, 记为 $2^{\mathcal{D}}$, 映射到某个希尔伯 特空间 \mathcal{H}^m , 使得对 $\mathcal{D}' \subset \mathcal{D}$, $f(\mathcal{D}') \in \mathcal{H}^m$. 称 f 为 态制备, 其中 \mathcal{H}^m 中的元素都视为单位向量, 对应 量子态. 例如, $\mathcal{D} = \{x_1 = (1,0), x_2 = (0,1)\}$, 可以 找到一种态制备的映射, 使得 $f(\{x_1\}) = |10\rangle$, $f(\{x_2\}) = |01\rangle$, $f(\mathcal{D}) = 1/\sqrt{2}(|10\rangle + |01\rangle)$.

2.1 基底编码

这类编码中,限定所有数据是二元向量,或二 值化处理后的经典数据是二元向量,即 $\mathcal{D} \subset \{0,1\}^n$. 对任意数据集的子集 $\mathcal{D}' \subset \mathcal{D}, \mathcal{D}' = \{x^j\}_{j=1}^{l'}, x^j = (x_1^j, \cdots, x_n^j),$

$$f(\mathcal{D}') = \frac{1}{\sqrt{l'}} \sum_{j=1}^{l'} |x^j\rangle = \frac{1}{\sqrt{l'}} \sum_{j=1}^{l'} |x_1^j, \cdots, x_n^j\rangle.$$

这种编码方式将数据集中的所有数据,编码到量子 态的计算基上,等权叠加.制备过程中用到的量子 比特数为*O*(*n*).制备的思路是运用迭代法.

1) 制备 $|\psi_1\rangle = \frac{1}{\sqrt{l'}} |x^1\rangle |0\rangle_1^{\otimes n} |0\rangle_2 + \sqrt{\frac{l'-1}{l'}} |0\rangle^{\otimes n}$ $|0\rangle_1^{\otimes n} |1\rangle_2$. 注意到这里增加了两个寄存器,其中第 一个寄存器为了存放 $|x^2\rangle$,第二个寄存器放置辅助 比特标记位置.

2) 制备 $|\psi'_1\rangle = \frac{1}{\sqrt{l'}} |x^1\rangle |0\rangle_1^{\otimes n} |0\rangle_2 + \sqrt{\frac{l'-1}{l'}} |x^2\rangle^{\otimes n} |x^2\rangle_1^{\otimes n} |1\rangle_2.$ 这里的操作是一系列并行的 *X* 门和单比特受控门.

3) 得 到 $|\psi_{12}\rangle = \frac{1}{\sqrt{l'}}(|x^1\rangle + |x^2\rangle)|0\rangle_1^{\otimes n}|0\rangle_2 + \sqrt{\frac{l'-2}{l'}}|0\rangle^{\otimes n}|0\rangle_1^{\otimes n}|1\rangle_2.$ 这里的操作涉及对上一步 $3|x^2\rangle$ 的态进行受控 Y操作, $3|x^2\rangle$ 态的振幅也成 $51/\sqrt{l'}$, 再进行退计算的操作还原即可得到.

以同样的迭代方法,可以得到 $|\psi_{1...l'}\rangle$,即得到 f(D').

2.2 振幅编码

这类编码要求的数据不再是二元向量,可以是 任意实数.对于任何 $D' = \{x^j\}_{j=1}^{l'}, x^j = (x_1^j, \cdots, x_n^j),$

$$f(\mathcal{D}') = \frac{1}{C} \sum_{j=1}^{l'} \sum_{i=1}^{n} x_i^j |i, j\rangle,$$

这里 *C*为归一化常数, $C = \sqrt{\sum_{j=1}^{l'} \sum_{i=1}^{n} (x_i^j)^2}$. 可以看出, 如果对数据集中所有数据振幅编码, 当 *ln*是 2的幂次时, 只需要 log(*ln*)个量子比特便可以 编码*ln*个振幅. 例如, *ln* = 4 时, 只需制备一个 2 bit 的量子态, 使得在四个计算基 |00>, |01>, |10>和 |11> 上的振幅为数据大小即可. 振幅编码问题可以简化 为, 给定一个单点集合 $X = \{x = (x_0, \dots, x_{N-1})\} \subset \mathcal{R}^N, N$ 为 2的幂次, 使得在忽略归一化 常数的条件下

$$f(X) = \sum x_i |i\rangle.$$

2.2.1 显式的编码

1) 用 log N 个量子比特编码.

基本的想法是利用迭代法,用部分量子态对新 粒子多重控制操作,直到全部粒子完成态制备.这 个算法的执行时间是 $\mathcal{O}(N)$.假定制备出的量子 态的每个振幅的大小已知,即每个计算基上测量得 到相应的结果的概率 $p_i^{\log N} = x_i^2/||x||^2$ 已知,并且我 们定义边际概率 $p_i^k = \sum_{i'=2i,2i+1} p_{i'}^{k+1}, k = 1, \cdots,$ log N - 1,如图 1 所示.



图 1 当 N = 8时所有的边际概率 Fig. 1. The marginal probabilities for N = 8.

假定 $\theta_i^k = 2 \arccos \sqrt{p_{i/2}^{k+1}/p_i^k}$ ($k = 0, \dots, n-1$; $i = 0, \dots, 2^k - 1$), 则态制备的整个迭代流程图可 以参看图 2.

$$f(X) = \sum_{i=0}^{N-1} x_i |i\rangle$$



图 2 N = 8时用 $\mathcal{O}(N)$ 的时间制备f(X)Fig. 2. Preparation for f(X) in $\mathcal{O}(N)$ time for N = 8.

在用 log N 个量子比特编码时,每个基底前的 振幅都不能并行运算,导致了这个方法的运行时间 为 O(N).如果这些多重受控操作可以并行操作, 运行时间将大大降低.

2) 用 N个量子比特编码. 基于减少运行时间 的考量,可以增加量子比特,使得编码振幅的基底 选择性更多,从而增加并行运算的可行性. 这里选 取 W态的基底,得到态制备的映射为

$$f(X) = x_0 |0 \cdots 01\rangle + \dots + x_{N-1} |10 \cdots 0\rangle$$
$$= \sum_{i=1}^{N} x_i |2^i\rangle.$$
$$Y(\theta) = \begin{bmatrix} 1 & & \\ & \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ & \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \\ & & & 1 \end{bmatrix} \triangleq \underbrace{-}_{R_Y(\theta)}$$

솏

可以注意到 $Y(\theta)|01\rangle = \cos(\theta/2)|01\rangle + \sin(\theta/2)|10\rangle$, $Y(\theta)|10\rangle = -\sin(\theta/2)|01\rangle + \cos(\theta/2)|10\rangle$,这类似于 对单比特量子门 $R_Y(\theta)$ 在 $|0\rangle$ 和 $|1\rangle$ 上的操作,故将 $Y(\theta)$ 定义为由符号"×"控制的 $R_Y(\theta)$ 量子门.这里 直接给出N = 8时的整个迭代过程,详见图 3.



图 3 N = 8时用 $\mathcal{O}(\log N)$ 的时间获取振幅 Fig. 3. Acquiring the amplitudes in $\mathcal{O}(\log N)$ time for N = 8.

2.2.2 含黑箱的编码

这类编码不考虑黑箱构造的问题,有两大类制备方案.

I) Grover 等^[16]和 Kaye 等^[17]的工作

Grover 在 2002 年提出将满足条件可积的经 典数据制备成量子态的方法. 给定一个离散概率分 布 $\{p_i\}_{i=0}^{N-1}$, 目标制备 $|\psi\rangle = \sqrt{p_i}|i\rangle$. 等价于给定单 点集 $\{x = (x_0, \dots, x_{N-1}), x_i \ge 0\}$, 满足归一化条 件 $\sum_i x_i^2 = 1$. 制备的思路与显式编码相同, 都是 运用迭代法, 为了描述方便, 仍然采用边际概率的 记号 $p_i^k = \sum_{i'=2i, 2i+1} p_{i'}^{k+1}$, $k = 1, \dots, \log N - 1$.

1) 制备 $|\psi_1\rangle = \sqrt{p_0^1}|0\rangle + \sqrt{p_1^1}|1\rangle$. 这一步操作 只需要一个单比特门即可.

2) 制备 $|\psi_2\rangle = \sqrt{p_0^2}|00\rangle + \sqrt{p_1^2}|01\rangle + \sqrt{p_2^2}|10\rangle + \sqrt{p_3^2}|11\rangle$. 该步骤的思路是运用迭代法,具体来讲, 首先,令 $f(0) = \frac{p_0^2}{p_0^2 + p_1^2}$, $f(1) = \frac{p_2^2}{p_2^2 + p_3^2}$ 表示条件 概率,在 1) 的基础上加上一个寄存器,用来存储条 件概率,即

$$\sqrt{p_i^1}|i\rangle|0
angle o \sqrt{p_i^1}|i\rangle|\theta_i
angle,$$
 (1)

这里 $\theta_i = \arccos \sqrt{f(i)}$.这一步操作并未指明寄存 器量子比特的数量与执行时间.可以视为以f为黑 箱的一步操作.原文中该步骤的有效实现需要f为 条件可积函数,并要求在经典计算中对f这步黑箱 操作是有效计算的.那么从经典计算机到量子计算 机,通过辅助量子比特将经典计算转变为可逆计算 并退计算等一系列操作,一定存在时间复杂度与经 典运算的时间复杂度相同的量子线路,使得该步骤 可以有效实现^[18].接着,再增加一个寄存器,受控 于 θ_i 的操作,得到 $|\psi_2\rangle$ 的振幅信息,即

$$\sqrt{p_i^1}|i\rangle|\theta_i\rangle|0\rangle \to \sqrt{p_i^1}|i\rangle|\theta_i\rangle(\cos\theta_i|0\rangle + \sin\theta_i|1\rangle)$$

之后进行退计算操作,将 θ_i 擦除,这步操作的量子 比特数量与执行时间及存储 θ_i 相同,同样是含 f的 黑箱操作,得到 $|\psi_2\rangle$.

运用迭代法,由 $|\psi_k\rangle$ 得到 $|\psi_{k+1}\rangle$,最终得到 $|\psi_{\log_0 N}\rangle$,即目标量子态 $|\psi\rangle$.

Kaye 等^[17] 的态制备方法与 Grover 类似, 给 出了量子线路的语言. 其中对存储 θ_i 的步骤进行细 化, 在已知 $|\psi\rangle$ 的前提下, 将 (1) 式改写为

$$|\psi\rangle\sqrt{p_i^1}|i\rangle|0\rangle \to |\psi\rangle\sqrt{p_i^1}|i\rangle|\theta_i\rangle, \qquad (2)$$

这步黑箱操作表明θ_i的获得需要整个态|ψ⟩的各个 分量的值,并未给出黑箱操作的具体构造,在这一 点上与 Grover 的算法没有本质区别.

评价含黑箱的算法复杂度,通常不考虑黑箱以 外的线路,这是由于黑箱的结构相比于显式的量子 线路更为复杂. 如果同时考虑黑箱内部的执行时间 和黑箱外的量子门执行时间,对于任意 N 规模的 量子态,是不可能用*O*(*n*)的量子比特在有效时间 完成的.因此,我们往往考虑黑箱的执行次数,称 为质询复杂度,以此来衡量含黑箱算法的计算复杂 度. Kaye 的算法对于任意的量子态都可以制备, 并且从含黑箱的角度看出是以(2)式为黑箱的振 幅编码,该编码方式具有有效的质询复杂度,不过, 值得说明的是 Grover 和 Kaye 的算法原文中并没 有指明是含黑箱的算法. 给定数据集 X, 制备过程 可以视为含黑箱的量子算法. 若是未指明某个数据 集上的经典数据,对数据集中的元素随机化处理, 如数据集中的元素满足某种概率分布函数 g, 对这 种分布的态制备问题可能是有效的,因为 q 的参数 可能不依赖于 n. 这种含黑箱的编码比较广泛, 将 在 2.3 节的量子抽样编码中再次提及.

II) Soklakov 和 Schack^[11] 的工作

真正意义上经典的含黑箱的振幅编码可参看 Soklakov 等的工作. 这类算法属于概率性的量子 算法, 态制备给出了理想态的近似量子态. 数据向 量 不 局 限 于 实 空 间 , 即 $X = \{x = (x_0, \dots, x_{N-1})\} \subset C^N, N$ 为2的幂次, 这里 $x_i = |x_i| e^{i\alpha_i}, \alpha_i \in [0, 2\pi), \ (|x_i|$ 不可以全相等. 理想的量子态为

$$f(X) = \frac{1}{\|x\|} \sum_{i=0}^{N-1} x_i \mathrm{e}^{\mathrm{i}\alpha_i} |i\rangle \triangleq |\psi\rangle.$$
(3)

该编码的执行时间受限于两个因素.一方面是数据 集本身 x各个分量实部的差异,如果各个分量|x_i| 大小都比较接近,那么编码执行时间会很快.另一 方面是对制备量子态结果保真度和成功率的要求, 如果对态制备的结果要求严苛,会导致执行时间变 慢.令 $\gamma, \lambda, \eta \in (0,1)$,如果对任意的数据分量 $x_i^2 < 1/\eta N$,以不小于 $1 - \gamma$ 的成功率制备的近似 态为 $|\tilde{\psi}\rangle$,满足 $|\langle \tilde{\psi} | \psi \rangle| = 1 - \lambda$,所需要的计算复杂 度为 $\mathcal{O}(P(\log_2 N \cdot \gamma^{-1} \lambda^{-1} \eta^{-1})).$

算法的核心内容是选择合适的黑箱,对获取目标量子态所有分量振幅的大小做分割,并从振幅大分量向振幅小的分量标记,最终用 Grover 搜索算法,将目标态的近似态以一定的成功率找到.

2.2.3 QRAM

量子随机存取存储器 (QRAM) 是类比于经典

内存存取数据的一种装置,可以将经典数据存储到 相干的量子态各个分量地址中.在读取量子态的任 意一个分量时,每个分量地址上都需要附带经典数 据的信息:

$$\sum_{i} \psi_{i} |i\rangle \to \sum_{i} \psi_{i} |i\rangle |D_{i}\rangle.$$
(4)

QRAM 存取数据的过程中, 第一个寄存器存储经 典数据作为指标,要求对任意量子态 $\sum \psi_i |i\rangle$,分 量都需要存储经典数据地址信息. 第二个寄存器是 数据寄存器,用于存储经典数据 X. 这种装置类似 于(1)式,(2)式的操作方式,故一定程度上,QRAM 的模型包含了 Grover 和 Kaye 等的态制备工作. 例如在量子推荐系统算法中^[19],概率分布被提前 储存到 QRAM 中, Grover 的算法也可以实现. 理 论上, QRAM 的模型可以通过增加大量的比特 数来减小执行时间. QRAM 通过二分的树状图和 桶队结构 (bucket-brigade) 来实现, 这种实现方 式可以做到O(n)的时间复杂度,但量子比特数是 O(N)的. QRAM 的量子线路语言实现方式种类较 多^[20,21]. 人们在后续的工作中更关心哪一种 QRAM 的实现方式更具有噪声的抗性和可拓展性, Hann^[22] 给出了一种关于噪声抗性的论证.

(1) 式, (2) 式以及 QRAM 的直接形式 (4) 式 都是将经典数据存储到辅助比特上,每个分量对应 的辅助比特上都有经典数据的信息.特别地,如果 辅助比特可以写成二进制数,这种变换称为数字 编码.与之对应的, Mitarai 称振幅编码为模拟编 码^[23],并介绍了数字编码和模拟编码的转换关系. 利用这种数模转换的关系,可以得到振幅编码的具 体形式.

具体来讲,如果 QRAM 的操作完成后,振幅 编码可以通过条件受控和后选择的方式得到振幅 编码的概率性量子算法.给定数据集 *X*, QRAM 可 以将 $x = (x_0, \cdots, x_{N-1})$ 存储到等权叠加的量子态 上,忽略归一化,得到

$$\sum_{i=0}^{N-1} |i\rangle|0\rangle \to \sum_{i=0}^{N-1} |i\rangle|x_i\rangle.$$
(5)

进行条件受控操作,通过增加辅助比特和受控操作 实现,

$$\sum_{i=0}^{N-1} |i\rangle |x_i\rangle \to \sum_{i=0}^{N-1} |i\rangle |x_i\rangle (x_i|0\rangle + \sqrt{1 - |x_i|^2}|1\rangle), \quad (6)$$

(6) 式中的受控旋转操作可以将*x_i*表示为*t*比特的 二进制数,分别对辅助比特做控制*R_Y*(π/2^t)类似的 操作来实现.最后一步进行后选择的操作,对辅助 比特进行测量,当测量到^{|1})态时,制备失败.需要 重复这个算法的流程,直到测量值^{|0}/. 当测量值为 ^{|0}/态时,成功制备,成功率可以计算,得到

$$\sum_{i=0}^{N-1} x_i |i\rangle |x_i\rangle.$$
(7)

(7) 式得到的量子态与目标态还多了数据寄存器的数据,需要擦除.这步擦除数据是退计算的过程,
也是 QRAM 里 (5) 式的逆操作,即

$$\sum_{i=0}^{N-1} x_i |i\rangle |x_i\rangle \to \sum_{i=0}^{N-1} x_i |i\rangle.$$
(8)

2.3 量子抽样编码

本节介绍基底编码和振幅编码的一种混合编 码方式——量子抽样编码.振幅编码的数据集 *X*是单点集, *X* = {*x* = (*x*₀, · · · ,*x*_{*N*-1})},如果用log*N* 个量子比特,时间复杂度为O(N).在量子抽样编 码中,给定一种概率分布,不妨假定为 $g(x'), x' \in$ [0,*N*],表示量子态的在每个基底的概率为 $p_i = \int_{i}^{i+1} g(x')dx', p_i \ge 0, \sum p_i = 1.数据集仍为单点$ 集*X* $= {$ *x*= (*x*₀, · · · ,*x*_{*N* $-1})},$ *x* $_i = <math>\sqrt{p_i}$.目标量子 态是 *X* 对应的振幅编码

$$f(x) = \sum_{i=0}^{N-1} x_i |i\rangle.$$
 (9)

这类编码的编码技术是从已有的量子态出发,根据 经典数据的分布 g(x')得到新的量子态,与一般的 振幅编码相比,这类编码的 g(x)参数可能与 N 无 关.可以通过对分布函数 g(x')做一些限制,得到关 于某些函数性质有关的态制备方法.

在 Grover 等^[16]的工作中,作者进一步提出对 于很大一类被称为"对数凸"的函数,都可以通过这 种编码方式进行制备,这其中包括常见的正态分 布和指数分布.除了 Grover 的工作,Kitaev 和 Webb^[24]也分析了高斯分布的量子态制备.文献 [25] 给出了一种基于矩阵积态 (matrix product state) 方法,得到了当g(x')为光滑可微且导函数有界时 的编码方式.该算法需要O(n)个量子比特,执行时 间为O(n).实验方面 Vazquez 和 Woerner^[26]给出 了基于量子振幅放大算法简化态制备的方法,并 在 IBM 的物理比特上展示了实现了该算法.

以机器学习的方式研究这类情形的态制备问题有大量的工作,如生成对抗网络^[27],给定经典数据的分布,利用含参数的量子线路生成一种量子分布,再由对抗识别器对量子分布采样,反复调整,直到识别不出经典数据分布与生成的分布,训练完毕.数值模拟过程中所需要的量子门数量控制在 O(P(n)).也有其他机器学习相关的工作,如Arrazola等^[28]用含参线路在光量子计算机模拟器演示了许多量子态的生成,如 ON 态和 GKP 态. 值得一提的是,此类统计分布的态制备问题,在量子蒙特卡罗模拟算法中处于非常核心的地位^[29],而后者已经被证明在很多金融和其他模拟问题中显示出量子优越性^[30-32].

2.4 哈密顿量编码

经典数据转为量子态的另一种方案是将经典数据的信息编码到某个量子系统的哈密顿量中,运用哈密顿模拟的方式代替将经典数据转为量子态的方法^[33]. 记经典数据集为 $X = \{x = (x_0, \cdots, x_{N-1})\}$,记对应的哈密顿量为 H_x ,表明哈密顿量依赖于经典数据的选择.则对于量子系统的初态 $|\psi\rangle$,演化时间为 t,得到演化后的量子态为

$$|\psi'\rangle = \mathrm{e}^{-\mathrm{i}H_x t} |\psi\rangle.$$

哈密顿量的演化过程可以由量子线路语言实现^[18].考虑一个 n量子比特的量子系统,哈密顿量可以分解为一些哈密顿量的和,即 $H = \sum_{i=0}^{L-1} H_i$,其中 H_i 为较易模拟的哈密顿量,L = O(P(n)).由Trotter 公式^[33],

$$e^{-iHt} = (e^{-iHt/n})^n = \left(\prod_{i=0}^{L-1} e^{-iH_it/n} + \mathcal{O}((t/n)^2)\right)^n.$$

当 *n*充分大时,可以用多次的演化 e^{-iH_it/n}来实现 量子模拟. H_i常见的选择是泡利算符.

哈密顿量编码的步骤分为两大类. 一类是从哈 密顿量 *H* 出发. 1) 由经典数据确定哈密顿量 *H*, 如 经典数据为动能、势能函数决定的参量; 2) 在某个 量子系统中选定基底, 确定哈密顿量的矩阵元; 3) 哈密顿量模拟. 模拟过程包含哈密顿量的分解, 需要确定 *H*_i. Matto 等 ^[34] 用格雷码序的方式对经 典数据进行哈密顿量编码, 是一种比特数有效的编 码方案. 另一类是从分解后的哈密顿量 H_i出发. 1) 选定 H_i , 由经典数据得到每个 H_i 的系数; 2) 在 某个量子系统中选定基底,确定哈密顿量H_i的矩 阵元; 3) 得到总的哈密顿量, 即为经典数据的哈密 顿量编码. 例如用第二类的方法, 假定经典数据 x = (x_0, \dots, x_{4^n-1}) ,我们将经典数据 x编码到 n粒子 哈密顿量中,以泡利算符和单位算符在计算基上的 表示为基底,基底个数为4ⁿ,总的哈密顿量为

 $H = x_0(I_1 \otimes \cdots \otimes I_n) + x_1(\sigma_{1x} \otimes I_2 \otimes \cdots \otimes I_n)$ $+ x_{4^n-1}(\sigma_{1z} \otimes \sigma_{2z} \otimes \cdots \otimes \sigma_{nz}).$

复杂度 3

当数据规模为O(N)时,基底编码的执行时间 为O(N)次, 需要的量子比特数为O(N), 复杂度为 O(N²). 振幅编码适用范围广, 其中显式振幅编码 的复杂度为 $\mathcal{O}(N)\log N$,对于含黑箱的振幅编码, 仅考虑质询复杂度,可以做到有效制备,即在一定 条件下质询复杂度可以达到O(log(N)), 但黑箱的 执行时间在实际操作中需要考量. QRAM 的编码 方式从已知量子态获取经典数据得到新的量子态, 这个过程中需要O(N)个量子比特,但执行时间可 以做到*O*(*n*). 而量子抽样编码也是直接从量子态 出发,比较目标量子态与量子初态的差异得到新的 量子态,比特数和时间有效性都可以实现,这是由 于给出的分布函数可以不依赖于数据量规模.在哈 密顿量编码中,哈密顿量的合理选择可使得编码方 式中的比特数有效,执行时间取决于哈密顿量的矩 阵形式和哈密顿量模拟的精度.进一步的,考虑到 量子比特数和时间执行次数的平衡和取舍,噪声抗 性等因素,态制备问题应该被仔细斟酌.以上的复 杂度分析可参看表1.

4 研究前景和展望

在未来,经典计算机芯片的工艺制程接近摩尔 定律极限,经典计算机的算力发展达到瓶颈期.而 大数据的处理使得算力需求呈快速增长趋势, 这之 间算力的供需矛盾关系使得人们迫切地寻找新的 计算模式,研究量子机器学习的出发点是解决这种 矛盾关系.具体来说是希望在处理某一类问题时, 量子机器学习的方法能够大大缩短传统经典机器 学习需要的时间,继而在更广泛的问题中表现出加 速能力.量子计算机的实用化受限于量子比特、量 子门的质量和量子操作系统等诸多实验因素,故量 子机器学习的研究大多停留在数值模拟或是构建 理论模型的阶段.在这个大背景下,人们不过多关 注量子机器学习中的物理实现.

目前态制备问题里更受关注的是量子抽样编 码,其中涌现出了许多利用量子机器学习研究态制 备的工作. 这种编码方式通过经典神经网络与含参 量子线路的结合,以监督学习的方式训练参数,不 断优化量子线路得到近似的目标量子态.复杂度的 分析通常考虑参数的数量,但含参量子线路的表示 能力与学习方式的选择都会影响其编码的有效性. 这类工作比较丰富,例如生成对抗网络^[27],利用含 参数的量子线路生成一种分布并由对抗识别器采 样,机器学习的方式训练参数,直到对抗识别器识 别不出目标分布与生成的分布; Arrazola 等^[28] 采 用的是光量子计算机模拟器,利用自动微分法优化 得到目标量子态;最近 Zhou 等^[35]提出了一种自

表 1	态制备的不同编码方式的复杂度分析

Table 1. Complexity analysis of kinds of encoding methods for state preparations.					
编码方式		数据类型	比特数	执行时间	
基底编码		二值数据	$\mathcal{O}(N)$	$\mathcal{O}(N)$	
日子拒桓纪初	计算基 i> 上编码	连续变量	$\mathcal{O}(\log N)$	$\mathcal{O}(N)$	
业工行水中的细节	$ 2^i\rangle$ 上编码	连续变量	$\mathcal{O}(N)$	$\mathcal{O}(\log N)$	
	Grover和Kaye	连续变量	$\mathcal{O}(\log N)$	_	
含黑箱振幅编码	Soklakov和Schack	连续变量	$\mathcal{O}(\log N)$	$\mathcal{O}(\operatorname{Poly}(\log N))^*$	
	QRAM	连续变量	$\mathcal{O}(N)$	$\mathcal{O}(\log N)$	
量子抽样编码		分布函数	$\mathcal{O}(\log N)$	$\mathcal{O}(\log N)$	
哈密顿量编码		连续变量	$\mathcal{O}(N)/\mathcal{O}(\log N)$	$\mathcal{O}(\operatorname{Poly}(\log N))^*$	

注:*同时受保真度和成功率的影响.

动微分的量子含参线路,可以优化得到任意的量 子态.

另一方面,态制备问题作为经典数据和量子态 的桥梁,在量子机器学习中的使用不可避免.相较 于经典计算机编码数据的方式,量子计算机在态制 备时编码数据的指数级加速能力是没有问题的,但 这是以大量量子比特数为前提的实现方式. 研究量 子机器学习的初衷是实用化解决经典问题,更应该 考虑其中的态制备方案的时间计算资源和空间计 算资源.对于复杂度的分析,态制备的算法复杂度 至少是数据自由度的量级,既要分析时间复杂度, 也要考虑量子比特数的规模. 单看时间复杂度, 得 出具有加速能力的结论还不足以体现量子机器学 习的能力,分析时应该谨慎.但同时也要乐观对待 量子机器学习的能力,至少以发展的眼光去看待. 例如,大数分解的量子算法复杂度比已知最优的经 典算法有指数级的提高,而人们在大数分解算法提 出前也不清楚量子计算的加速能力. 总的来说, 随 着量子计算机的发展特别是硬件水平的提升,相信 会有更多的人关注态制备问题.

参考文献

- [1] Jordan M I, Mitchell T M 2015 Science 349 255
- [2] Lay K T, Katsaggelos A K 1990 Opt. Eng. 29 436
- [3] Lu D, Weng Q 2007 Int. J. Remote Sens. 28 823
- [4] Samaria F S, Harter A C 2002 Proceedings of 1994 IEEE Workshop on Applications of Computer Vision Sarasota, December 5–7, 1994 p138
- [5] Guillaumin M, Verbeek J, Schmid C 2009 In 2009 IEEE 12th international conference on computer vision Kyoto, Japan, September 29–October 2, 2009 p498
- [6] Sun Y 2015 Deep Learning Face Representation by Joint Identification-verification (Ann Arbor: ProQuest LLC) pp40-57
- [7] Silver D, Huang A, Maddison C J, Guez A, Sifre L, Driessche G V, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel

T, Hassabis D 2016 Nature 529 484

- [8] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, Driessche G V, Graepel T, Hassabis D 2017 Nat. Nature 550 354
- Le Q V 2013IEEE International Conference on Acoustics, Speech and Signal Processing Vancouver, Canada, May 26-31, 2013 p8595
- [10] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S 2017 Nature 549 195
- [11] Soklakov A N, Schack R 2006 Phys. Rev. A 73 012307
- [12] Schuld M, Petruccione F 2018 Supervised Learning with Quantum Computers (Vol. 17) (Berlin: Springer) pp139–171
- [13] Harrow A W, Hassidim A, Lloyd S 2009 Phys. Rev. Lett. 103 150502
- [14] Lloyd S, Mohseni M, Rebentrost P 2014 Nat. Phys. 10 631
- [15] Rebentrost P, Mohseni M, Lloyd S 2014 Phys. Rev. Lett. 113 130503
- [16]~ Grover L, Rudolph T 2002 arXiv: 0208112 v1 [quant-ph].
- [17] Kaye P, Mosca M 2001International Conference on Quantum Information New York, USA, June 13, 2001 p28
- [18] Nielsen M A, Chuang I 2002 Quantum Computation and Quantum Information (Cambridge: Cambridge University Press) pp120-215
- [19] Kerenidis I, Prakash A 2016 arXiv: 1603.08675 v3 [quant-ph].
- [20] Matteo O D, Gheorghiu V, Mosca M 2020 IEEE Trans. Quantum Eng. 1 4500213
- [21] Paler A, Oumarou O, Basmadjian R 2020 Phys. Rev. A 102 032608
- [22] Hann C T, Lee G, Girvin S M Jiang L 2021 PRX Quantum 2 020311
- [23] Mitarai K, Kitagawa M, Fujii K 2019 Phys. Rev. A 99 012301
- [24] Kitaev A, Webb W A 2008 arXiv: 0801.0342 [quant-ph].
- [25] Holmes A, Matsuura A Y 2020 In 2020 IEEE International Conference on Quantum Computing and Engineering (QCE) Denver, CO, USA, October 12–16, 2020 p169
- [26]~ Vazquez A C, Woerner S 2021 Phys. Rev. A 15 034027
- [27] Zoufal C, Lucchi A, Woerner S 2019 npj Quantum Inf. 5 103
- [28] Arrazola J M, Bromley T R, Izaac J, Myers C R, Brádler K, Killoran N 2019 Quantum Sci. Technol. 4 024004
- [29] Montanaro A 2015 Proc. R. Soc. A 471 20150301
- [30] Orus R, Mugel S, Lizaso E 2019 Rev. Phys. 4 100028
- [31] Stamatopoulos N, Egger D J, Sun Y, Zoufal C, Iten R, Shen N, Woerner S 2020 Quantum 4 291
- [32] Woerner S, Egger D J 2019 npj Quantum Inf. 5 15
- [33] Lloyd S 1996 Science 273 1073
- [34] Matteo O D, McCoy A, Gysbers P, Miyagi T, Woloshyn R M, Navrátil P 2021 Phys. Rev. A 103 042405
- [35] Zhou P F, Hong R, Ran S J 2021 arXiv: 2104.14949[quantph].

SPECIAL TOPIC—Machine learning and physics

Quantum state preparation and its prospects in quantum machine learning^{*}

Zhao Jian¹⁾ Chen Zhao-Yun¹⁾ Zhuang Xi-Ning¹⁾³⁾ Xue Cheng¹⁾ Wu Yu-Chun^{1)2)†} Guo Guo-Ping¹⁾²⁾³⁾

1) (CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China)

2) (Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China)

3) (Origin Quantum Computing Company Limited, Hefei 230026, China)

(Received 21 May 2021)

Abstract

The development of traditional classic computers relies on the transistor structure of microchips, which develops in accordance with Moore's Law. In the future, as the distance between transistors approaches to the physical limit of manufacturing process, the development of computation capability of classical computers will encounter a bottleneck. On the other hand, with the development of machine learning, the demand for computation capability of computer is growing rapidly, and the contradiction between computation capability and demand for computers is becoming increasingly prominent. As a new computing model, quantum computing is significantly faster than classical computing for some specific problems, so, sufficient computation capability for machine learning is expected. When using quantum computing to deal with machine learning tasks, the first basic problem is how to represent the classical data effectively in the quantum system. This problem is called the state preparation problem. In this paper, the relevant researches of state preparation are reviewed, various state preparation schemes proposed at present are introduced, the processes of realizing these schemes are described, and the complexities of these schemes are summarized and analyzed. Finally, some prospects of the research work in the direction of state preparation are also presented.

Keywords: state preparation, quantum machine learning, encoding

PACS: 03.67.Ac, 03.67.Lx, 03.67.–a

DOI: 10.7498/aps.70.20210958

^{*} Project supported by the National Key Research and Development Program of China (Grant No. 2016YFA0301700), the National Natural Science Foundation of China (Grant No. 11625419), the Anhui Initiative in Quantum Information Technologies, China (Grant No. AHY080000), and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB24030600).

[†] Corresponding author. E-mail: wuyuchun@ustc.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

基于波动与扩散物理系统的机器学习

陈江芷 杨晨温 任捷

Machine learning based on wave and diffusion physical systems Chen Jiang-Zhi Yang Chen-Wen Ren Jie 引用信息 Citation: Acta Physica Sinica, 70, 144204 (2021) DOI: 10.7498/aps.70.20210879 在线阅读 View online: https://doi.org/10.7498/aps.70.20210879

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

您可能感兴趣的其他文章

Articles you may be interested in

基于人工神经网络在线学习方法优化磁屏蔽特性参数

Online learning method based on artificial neural network to optimize magnetic shielding characteristic parameters 物理学报. 2019, 68(13): 130701 https://doi.org/10.7498/aps.68.20190234

铅基钙钛矿铁电晶体高临界转变温度的机器学习研究

High critical transition temperature of lead-based perovskite ferroelectric crystals: A machine learning study 物理学报. 2019, 68(21): 210502 https://doi.org/10.7498/aps.68.20190942

基于经验知识遗传算法优化的神经网络模型实现时间反演信道预测

Prediction of time reversal channel with neural network optimized by empirical knowledge based genetic algorithm 物理学报. 2019, 68(17): 170503 https://doi.org/10.7498/aps.68.20190327

多气隙电阻板室飞行时间谱仪技术

Time of flight technology based on multi-gap resistive plate chamber 物理学报. 2019, 68(10): 102901 https://doi.org/10.7498/aps.68.20182192

基于机器学习构建的环三亚甲基三硝胺晶体势

Energetic potential of hexogen constructed by machine learning 物理学报. 2020, 69(23): 238702 https://doi.org/10.7498/aps.69.20200690

不确定分数阶时滞混沌系统自适应神经网络同步控制

Synchronization of uncertain fractional-order chaotic systems with time delay based on adaptive neural network control 物理学报. 2017, 66(9): 090504 https://doi.org/10.7498/aps.66.090504

专题: 机器学习与物理

基于波动与扩散物理系统的机器学习*

陈江芷1) 杨晨温1) 任捷1)2)†

1) (同济大学物理科学与工程学院,声子学与热能科学中心,上海市特殊人工微结构材料与技术重点实验室,上海 200092)

2) (同济大学, 上海自主智能无人系统科学中心, 上海 200092)

(2021年5月10日收到; 2021年7月1日收到修改稿)

物理学在机器学习中的应用以及两者的交叉融合正引起广泛关注,尤其是在波动系统和扩散系统中.本 文重点关注波动与扩散物理系统和机器学习之间的内在联系以及对机器学习算法和物理实现的推进作用, 综述了波动系统和扩散系统中的机器学习研究,介绍了部分最新研究成果.文中首先讨论了监督学习的波动 系统实现,包括神经网络的波动光学实现、量子搜索的波动实现、基于波动系统的递归神经网络以及神经形 态的非线性波动计算.接着,文中继续讨论了受扩散系统启发的机器学习算法,如基于扩散动力学的分类算 法,基于热扩散的数据挖掘和信息过滤,以及基于群体扩散的搜索优化等.波动系统以其天然的并行性、高 效、低能耗等优势,通过丰富的波动力学和波动物理现象进行计算或算法模拟,正成为机器学习的新型物理 载体.扩散系统中的物理机制可以启发构建高效的机器学习算法,用于复杂系统和物理学研究中的分类、优 化等问题.期望通过对波动、扩散物理系统与机器学习内在联系的讨论,能够为开发物理启发的新算法和硬 件实现甚至软硬一体化带来抛砖引玉的启示.

关键词: 波动系统,扩散系统,机器学习,人工神经网络 PACS: 42.79.Ta, 05.40.Fb, 05.90.+m

DOI: 10.7498/aps.70.20210879

1 引 言

机器学习与人工智能被认为是这个时代的突 破性技术之一,应用领域广泛^[1-4].机器学习主要 用大量数据训练结构,以此模拟人类的学习行为, 进而对复杂事件作出推理和决策.这与物理系统的 思维方式十分接近.因为任一物理系统都可视作将 输入信息映射到输出结果的一种信息处理方式,于 是机器学习能够作为物理学研究的工具^[5].但是机 器学习与物理学的交叉互融不仅限于此,物理学也 可以促进机器学习.物理学中的波动系统和扩散系 统可以构建更高效的机器学习硬件平台和算法,从 而满足人们日益增长的对计算机算力的需求. 自从 20 世纪 80 年代机器学习成为一个独立 的研究方向以来,各种机器学习算法被大量提出, 但是如何直接利用硬件设备实现更高效的机器学 习仍是一个值得探索的命题.尽管电子器件的发展 使机器学习在集成电路上能达到令人满意的效果, 这种方式需要比较高的能耗和带宽.同时,通过传 统的电子设备实现机器学习通常需要大量时间和 较大尺寸的硬件设备.这使得我们在处理复杂问题 和边缘计算时将会十分困难^[6].波动系统实现监督 学习 (如实现人工神经网络),具有天然的优势.它 保证了大规模的并行性^[7,8]和片上集成后较小的设 备尺寸.信息将以非常快的速度传输,例如光速, 并且这种传输方式极大地减少了能耗,甚至能够达 到零能耗^[9].另一方面,物理学和机器学习都试图

* 国家自然科学基金 (批准号: 11935010, 11775159)、上海市自然科学基金 (批准号: 18JC1410900)、上海市特殊人工微结构材料 与技术重点实验室项目和中央高校基本科研业务费专项资金资助的课题.

© 2021 中国物理学会 Chinese Physical Society

[†] 通信作者. E-mail: xonics@tongji.edu.cn

分析数据的规律来建立模型,从而预测系统的行为,两者之间存在着一些本质联系.可以用物理学的机制来理解和构建机器学习方法,例如基于扩散系统实现分类、信息过滤、优化等无监督算法.

物理学和机器学习的交叉互融具有悠远的 历史和广泛的应用, 涉及范围十分宽广. 更多有 关机器学习与物理学的讨论可以参考《Review of Modern Physics》以及《Physics Reports》的两 篇综述文章[5,10],它们更一般地回顾了机器学习技 术在物理学的各大领域的典型应用.由于能力和篇 幅限制,本文主要基于课题组自身的研究积累,尝 试从波动与扩散动力学的物理视角,来统一地阐述 和理解机器学习相关研究. 特别地, 本文重点关注 波动、扩散物理系统对物理实现的推进作用,以及 机器学习算法启发,主要讨论波动物理作为人工神 经网络的硬件平台以及两者之间的内在联系,以及 受扩散物理启发的分类、优化等机器学习算法.扩 散与波动过程是物理学中的基本动力学过程,我们 希望通过这一独特的切入点,为后续物理启发的机 器学习研究带来新的思路.

2 波动与扩散系统中的机器学习算法

根据训练数据是否已经被人为标记,可以将机 器学习算法大致地分为两类——监督学习和无监 督学习 (图 1(a)). 监督和无监督学习算法不仅能成 为解决物理系统中特定问题的机器学习工具,波动物理系统和扩散物理系统还可以反过来为机器学习提供硬件实现平台和启发新的算法.

物理学中的波动系统既能实现监督学习,也能 实现无监督学习[11,12]. 然而随着近年来计算机视觉 等领域的发展,波动系统执行推理任务的能力尤为 重要,于是波动系统被频繁地与一些监督学习算法 联系在一起. 在监督学习中, 假设有 n个数据样本, 用 X_i ($i = 1, 2, \cdots, n$)来表示这些数据的特征向量. 同时,各数据样本的标记是已知的,用 Y_i (i=1, 2,…,n)表示. 监督学习的目标是建立一个函数 f, 使 $f(X_i)$ 尽可能地接近 Y_i . 过去二十年间, 人工 神经网络[13] 成为最热门的监督学习方法之一. 1982年,美国科学家 Hopfield^[14]提出了适应于集 成电路的人工神经网络,引起了巨大的反响.图1(b) 展示了一个单隐藏层神经网络的基本结构.其中 W_1 , W_2 代表权重矩阵, b_1 , b_2 代表偏差. 隐藏层 和输入层之间的连接满足如下关系: $H = W_1X_+$ b_1 . 神经网络之所以区别于传统的线性回归, 并具 有更强的计算能力的原因之一在于,神经网络包含 了非线性激活元. 若以σ(·)来表示非线性激励函 数, 则输出结果为 $Y = \sigma(W_2 \cdot \sigma(H) + b_2)$. Farhat 等^[15]于 1985 年提出 Hopfield 模型可以在光学系 统中实现,奠定了波动系统实现人工神经网络的 基础.



图 1 (a) 监督学习与无监督学习; (b) 一个单隐藏层神经网络的基本结构, 各层神经元根据权重系数相互连接; (c) 利用无监督 学习进行社交网络分析

Fig. 1. (a) Supervised learning and unsupervised learning; (b) the structure of a single hidden-layer neural network, where neurons are connected by weight coefficients; (c) unsupervised learning can be applied in social network analysis.

物理学中的扩散现象描述了大量粒子集体运 动的统计结果. 对物理学中的扩散机制的研究和深 入了解启发了新的机器学习算法的诞生,尤其是一 些改进的无监督学习算法. 与监督学习不同, 无监 督学习的训练样本是没有已知标记的.因此,无监 督学习不再依赖于"经验",而是更注重数据样本 的内在模式和统计规律,这与物理学中的扩散机 制存在本质联系. 传统的无监督学习方法包括聚 类——*k*-means算法^[16]和EM算法^[17]等,以及数 据降维——主成分分析 (PCA)^[18] 和流形学习^[19,20] 等方法. 基于对物理学中的热传导扩散和概率扩散 等系统的研究,逐渐开发出各种改进的机器学习算 法. 例如, 基于扩散动力学实现数据降维, 并根据 数据的内在规律进行分类[21]: 基于热传导扩散实 现数据挖掘,建立推荐模型并应用于社交网络分 析^[22](图 1(c));以及基于全局扩散搜索算法建立优 化模型,实现材料搜索和结构预测,等等.

下面将分别介绍波动系统中的监督学习,包括 神经网络的波动光学实现、波动系统的递归神经网 络、神经形态的非线性波动计算,以及基于扩散系 统的无监督学习,包括基于扩散动力学的分类模 型、基于热传导扩散的推荐模型、基于全局和局部 扩散搜索算法的优化模型.

3 波动系统中的监督学习

数学上, 波在自由空间中的传播可以用菲涅 耳-基尔霍夫衍射积分公式来描述, 它相当于对场 进行卷积. 在实际应用中, 这种卷积行为可以通过 傅里叶光学实现. 输入场通过特定的透镜进行傅里 叶变换, 与光学元件的振幅和相位的按元素相乘. 这就完成了输入场与元素的傅里叶逆变换的卷 积. 于是, 波在自由空间和不同介质中的传播可以 高速、低能耗地实现卷积、矢量-矩阵乘法等运算 (图 2)⁶, 从而在波动系统中实现线性回归^[23]等, 乃至更复杂的计算. 我们的宇宙简直就是一个天然 的计算器.

随着深度学习的兴起,近年来人们更关注人工 神经网络的波动物理实现.人工神经网络需要在矩 阵乘法运算的基础上引入非线性激活函数,这在波 动系统中是个挑战.如何用波的传输模拟神经元中 的信息传输,如何控制权重也是亟待解决的问题. 接下来,简要介绍近期与波动系统中的神经网络有 关的部分工作.

3.1 神经网络的波动光学实现

对于最常见的全连接神经网络,输出层中的 每个元素都可以视作输入层中所有元素的加权和.





Fig. 2. Wave propagation through different media and the corresponding linear matrix operations: (a) A traditional optical 4f system realizes multiplication in Fourier space, which corresponds to the convolution in the original space; (b) modified 4f systems can copy the input field with a grating and use different kernels for convolution; (c) a 4f-type system can implement the vector-matrix multiplication^[6].

这种矢量-矩阵乘法的运算,可以通过马赫-曾德干 涉仪在光学领域实现^[24].随着近年来光子集成电 路的迅速发展,科学家用一个可编程纳米光子处理 器来实现基于相干光和全光学矩阵乘法的硅光子 神经形态电路(图3(a))^[25].光子神经形态计算的另 一个重要方法是基于相变材料和器件的整合.最近 一项研究中,科学家利用微米级环型谐振器将输入 信息调制成不同波长,并通过相变材料实现权重调 节 (图 3(b)). 这种方式通过相变材料和环型谐振器的耦合来实现非线性激活,最终在光子集成系统中构建出了脉冲神经网络^[12,26],有效地减小了光子芯片的体积.

除了光子集成电路,还可以利用波通过透镜的 传播来构建神经网络. Chang 等^[27]提出了一种光 电混合的卷积神经网络,即在电子计算之前加入 一层光学卷积层,减小电子计算成本和处理时间,



图 3 (a) 基于相干光和全光学矩阵乘法的硅光子神经形态电路可用于实现元音分类^[25]; (b) 基于波分复用 (WDM) 的分层结构 构成的全光脉冲神经网络, 能够实现图像和语言识别^[12]; (c) 改进的光学 4*f*系统实现卷积神经网络, 提高图像分类性能^[27]; (d) 全 光衍射深度神经网络实现数字分类^[20]; (e) 光学衍射元件与图像处理算法端对端协同设计^[32]

Fig. 3. (a) Nanophotonic circuits based on coherent light and all-optical matrix multiplication is capable for vowel recognition^[25]; (b) image and language recognition are achieved by an all-optical spiking neural networking with wavelength division multiplexing (WDM)^[12]; (c) a design for an optical convolutional layer using a modified optical 4f system^[27]; (d) an all-optical diffractive deep neural network that implements the digit classification^[29]; (e) end-to-end learning paradigms of diffractive optics and processing algorithms^[32].

同时提高处理图像分类任务的性能(图 3(c)).这种 方式是分离地在光计算层进行线性计算,在电子计 算层实现非线性激活函数,但是光电转换过程的效 率就成为主要瓶颈.构建全光神经网络可以解决这 一问题,即神经网络中的线性和非线性操作都在波 动系统中实现.例如,用空间光调制器和傅里叶透 镜实现线性操作,电磁诱导透明的激光冷却原子实 现非线性光学激活函数^[28].

衍射层可以用来代替透镜调节与波的相互作 用,在缩小系统外形的前提下构建更高效的全光神 经网络. 衍射层上的每个点透射或反射入射波, 并 通过波的衍射连接到下一衍射层. 根据惠更斯-菲 涅耳原理,输入波经过衍射层上各个点的透射或反 射后成为次波源,次波源的振幅和相位由输入波与 该点的复数透射或反射系数的乘积决定.因此,衍 射层上各点的透射或反射系数可被视作神经网络 的权重,通过设计固定结构的多层衍射层形成全光 神经网络,实现手写数字分类(图 3(d))^[29]. 类似的 衍射神经网络还可以用来执行光学逻辑运算,实现 小型化的光学逻辑门^[30].为了进一步提高核心计 算模块的训练速度和能量效率,科学家们提出了一 种原位光学学习结构. 通过这种结构, 可以在光学 系统中实现衍射神经网络的训练过程[31]. 衍射光 学元件与计算成像结合,有望实现轻薄的高性能成 像系统. 同济大学程鑫彬课题组^[32]提出基于同心 圆环分解的成像模型计算降维理念,成功地将衍射 光学元件和图像处理算法端到端设计框架的内存 需求降低了一个数量级,有助于发展基于衍射光学 元件的轻薄计算成像系统 (图 3(e)).

除此之外, 衍射神经网络也可以在声学超构材 料^[33] 或者傅里叶空间中^[34] 实现. 已有研究在声波

系统中利用超材料控制声波的相位和透射率进 行模拟计算[35],可以实现空间上的微分、积分和卷 积^[36,37]以及常微分方程的求解^[38,39]等. Weng 等^[33] 从理论上提出并实验证明了一个纯粹的被动神经 网络,由于它的超材料单元产生深亚波长相移,该 声学神经网络能够通过分析声散射实时识别复杂 物体 (图 4(a)). 特别地, 由于线性波动系统本身具 有并行运算能力,即两个或多个波包的传播不互相 影响,利用波动系统构建机器学习算法时还可以考 虑结合量子算法进一步提升计算速度^[40].例如,以 波元的振幅代表量子态的概率幅, 波元的相位代表 量子态的相位,可以实现有别于经典搜索的量子搜 索算法.对于一个包含有 N个数据的数据库而言, 找到一个指定数据, 经典算法成功搜索一个数据需 要根据搜索条件在数据库中逐一进行比对,平均需 要N/2次迭代计算,其量级正比于N. 而 Grover 量子搜索算法成功搜索数据所需的迭代次数r_f为: $r_f = \frac{\pi}{4}\sqrt{N} - \frac{1}{2}$. 显然, 当数据库足够大时, Grover 算法所需迭代次数 r_f 正比于 \sqrt{N} ,远小于N/2.具 体到波动系统中(图 4(b))^[40],可以利用超材料调 节波的传播波速和方向以实现相对应的量子逻辑 门操作, 如果将经典波动系统中实现量子算法的思 路与机器学习算法相结合,可以进一步提升受到经 典计算机算力限制的机器学习算法的性能.

3.2 波动系统的递归神经网络

波的动力学与递归神经网络 (recurrent neural network, RNN) 之间具有强烈的映射关系 (图 5)^[41]. 包括声学和光学在内的波动物理可以自然地为时 变信号构建模拟处理器. 如图 5(a) 所示, 递归神经 网络中的更新过程可以描述为



图 4 (a) 基于声学超构神经网络的被动目标识别^[33]; (b) 利用声表面波系统实现量子搜索算法,实现与量子逻辑门相似的操作^[40] Fig. 4. (a) Passive object recognition with acoustic meta-neural-network^[33]; (b) realize quantum search algorithm with acoustic system, achieving operations similar to quantum logic gates^[40].



图 5 标准 RNN 和波物理的对比 (a) 具有离散的输入、输出序列的 RNN 的更新过程; (b) 具有连续输入、输出序列的波动系统系统的更新过程^[41]

Fig. 5. Comparison of a standard RNN and a wave system: (a) The update process of an standard RNN with discrete input and output sequence; (b) the update process of a wave-based physical system with continuous input and output sequence^[41].

$$\boldsymbol{h}_{t+1} = \sigma^{(h)} \big(\boldsymbol{W}^{(h)} \cdot \boldsymbol{h}_t + \boldsymbol{W}^{(x)} \cdot \boldsymbol{x}_{t+1} \big), \quad (1)$$

$$\boldsymbol{y}_{t+1} = \sigma^{(y)} (\boldsymbol{W}^{(y)} \cdot \boldsymbol{h}_{t+1}), \qquad (2)$$

其中 x_t , h_t , y_t 分别代表 t时刻的输入向量、隐藏 态向量和输出向量; $W^{(h)}$, $W^{(x)}$, $W^{(y)}$ 代表权重 矩阵, 上标(h), (x), (y)分别对应于隐藏层, 输入 层和输出层; $\sigma^{(h)}(\cdot)$ 和 $\sigma^{(y)}(\cdot)$ 代表非线性激励函数.

波场分布 u(x,y,z)可以由二阶偏微分方程 表示:

$$\frac{\partial^2 u}{\partial t^2} - c^2 \cdot \nabla^2 u = f, \qquad (3)$$

其中∇²是拉普拉斯算子; c = c(x, y, z)表示波速的 空间分布; f = f(x, y, z)表示源. 对 (3) 式进行有限 差分, 得到递归关系:

$$\frac{u_{t+1} - 2u_t + u_{t-1}}{\Delta t^2} - c^2 \cdot \nabla^2 u_t = f_t.$$
 (4)

将(4)式写成矩阵形势:

$$\begin{bmatrix} u_{t+1} \\ u_t \end{bmatrix} = \begin{bmatrix} 2 + \Delta t^2 \cdot c^2 \cdot \nabla^2 & -1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} u_t \\ u_{t-1} \end{bmatrix} + \Delta t^2 \cdot \begin{bmatrix} f_t \\ 0 \end{bmatrix}.$$
(5)

波系统的隐藏态定义为当前时刻和前一时刻的场

分布相连 $h_t \equiv [u_t, u_{t-1}]^T$, 对应于 RNN 中的隐藏 态向量 h_t , 于是波动方程的更新过程可写作:

$$\boldsymbol{h}_{t+1} = A\left(\boldsymbol{h}_{t}\right) \cdot \boldsymbol{h}_{t} + \boldsymbol{P}^{(i)} \cdot \boldsymbol{x}_{t+1}, \quad (6)$$

$$\boldsymbol{y}_{t+1} = \left| \boldsymbol{P}^{(0)} \cdot \boldsymbol{h}_{t+1} \right|^2.$$
(7)

见 (图 5(b)). 显然, (6) 式和 (7) 式有与 (1) 式和 (2) 式一样的形式, 其中 h_t 都是隐藏态向量, y_{t+1} 都是输出向量.类比于标准 RNN 中的权重矩阵 W^(x), W^(y), 波动方程的隐藏态与输入输出之间 的关系也由线性算子 P⁽ⁱ⁾, P^(o)给出. 但不同之处 在于, **P**⁽ⁱ⁾, **P**^(o)式只在波源的入射处和输出波 的测量处为1,其余位置为0的矩阵.于是通过 $P^{(i)} \cdot x_{t+1}$ 将输入向量 x_{t+1} 编码成空间内特定位 置入射的波源 f_t , 对应于 (1) 式中的 $W^{(x)} \cdot x_{t+1}$. 稀疏矩阵 $A(h_t)$ 描述了无源条件下波场 u_t 的更新, 对应于 (1) 式中的隐藏态权重矩阵 $W^{(h)}$. 在线性 波中, A 实际上不依赖于 h_t . 为了实现 (1) 式中的 非线性激励函数 $\sigma^{(h)}(\cdot)$,可以进一步引入非线性波, $A(h_t)$ 中的波速的形式为 $c = c_{\text{lin}} + u_t^2 \cdot c_{\text{nl}}, c_{\text{nl}}$ 对应 非线性响应区域. 这种形式的非线性在各种各样的 波物理中都会遇到,包括水波、非线性光学材料、 气泡流体和软材料中的声学. 而通过测量波的强度 得到输出结果时, (7) 式中的 $|\cdot|^2$ 自然地完成了 (2) 式 中的非线性操作 $\sigma^{(y)}(\cdot)$.

通过这种递归神经网络与波动物理的映射表 明,神经网络学习时间数据中的复杂特征,可以通 过特定的波动物理系统来实现. 例如, 通过波在非 均匀介质中的散射和传播实现对音频信号的元音 分类,实现了与递归神经网络的标准数字实现相当 的性能[41]. 除此之外, 在小型硬件构成的神经网络 上加入非线性动力学特征,如振荡和同步,可以实 现特殊的分类任务,例如训练一个由四个自旋转矩 纳米振荡器组成的硬件网络,通过自动实时学习规 则调整语音元音的频率来识别语音元音[42].另外, 波在非均匀纳米光子介质中的散射也可以实现连 续无分层的方式的人工神经网络计算.非均匀介质 通过变换波前来实现复杂的计算任务,如图像识 别. 这些计算介质可以小到几十个波长, 并提供超 高的计算密度,这种方式利用亚波长散射体来实现 复杂的输入/输出映射,超越了传统纳米光子器件

的能力^[43].除了经典的时间序列学习,波动系统有 望应用于更复杂的系统学习,如厦门大学赵鸿^[44] 提出的利用时序数据的自演化学习机,可以解决 "黑箱"系统周期动力学,甚至混沌动力学的推断问 题,并有望推广到复杂耦合体系的系统重构^[45].

3.3 神经形态的非线性波动计算

神经网络计算同样可以通过非线性波实现. Marcucci 等^[46] 最近研究了非线性波具有进行神经 形态计算的潜力.非线性波,如孤子、冲击波和怪 波的发散行为能够提供足够的复杂度来进行机器 学习,它们被有效地应用到储蓄池计算中. Marcucci 等^[46] 提出了一个由非线性偏微分方程驱动的计算 模型,称为单波层前馈网络 (single wave-layer feedforward network, SWFN)(图 6(a)). SWFN 结构 由三层组成:编码层,将输入向量编码成波的初始 振幅或相位;储蓄层,波按照非线性波动方程演化; 读出层,通过波动演化后,从最终状态读出结果. 由于该网络是储蓄池计算网络,只需对读出层的权



图 6 神经形态的非线性波动网络 (a) 单波层前馈网络 (SWFN) 包含编码层、储蓄层和读出层,其中波按照非线性偏微分方程 演化; (b) 偏差 ψ_b 的演化、波的演化以及近似计算一维函数 $y = \sin(\pi x) / (\pi x)$ 的结果; (c) SWFN 用于学习鲍鱼数据集; (d) 用孤 子训练通用的布尔逻辑门^[46]

Fig. 6. Neuromorphic computing by nonlinear waves: (a) Single wave-layer feed forward neural network (SWFN) with input layer, reservoir and readout layer, where the wave evolves according to a nonlinear partial differential equation; (b) the bias and wave evolution and results of learning the function $y = \sin(\pi x) / (\pi x)$; (c) results of learning the abalone dataset; (d) training a universal logic gates by soliton gases^[46].

值进行训练.除了该系统中用到的非线性薛定谔方程,其他任何非线性波动微分方程都可用于波的演化.事实上,任何具有非线性波动动力学特征的系统都可以用来建立神经形态的非线性波动网络.研究人员用不同的编码方法实现了三种具体应用:近似计算一维函数(图 6(b)),学习一个八维数据集(图 6(c)),实现布尔逻辑门(图 6(d)).三个例子中,SWFN都能与传统神经网络一样.这说明了 SWFN的通用性,它能够用于近似计算任意函数和学习高维数据集.这项基础工作阐明了非线性波与机器学习之间的联系,为电子学、光子学、自旋电子学、流体力学、玻色-爱因斯坦凝聚等领域的各种非线性波现象用作神经形态计算打开了大门.

储蓄池计算^[47,48] 是一类特殊的人工神经网络, 其作为中间层的储蓄层是随机生成的,且生成后就 保持不变,只需要训练输出层.科研工作者提出了 一个多功能的基于孤子的计算系统^[49],使用离散 孤子链作为储蓄池,通过利用其可调的控制动力 学,证明了足够强的非线性动力学能够实现对非线 性可分离数据集执行精确的回归和分类任务.由于 近年来科研工作者们才关注到非线性波中的机器 学习,相关的工作还很少.但是基于非线性波实现 储蓄池计算,通过储蓄层中的波传输携带大量信 息,能够学习更大尺度的数据集,并且这种方式往 往不需要严格控制传播介质,因此该方向值得深入 的研究探索.

4 基于扩散系统的无监督学习

扩散过程是物理学的经典过程之一. 微观上是 粒子无规则运动而导致了宏观迁移. 根据傅里叶定 理, 单位时间内通过垂直于扩散方向的单位面积截 面的扩散热量, 与扩散物质的温度梯度成正比. 因 此, 物质总趋向于由高势能区域流向低势能区域, 一段时间后达到稳态. 扩散系统在给定初始条件和 边界条件的情况下, 根据自身的规律进行演化, 最 终呈现稳定的分布, 这启发了许多新的无监督机器 学习算法. 这些算法通过计算、分析扩散过程稳态 时的概率分布实现对数据的分类、评级等功能. 下 面介绍若干扩散系统用于开发无监督学习算法的 示例, 包括概率扩散用于数据降维和分类, 热扩散 用于数据挖掘与社会网络推荐机制, 以及基于多体 扩散的扩散搜索算法用于结构搜索和材料预测.

4.1 基于扩散动力学的分类模型

分类是机器学习的重要任务之一. 流形学习能 够将真实世界中的高维数据映射到一个低维特征 空间, 从而根据数据的内在规律进行分类. 但是对 于非线性流形, 传统的线性映射方法并不可行, 因 此科学家们提出了扩散映射^[50,51]. 扩散映射的基本 思想是在数据图上定义一个扩散行为, 通过一段时 间的扩散, 逐渐滤除数据集中不重要的信息, 并得 到数据之间的相似度关系.

在具有 N 个数据点的数据集 $\{x_n\}_{n=1}^N, x_n \in \mathbb{R}^p$ 上定义一个 Markov 随机行走,则任意两个数据点 $x_i 和 x_j$ 之间的距离为

$$A_{i,j} = A_{\varepsilon} \left(x_i, x_j \right). \tag{8}$$

A定义为对应参数 ε 的核矩阵,常选用 exp $\left(-\frac{||x_i-x_j||^2}{2\varepsilon}\right)$ 的形式.定义对角矩阵**D** = diag([$D_1 \cdots D_N$]),其中 $D_j = \sum_{i=1}^N A_{i,j} = \sum_{i=1}^N A_{\varepsilon}(x_i, x_j)$. 本征方程为

 $\boldsymbol{A}\boldsymbol{D}^{-1} |\psi\rangle = \lambda |\psi\rangle \quad \vec{\mathbf{x}} \quad \boldsymbol{P} |\psi\rangle = \lambda |\psi\rangle, \qquad (9)$

其中 $P = AD^{-1}$ 是对核矩阵进行列归一化之后的 转移矩阵,用来描述扩散图上的随机行走, λ 是本 征值, $|\psi\rangle$ 是本征右矢.经过一个时间步长 ε ,数据 点 x_j 到 x_i 的扩散概率为

$$p\left(x^{t+\varepsilon} = x_i | x^t = x_j\right) = P_{i,j} = \frac{A_{\varepsilon}\left(x_i, x_j\right)}{D_j}$$
$$= \frac{A_{\varepsilon}\left(x_i, x_j\right)}{\sum_{i=1}^{N} A_{\varepsilon}\left(x_i, x_j\right)}.$$
(10)

经过 *n*个步长后,从数据点*x_j*出发到终点 *y*的扩散概率为

 $p(t = n\varepsilon, y|x_j) = p(x^t = y|x^0 = x_j) = P^n e_j$, (11) 其中 e_j 是一个仅在 j^{th} 处为 1, 其余位置为 0 的列 向量.转移矩阵 P可转化为对称矩阵 $P_s = D^{-1/2}PD^{1/2}$, P和 P_s 具有M个相同的本征值, 定 义 P_s 的本征矢为 $|\varphi_k\rangle$ 是列向量,则转移矩阵P的 第k个本征左矢和本征右矢分别为

$$\langle \phi_k | = |\varphi_k\rangle^{\mathrm{T}} \boldsymbol{D}^{-\frac{1}{2}}, \ |\psi_k\rangle = \boldsymbol{D}^{\frac{1}{2}} |\varphi_k\rangle.$$
(12)

 $\langle \phi_k | \pi | \varphi'_k \rangle$ 是双正交的, 即:

$$\langle \phi_k | \psi_{k'} \rangle = \delta_{k,k'}. \tag{13}$$

经过时间 t 后, 从点 x 扩散到 y 的概率为
$$p(t, \boldsymbol{y}|\boldsymbol{x}) = \langle \boldsymbol{y}|\psi_0 \rangle + \sum_{k \ge 1} \langle \boldsymbol{y}|\psi_k \rangle \,\lambda_k^t \,\langle \phi_k | \boldsymbol{x} \rangle \,, \quad (14)$$

其中 $\langle y | \pi | y \rangle$ 分别是表示 N个数据点的行向量 和列向量, 仅在 y处为 1, 其余处为 0. $|\psi_0\rangle$ 是特 征值 $\lambda_0 = 1$ 对应的 P矩阵的本征右矢, $\langle x_i | \psi_0 \rangle =$ $D_{i,i} / \sum_j D_{j,j} \cdot \lambda_k^n$ 是矩阵 P^n 的第 k个本征值. 数据 点 $x_i \pi x_j$ 之间的扩散距离定义为

$$d_t^2(x_i, x_j) = \sum_{\boldsymbol{y}=\boldsymbol{x}_1}^{x_N} \frac{\left(p\left(t, \boldsymbol{y}|x_j\right) - p\left(t, \boldsymbol{y}|x_i\right)\right)^2}{\langle \boldsymbol{y}|\psi_0\rangle}.$$
 (15)

将 (14) 式代入 (15) 式, 并结合 (12) 式和 (13) 式 中的关系, 扩散距离 (15) 式等于扩散空间中的欧 氏距离:

$$d_t^2(x_i, x_j) = \sum_{k=1}^{M-1} \lambda_k^{2n} (\langle \phi_k | x_j \rangle - \langle \phi_k | x_i \rangle)^2.$$
(16)

通过扩散,原空间中的高维数据被映射到 k 个特征 向量上,扩散距离的大小代表着数据之间的相似度 大小 (图 7(a)).经过长时间的演化,数据自然地进 行聚类,并在数据分布图和势能绘景图上显示出分 类信息 (图 7(b)).

最近,同济大学声子学课题组通过扩散映射, 实现了基于实空间动力学性质相似性的拓扑声子 无监督流形聚类 (图 7)^[21].用一个 $L \times L$ 阶的对角 矩阵 \hat{P} 描述声子系统谐振子之间的相互作用,对应 的核矩阵为 $A_{\varepsilon}(i,j) = \exp\left(-\frac{||\hat{P}_i - \hat{P}_j||^2}{2\varepsilon L^2}\right)$.拓扑 不变量v是关于几何参数构型 \hat{P} 的函数,所以可以 根据构型 \hat{P} 的差异对拓扑性质 $v(\hat{P})$ 进行分类.这 种聚类方法只需要定义矩阵 \hat{P} 和核矩阵 $A_{\varepsilon}(i,j)$, 不需要针对具体系统定义特定的拓扑不变量,因此 可用于不同的拓扑声子系统,包括随机耦合的一 维 Su-Schrieffer-Heeger 声子链(图 7(d))、不规则 声子拓扑绝缘体(图 7(e))、随机耗散的非厄米声子 链(图 7(f))和具有高阶声子拓扑态(图 7(g)).

基于扩散动力学的分类算法有很强的可适应 性,通过定义式(8)式中的核矩阵 $A_{i,j} = A_{\varepsilon}(x_i, x_j)$ 的具体形式,能够快速转变成适合不同物理情境 的分类算法. Rodriguez-Nieva 和 Scheurer^[52]提出 了基于扩散映射的无监督机器学习算法,能够对 二维 XY模型的样品进行卷绕数分类,并捕获 Berezinskii-Kosterlitz-Thouless 跃迁(图 8(a)).这 种方法也可以应用于伊辛规范理论,从而通过扩散 图进行拓扑分类,实现了对物质奇异相的完全无监 督研究. 继我们的工作之后^[21], Scheurer 和 Slager^[53] 将扩散映射算法用于搜索哈密顿量之间的绝热路 径,从而根据它们的拓扑性质对它们进行聚类. Lustig 等⁵⁴将类似的方法应用于从实验数据中识 别拓扑相变,分析了经历拓扑相变的光学系统的实 验数据,证明了即使数据来源于系统的一小部分 甚至不包括边缘态,扩散映射也能识别拓扑相变 (图 8(b)). Lidiak 和 Gong^[55] 提出了适用于量子系 统中的扩散映射算法,作为学习各种量子相位和相 变的通用工具 (图 8(c)). 这种方法可以在单一基础 上测量局部可观测值,例如测量单个方向上的所 有自旋,因此很容易应用于许多实验量子模拟器. 另外, Che 等^[56]提出了适用于动量空间的扩散映 射方法,成功识别拓扑特征,并且在动量空间中的 典型 Su-Schrieffer-Heeger 模型、Qi-Wu-Zhang 模 型和淬火 Su-Schrieffer-Heeger 模型上证明了这种 方法.

4.2 基于热传导扩散的推荐模型

除了基于概率扩散的机器学习算法,热传导扩 散也启发了新的机器学习算法.热传导系统中,由 于介质与介质之间存在温度差而产生传热,使能量 从物体的高温部分传至低温部分,经过一段时间后 形成稳定的温度分布.热传导机制能够有效地应用 于建立社会网络中的信息挖掘和推荐模型.这种方 法通过用户已选择的偏好项目(高温部分)推测出 用户可能选择的其他项目(低温部分).例如淘宝网 通过用户已购买的产品推荐其他类似的产品.基于 物理学中的热传导^[57,58],科研工作者们提出了可以 处理个性化边界条件的推荐模型,用于处理社会网 络中庞大的数据信息.

物理学中的热传导过程可以用偏微分方程:

$$-\kappa \nabla^2 T\left(\boldsymbol{r}\right) = \boldsymbol{\nabla} \cdot \boldsymbol{J}\left(\boldsymbol{r}\right), \qquad (17)$$

描述,其中 κ 为导热系数;T(r)为温度;J(r)为热 流密度.将包含N个项目的推荐网络视作一个含 有N个节点的热传导模型,定义R表示N个节点 的温度向量,已知的高温节点处温度为1,低温节 点处温度为0,我们的目标是得到热平衡状态时其 余节点的温度.节点之间的连接关系由对称邻接矩 阵A给出(图9):若两个项目直接相关,即两个节 点之间直接相连,则元素 $A_{ij}=1$,反之 $A_{ij}=0$.构 建转移矩阵 $P = D^{-1}A$,其中D是表示权重的对角



图 7 利用扩散映射实现典型声子系统中的流形聚类 (a) 流形空间降维; (b) 流形空间的样本数据分布与势能绘景; (c) 流形空间的扩散与凝聚, 稳态显示出天然的聚类; (d) 随机耦合的无序 Su-Schrieffer-Heeger(SSH) 声子链; (e) 无序非晶态声子的拓扑分 类; (f) 一维非厄米声子链; (g) 高阶拓扑声子^[21]

Fig. 7. Diffusion mapping in typical phononic systems to realize manifold clustering: (a) Dimension reduction in manifold space; (b) the probability distribution of samples and the effective landscape; (c) along with evolution, the samples diffuse and finally concentrate on positions with minimum local potentials, which indicates the clustering; (d)–(g) applications in disordered photonic SSH chain, amorphous topological phononics, 1D non-Hermitian phononic chain, high-order topological phononics^[21].

矩阵. 这个网络中的离散拉普拉斯算子 $L = \hat{I} - P$ 类似于 (17) 式中的 $-\kappa\nabla^2$,其中 \hat{I} 是单位矩阵.于是只需要求解

$$LR = f. \tag{18}$$

(18) 式与 (17) 式相类似,其中 *f* 是表示外部热源
 的向量,*f* 对应于∇ · *J*(*r*);温度向量 *R* 对应于*T*(*r*).

格林函数可以用来处理图上的扩散型问题^[22]. 在推荐模型中,温度向量 R即用户对项目的评级 向量, $R_i \neq 0$ 代表用户对第 i个项目的评级,对应 于热传导模型中已知的节点温度,若该用户未对该 项目进行过评级,则 $R_i = 0$,对应于热传导模型中 待确定的节点温度.所有已评级的项目 ($R_i \neq 0$) 是这个系统的边界条件.将用户的已知评级 (边界条件) 和未知评级分别用 **R**_B和 **R**_U表示,则 (18) 式 可展开成:

$$\begin{pmatrix} \mathbf{R}_{\mathrm{B}} \\ \mathbf{R}_{\mathrm{U}} \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{\mathrm{BB}} & \mathbf{P}_{\mathrm{BU}} \\ \mathbf{P}_{\mathrm{UB}} & \mathbf{P}_{\mathrm{UU}} \end{pmatrix} \begin{pmatrix} \mathbf{R}_{\mathrm{B}} \\ \mathbf{R}_{\mathrm{U}} \end{pmatrix} + \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \end{pmatrix}.$$
(19)

只需要求解

$$\boldsymbol{R}_{\mathrm{U}} = \boldsymbol{P}_{\mathrm{UB}}\boldsymbol{R}_{\mathrm{B}} + \boldsymbol{P}_{\mathrm{UU}}\boldsymbol{R}_{\mathrm{U}} + 0. \tag{20}$$

预测得到的评级向量为

$$\boldsymbol{R}_{\mathrm{U}} = \left(\hat{\boldsymbol{I}} - \boldsymbol{P}_{\mathrm{UU}}\right)^{-1} \boldsymbol{P}_{\mathrm{UB}} \boldsymbol{R}_{\mathrm{B}}.$$
 (21)

经过 n个时间步长之后, (21) 式的解为



图 8 基于扩散映射的无监督学习方法适用于解决不同物理系统中的拓扑分类问题 (a) 一维 XY 模型拓扑序的检测^[52]; (b) 扩散映射能够不借助边缘态, 识别 Haldane 模型描述的拓扑相变点^[54]; (c) 量子系统中的扩散映射算法, 能够无监督地识别 Z₃ 横场 伊辛模型的量子相^[55]

Fig. 8. The unsupervised learning with diffusion map is applied to solve topology identification in different physical systems: (a) Identifying the topological order in 1-dimensional XY model^[52]; (b) detection of the phase transition for the Haldane model without the edge states^[54]; (c) diffusion maps in learning quantum phases with a \mathbb{Z}_3 transverse field Ising model^[55].



图 9 标号图表明 6 个点 (项目) 之间的连接关系, 右侧是对应的权重矩阵和邻接矩阵

Fig. 9. Labelled graphs show the connection of 6 points (items), and the corresponding degree matrix and adjacency matrix are on the right side.

$$\boldsymbol{R}_{\mathrm{U}} = \left(\hat{\boldsymbol{I}} + \boldsymbol{P}_{\mathrm{UU}} + \boldsymbol{P}_{\mathrm{UU}}^{2} + \dots + \boldsymbol{P}_{\mathrm{UU}}^{n-1}\right) \boldsymbol{P}_{\mathrm{UB}} \boldsymbol{R}_{\mathrm{B}}.$$
(22)

因为 **P**_{UU} 的特征值小于 1, 所以 **P**ⁿ_{UU} 很快收敛, 于 是经过几个步长之后就能得到稳定解.这种方式避 免了对评级矩阵 **R** 的迭代求解, 而是将已知信息 作为边界条件直接得到最终稳态解, 从而减少计算 时间.其中 **P**_{UB}代表着已知评级节点到未知评级节 点的转移矩阵, **P**_{UU} 代表着未知评级节点之间的转 移矩阵.已知评级 **R**_B即热传导过程中的热源,未知 评级 **R**_U即其余节点平衡态时的温度.

热传导扩散和概率扩散都能应用于机器学习. 值得注意的是, 热传导模型中, 定义行归一化的转移矩阵 $P = D^{-1}A$. 拉普拉斯矩阵 $L = \hat{I} - D^{-1}A$ 表示该点的温度变化仅与汇入的净总热流有关. 例如图 9 中, 节点 4 的温度变化量仅与相连的节点 3, 5, 6 流入节点 4 的净热流的算术平均值有关, $\frac{\partial T_4}{\partial t} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} J_1 & J_2 & J_3 \end{bmatrix}^{\mathrm{T}}, J$ 为流人的 净热流. 而在概率扩散过程中, 定义列归一化的转 移矩阵 $P = AD^{-1}$, 拉普拉斯矩阵 $L = \hat{I} - AD^{-1}$ 表示扩散概率守恒. 例如图 9 中, 当时间步长 Δt 很 短时, 假设仅从节点 4 扩散到节点 3, 5 和 6, 那么 这四个节点概率之和为 1.

4.3 基于全局和局部扩散搜索算法的优化 模型

扩散搜索算法能够在庞大的数据集中, 通过随机且分布均匀的搜索方式实现信息的最优化处理, 被广泛的应用于结构搜索^[59-61]. Pickard 和 Needs^[62] 将随机扩散与第一性原理相结合, 提出了从头算随 机结构搜索算法 (*ab initio* random structure searching, AIRSS). 该算法以最随机的方式生成初始 结构, 为了提高效率可以考虑引入基于化学、实验 或对称性的偏置条件, 然后在保持实验和对称约 束的同时演化起始结构. AIRSS 的计算量集中在 演化大量不同的初始结构, 直到多次获得相同的 最低能量结构, 以确保该结构的势能面位于全局最 小值.

粒子群优化算法 (particle swarm optimization, PSO)^[63] 模拟自然界中鸟群的捕食行为, 它不同于 普通的单一粒子扩散行为,群体中每个粒子的扩散 搜索不仅受到自身个体极值影响,还受到整个粒子 群的当前全局最优解影响,最终实现全局或局部 扩散搜索最优解(图 10(a)).利用 PSO 进行晶体结 构搜索, 吉林大学马琰铭教授团队 [64,65] 开发了 CALYPSO, 全称为基于粒子群优化算法的晶体结 构分析 (crystal structure analysis by particle swarm optimization). 该方法只需要给定材料的化学成分 和外部条件,如压力,就能预测材料稳定或亚稳结 构,大大减少了第一性原理密度泛函计算的计算量 (图 10(b)). CALYPSO 算法的开发启发了很多原 创性工作,在设计各种材料方面具有广泛应用,为功能 驱动的材料设计打开了大门,具体内容可以参考 《Journal of Physics: Condensed Matter》,《Computational Materials Science》以及《Chinese Physics B》上的相关综述文章[66-68].

Gao 等^[69] 通过晶体结构搜索, 找到了三种新型的具有平面内负泊松比的氧化硅结构, 并且确认了二维氧化硅结构的全局最小自由能, 这在纳米力学和纳米电子学中有巨大的潜在应用. 基于 PSO



图 10 (a) 全局及局部粒子群优化算法示意图; (b) 粒子群优化算法中速度及位置更新示意图^[68]; (c) 多目标优化的二维 SnSe 材料定向设计工作流程图; (d) 室温下 (300 K), 二维 SnSe 材料单层结构的自由能; (e) 图 (d) 中第一 Pareto 前沿 (红线) 上的四种新型单层结构的三视图, 深灰色和绿色的球分别表示 Sn 原子和 Se 原子^[70]

Fig. 10. (a) The diagram of PSO; (b) The schematic diagram of the velocity and position updates in $PSO^{[68]}$; (c) workflow of the multi-objective optimization for 2D SnSe materials design; (d) thermopower landscape at room temperature (300 K) versus the free energy of 2D SnSe materials; (e) four 2D SnSe structures on the first Pareto front, where the dark gray and green balls denote Sn and Se atoms, respectively^[70].

的扩散搜索也可以是多目标的,多目标约束下功能 材料的定向设计是一个很大的挑战,其中性能和稳 定性是由不同物理因素的复杂关联决定的. 闫申申 等^[70]基于帕累托最优和粒子群优化方法的多目标 优化方法,对新型功能材料进行定向设计. 该工作 利用第一性原理结合多目标优化算法同时预测了 具有低自由能和高热电势的多种新型二维硒化锡 (图 10(c)—(e)),并且揭示了这些新型二维材料高 热电势来源于其费米面附近能带的多简并度. 基于 粒子群扩散的多目标优化方法能为未来多目标、多 功能材料的一体化设计提供一个新的思路.

除此之外,物理学中的扩散机制可以延伸到更 为宽泛的领域,比如利用极小值跳跃^[71]和微分演 化^[72]进行材料结构预测也可视作广义上的扩散过 程.扩散搜索算法有望应用于更多凝聚态物理学领 域,比如解决文章^[73]中提到到的光子拓扑态逆设 计问题,以及文章^[74]中提到的分子热流分束问题.

5 总结与展望

针对波动和扩散系统中的机器学习研究方兴 未艾.近年来人们关注该领域两个重点内容:其一 是在利用丰富的波物理现象作为实现机器学习的 硬件平台,以实现波动系统中的人工神经网络为主; 另一个是扩散系统启发新的机器学习算法,通过物 理中的扩散机制分析数据的内在规律,从而实现分 类、优化等.本文围绕着这两个方面简要介绍了相 关的进展以及一些前沿工作.

首先,本文介绍了波动系统中的神经网络实现,包括线性的光学、声学系统以及非线性波系统. 一系列相关工作说明了依赖于波的并行性和快速 传输性,波动系统中的神经网络具有高效、低能 耗、高带宽的特点.本文重点介绍了几个示例,凸 显出波动系统在推断视觉任务、时序任务或大数据 集任务时的优越性.波动系统作为人工神经网络硬 件载体具有巨大潜力,为下一代芯片的开发提供了 启发性的思路.再者,本文介绍了由扩散系统启发 的无监督机器学习算法.扩散系统中物质根据一定 规律扩散,最终达到稳态分布,这一机制开发了许 多机器学习算法,解决了许多具有类似特点的实际 问题.例如,基于概率扩散的分类算法,基于热传 导的社会网络推荐模型,以及基于群体扩散的结构 搜索算法.

尽管这个方向的研究已经取得巨大的进展,但 是仍处于初步阶段,一些重要的基本问题尚未解 决.首先,如何在波动系统中实现非线性激励函数 仍是个重要的问题.考虑到波动系统实现非线性激 励函数的复杂度和局限性,在波动系统中实现神经 网络是否优于传统方法 (如线性回归) 值得商權 [23]. 另外,在光学系统中可以通过有效的电光转换机制 进行信号恢复^[75]避免散粒噪声,但是在其他的波 动系统中如何避免噪声还不明确. 除此之外, 现有 的研究主要通过波动系统实现人工神经网络,突破 经典硬件平台的限制,或是根据扩散机制分析数据 的内在规律,从而实现无监督学习.二者之间的交 叉结合却鲜有讨论,扩散物理是否能与人工神经网 络相结合,从而进一步实现扩散系统中的深度学 习,以及如何实现硬件和软件的结合优化,这些都 是值得继续深入研究的问题. 最后, 在经典波动系 统或者经典扩散系统中实现类量子或量子启发算 法模拟也是一个重要的研究方向.

参考文献

- Eslami S M A, Jimenez R D, Besse F, et al. 2018 Science 360 1204
- [2] Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S 2010 11th Annual Conference of the International-Speech-Communication-Association Makuhari, Japan, September 26-30, 2010 pp1045-1048
- Krizhevsky A, Sutskever I, Hinton G E 2012 Advances in Neural Information Processing Systems 25 1097
- [4] Liakos K G, Busato P, Moshou D, Pearson S, Bochtis D 2018 Sensors 18 2674
- [5] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, Zdeborov'a L 2019 *Rev. Mod. Phys.* 91 045002
- [6] Wetzstein G, Ozcan A, Gigan S, Fan S, Englund D, Solja J M, Denz C, Miller D, Psaltis D 2020 Nature 588 39
- [7] Miller D A 2019 Adv. Opt. Photonics 11 679
- [8] Graves A, Wayne G, Reynolds M, Harley T, Danihelka I, Grabska-Barwińska A, Colmenarejo S G, Grefenstette E, Ramalho T, Agapiou J 2016 Nature 538 471
- [9] Brunner D, Soriano M C, Mirasso C R, Fischer I 2013 Nat. Commun. 4 1364
- [10] Mehta P, Bukov M, Wang C H, Day A G, Richardson C, Fisher C K, Schwab D 2019 Phys. Rep. 810 1
- [11] Sarantoglou G, Skontranis M, Bogris A, Mesaritakis C 2020 2020 Optical Fiber Communications Conference and Exhibition (OFC) San Diego, CA, USA, March 8–12, 2020 p3
- [12] Feldmann J, Youngblood N, Wright C D, Bhaskaran H, Pernice W H P 2019 Nature 569 208
- [13] Hornik K, Stinchcombe M, White H 1989 IEEE Trans. Neural Networks 2 359
- [14] Hopfield J J 1982 PNAS 79 2554
- [15] Farhat N, Psaltis D, Prata A, Paek E 1985 Appl. Opt. 24 10 1469

- [16] MacQueen J B 1967 Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1) (California: University of California Press) pp281–297
- [17] Dempster A P, Laird N M, Rubin D B 1977 J. R. Stat. Soc. Ser. B-Stat. Methodol. 39 1
- [18] Jolliffe I T 1986 Principal Component Analysis (New York: Springer) pp129–155
- [19] Roweis S T, Saul L K 2000 Science 290 2323
- [20] Tenenbaum J B, De Silva V, Langford J C 2000 Science 290 2319
- [21] Long Y, Ren J, Chen H 2020 Phys. Rev. Lett. 124 18 185501
- [22] Chung F, Yau S T 2000 J. Comb. Theory Ser. A 91 191
- [23] Jiao S, Gao Y, Feng J, Lei T, Yuan X 2020 Opt. Express 28 3 3717
- [24] Miller D 2013 Photonics Res. 1 1
- [25] Shen Y, Harris N C, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S, Larochelle H, Englund D, Soljačić M 2017 Nat. Photonics 11 441
- [26] Chakraborty I, Saha G, Roy K 2019 Phys. Rev. Appl. 11 014063
- [27] Chang J, Sitzmann V, Dun X, Heidrich W, Wetzstein G 2018 Sci Rep 8 12324
- [28] Zuo Y, Li B, Zhao Y, Jiang Y, Chen Y C, Chen P, Jo G B, Liu J, Du S 2019 *Optica* 6 1132
- [29] Lin X, Rivenson Y, Yardimci N T, Veli M, Luo Y, Jarrahi M, Ozcan A 2018 Science 361 1004
- [30] Qian C, Lin X, Lin X, Xu J, Sun Y, Li E, Zhang B, Chen H 2020 Light-Sci. Appl. 9 59
- [31] Zhou T, Fang L, Yan T, Wu J, Li Y, Fan J, Wu H, Lin X, Dai Q 2020 Photonics Res. 8 940
- [32] Dun X, Ikoma H, Wetzstein G, Wang Z, Cheng X, Peng Y F 2020 Optica 7 913
- [33] Weng J, Ding Y, Hu C, Zhu X F, Liang B, Yang J, Cheng J 2020 Nat. Commun. 11 6309
- [34] Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J, Fang L, Lin X, Dai Q 2019 Phys. Rev. Lett. 123 2 023901
- [35] Zangeneh-Nejad F, Sounas D L, Alù A, Fleury R 2021 Nat. Rev. Mater. 6 207
- [36] Zuo S Y, Wei Q, Cheng Y, Liu X J 2017 Appl. Phys. Lett. 110 011904
- [37] Zuo S Y, Tian Y, Wei Q, Cheng Y, Liu X J 2017 J. Appl. Phys. 123 091704
- [38]~ Zuo S, Wei Q, Tian Y, Cheng Y, Liu X 2018 $Sci \ Rep \ 8 \ 10103$
- [39] Zangeneh-Nejad F, Fleury R 2018 New J. Phys. 20 073001
- [40] Yang C, Liu T, Zhu J, Ren J, Chen H 2021 Phys. Rev. Appl. 15 044040
- [41] Hughes T, Williamson I A D, Minkov M, Fan S 2019 Sci. Adv. 5 eaay6946
- [42] Romera M, Talatchian P, Tsunegi S, Abreu Araujo F, Cros V, Bortolotti P, Trastoy J, Yakushiji K, Fukushima A, Kubota H, Yuasa S, Ernoult M, Vodenicarevic D, Hirtzlin T, Locatelli N, Querlioz D, Grollier J 2018 Nature 563 230
- [43] Khoram E, Chen A, Liu D, Ying L, Wang Q, Yuan M, Yu Z 2019 Photonics Res. 7 823
- [44] Zhao H 2021 Sci. China-Phys. Mech. Astron. 64 270511

- [45] Ren J, Wang W X, Li B, Lai Y C 2010 Phys. Rev. Lett. 104 058701
- [46] Marcucci G, Pierangeli D, Conti C 2020 Phys. Rev. Lett. 125 093901
- [47] Jaeger H 2001 GMD-Report 148, German National Research Institute for Computer Science
- [48] Maass W, Natschläger T, Markram H 2002 Neural Comput 14 2531
- [49] Silva N A, Ferreira T D, Guerreiro A 2021 New J. Phys. 23 023013
- [50] Nadler B, Lafon S, Coifman R, Kevrekidis I 2005 Appl. Comput. Harmon. Anal. 21 113
- [51] Coifman R R, Lafon S, Lee A B, Maggioni M, Nadler B, Warner F, Zucker S W 2005 Proc. Natl. Acad. Sci. U.S.A. 102 7426
- [52] Rodriguez-Nieva J, Scheurer M 2018 Nat. Phys. 15 790
- [53] Scheurer M S, Slager R J 2020 Phys. Rev. Lett. 124 226401
- [54] Lustig E, Yair O, Talmon R, Segev M 2020 Phys. Rev. Lett. 125 12 127401
- [55] Lidiak A, Gong Z X 2020 Phys. Rev. Lett. 125 225701
- [56] Che Y, Gneiting C, Liu T, Nori F 2020 Phys. Rev. B 102 134213
- [57] Zhang Y, Blattner M, Yu Y 2007 Phys. Rev. Lett. 99 154301
- [58] Ren J, Zhou T, Zhang Y 2008 EPL 82 58007
- [59] Oganov A, Pickard C, Zhu Q, Needs R 2019 Nat. Rev. Mater. 4 331
- [60] Schmidt J, Marques M, Botti S, Marques M 2019 NPJ Comput. Mater. 5 83
- [61] Zhang L, Wang Y, Lü J, Ma Y 2017 Nat. Rev. Mater. 2 17005
- [62] Pickard C J, Needs R J 2011 J. Phys. Condens. Matter 23 053201
- [63] Eberhart R, Kennedy J MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, October 4–6 1995 pp39–43
- [64] Wang Y, Lü J, Zhu L, Ma Y 2010 Phys. Rev. B 82 094116
- [65] Wang Y, Lü J, Zhu L, Ma Y 2012 Comput. Phys. Commun. 183 2063
- [66] Wang Y, Lü J, Zhu L, Lu S, Yin K, Li Q, Wang H, Zhang L, Ma Y 2015 J. Phys. Condens. Matter 27 20 203203
- [67] Wang H, Wang Y, Lü J, Li Q, Zhang L, Ma Y 2016 Comput. Mater. Sci. 112 406
- [68] Tong Q, Lv J, Gao P, Wang Y 2019 Chin. Phys. B 28 106105
- [69] Gao Z B, Dong X, Li N B, Ren J 2017 Nano Lett. 17 772
- [70] Yan S, Wang Y, Gao Z, Long Y, Ren J 2021 Chin. Phys. Lett. 38 027301
- [71] Geodecker S 2004 J. Chem. Phys. 120 9911
- [72] Wang L, Liu J G 2021 *Physics* 50 69 (in Chinese) [王磊, 刘金 国 2021 物理 50 69]
- [73] Long Y, Ren J, Li Y, Chen H 2019 Appl. Phys. Lett. 114 181105
- [74] Tan Y T, Wang L Q, Wang Z, Peng J, Ren J 2021 Chin. Phys. B 30 036301
- [75] Miller D 2017 J. Lightwave Technol. 35 346

SPECIAL TOPIC—Machine learning and physics

Machine learning based on wave and diffusion physical systems^{*}

Chen Jiang-Zhi¹⁾ Yang Chen-Wen¹⁾ Ren Jie^{1)2)†}

 (Shanghai Key Laboratory of Special Artificial Microstructure Materials and Technology, Center for Phononics and Thermal Energy Science, School of Physics Science and Engineering, Tongji University, Shanghai 200092, China)

2) (Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, China)

(Received 10 May 2021; revised manuscript received 1 July 2021)

Abstract

Recently, the application of physics to machine learning and the interdisciplinary convergence of the two have attracted wide attention. This paper focuses on exploring the internal relationship between physical systems and machine learning, and also on promoting machine learning algorithm and physical implementation. We summarize the researches of machine learning in wave systems and diffusion systems, and introduce some of the latest research results. We first discuss the realization of supervised learning for wave systems, including the wave optics realization of neural networks, the wave realization of quantum search, the recurrent neural networks based on wave systems, and the nonlinear wave computation of neural morphology. Then, we discuss the machine learning algorithms inspired by diffusion systems, such as the classification algorithm based on diffusion dynamics, data mining and information filtering based on thermal diffusion, searching for optimization based on population diffusion, etc. The physical mechanism of diffusion system can inspire the construction of efficient machine learning algorithms for the classification and optimization of complex systems and physics research, which may create a new vision for the development of physics inspired algorithms and hardware implementation, and even the integration of software and hardware.

Keywords: wave systems, diffusion systems, machine learning, artificial neural network

PACS: 42.79.Ta, 05.40.Fb, 05.90.+m

DOI: 10.7498/aps.70.20210879

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 11935010, 11775159), the Natural Science Foundation of Shanghai Science and Technology Committee, China (Grant No. 18JC1410900), the Opening Project of Shanghai Key Laboratory of Special Artificial Microstructure Materials and Technology, China, and the Fundamental Research Funds for the Central Universities, China.

 $[\]dagger$ Corresponding author. E-mail: <code>xonics@tongji.edu.cn</code>

物理学报Acta Physica Sinica





Institute of Physics, CAS

自动微分及其在物理模拟中的应用

刘金国 许开来

Automatic differentiation and its applications in physics simulation

Liu Jin-Guo Xu Kai-Lai

引用信息 Citation: Acta Physica Sinica, 70, 149402 (2021) DOI: 10.7498/aps.70.20210813 在线阅读 View online: https://doi.org/10.7498/aps.70.20210813 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

道路交通的流体物理模型与粒子仿真方法

Physics model of fluid and particle simulation method for road traffic 物理学报. 2017, 66(22): 224501 https://doi.org/10.7498/aps.66.224501

量子计算与量子模拟

Quantum computation and quantum simulation 物理学报. 2018, 67(12): 120301 https://doi.org/10.7498/aps.67.20180710

空域模拟光学计算器件的研究进展

Advances in spatial analog optical computing devices 物理学报. 2020, 69(15): 157803 https://doi.org/10.7498/aps.69.20200283

新型超导量子比特及量子物理问题的研究

Novel superconducting qubits and quantum physics 物理学报. 2018, 67(22): 228501 https://doi.org/10.7498/aps.67.20180845

非晶材料与物理近期研究进展

Recent progress of the glassy materials and physics 物理学报. 2018, 67(12): 126101 https://doi.org/10.7498/aps.67.20180681

温度梯度区域熔化作用下熔池迁移的元胞自动机模拟

Cellular automaton simulation of molten pool migration due to temperature gradient zone melting 物理学报. 2019, 68(4): 048102 https://doi.org/10.7498/aps.68.20181587

专题: 机器学习与物理

自动微分及其在物理模拟中的应用*

刘金国1)† 许开来2)

1) (哈佛大学物理系, 坎布里奇 02138)

2) (斯坦福大学, 斯坦福 94305)

(2021年4月27日收到; 2021年6月17日收到修改稿)

自动微分是利用计算机自动化求导的技术,最近几十年因为它在机器学习研究中的应用而被很多人了 解.如今越来越多的科学工作者意识到,高效、自动化的求导可以为很多科学问题的求解提供新的思路,其中 自动微分在物理模拟中的应用尤为重要,而且具有挑战性.物理系统的可微分模拟可以帮助解决混沌理论、 电磁学、地震学、海洋学等领域中的很多重要问题,但又因为其对计算时间和空间的苛刻要求而对自动微分 技术本身提出了挑战.本文回顾了将自动微分技术运用到物理模拟中的几种方法,并横向对比它们在计算时 间、空间和精度方面的优势和劣势.这些自动微分技术包括伴随状态法,前向自动微分,后向自动微分,以及 可逆计算自动微分.

关键词:自动微分,科学计算,可逆计算,最优检查点,物理模拟 PACS: 02.60.Pn, 02.30.Jr, 91.30.-f DO

DOI: 10.7498/aps.70.20210813

1 引 言

自动微分^[1]是指自动获取一段计算机程序导数的技术,很多人对它的了解源自它在机器学习中的成功应用,人们可以用它优化包含数千亿参数的神经元网络^[2].与很多人印象不同的是,自动微分其实是个很古老的技术.Nolan^[3]早在他1953年的博士论文中就提出过计算机自动化求导的构想,后来针对这一构想又出现了两种不同的实践,分别是1964年由Wengert^[4]实现的前向自动微分和1970年Linnainmaa^[5]实现的后向自动微分.而最近十几年,由于后向自动微分在机器学习中的广泛应用,相关技术越来越成熟并在科学计算中也得到了越来越多的应用.科学家们利用方便的、自动化的计算机辅助求导解决了很多重要的物理问题,其中包括帮助变分蒙特卡罗设计更加通用的多体物理波函数^[6-9],加速变分量子算法的模拟^[10],拓展

变分张量网络算法^[11],以及求解自旋玻璃基态构型^[12]等.

对物理模拟过程的自动微分是自动微分重要 的应用之一,其中最大的技术挑战是对电磁学、海 洋学[13] 和地震学[14,15] 等问题中最核心的微分方程 求解过程的自动微分,这些微分方程常见的求解方 法是先将问题的时空坐标离散化,并以数值积分的 形式完成求解. 要得到精确的结果, 离散化后的网 格需要设计得很稠密,从而对存储空间和计算时间 的需求巨大. 与此同时, 在自动微分计算过程中, 机器学习中主流的后向自动微分库,如 Jax^[16]和 PyTorch^[17], 需要很多额外的空间来缓存中间结 果,导致人们经常无法直接用它们对具有一定规模 的实际物理模拟问题求导.一种传统的解决常微分 方程求导的方案叫作伴随状态法[18,19], 它假设了在 较短时间内积分器可逆,并通过逆向积分来帮助自 动微分回溯状态.事实上,除了 Leapfrog 等少数辛 积分器外,大多数的积分器并不能保证严格的时间

^{*} QuEra computing 公司资助的课题.

[†] 通信作者. E-mail: jinguoliu@g.harvard.edu

^{© 2021} 中国物理学会 Chinese Physical Society

反演对称,所以伴随状态法往往存在与积分步长有 关的系统性误差.后来,有人把机器学习中的最优 检查点算法带到了物理模拟的反向自动微分中^[14], 到了严格求导.也有一些人直接利用包含最优检 查点算法^[20,21]的自动微分编译器,如 Tapenade^[22], OpenAD^[23]等,编译模拟代码实现类似的时间和 空间的权衡.在最近几年,可逆计算开始被用作一 种新的通用的微分的方案^[24],也可以做到对数的 额外空间开销.可逆计算提供了传统自动微分框架 所不具有的内存控制的灵活性,因此可以被用来微 分 GPU 核函数、复杂的控制流等.

本文将会回顾共轭态方法,前向自动微分以及 基于最优检查点算法和可逆计算的后向自动微分 等自动微分方法在物理模拟问题中的应用,并对比 不同方法的优劣以及适用的场景.第2节介绍了几 种自动微分方法的基本理论.第3节介绍了不同自 动微分技术在波的传播模拟中的应用.关于如何在 后向自动微分中权衡程序的运行时间和空间的理 论,虽然重要,但受篇幅限制我们将其放在附录.

2 自动微分方法

物理模拟过程的常见求解方案是先将偏微分 方程的空间部分离散化并做差分处理^[26],从而将 其转换为对时间的常微分方程

$$\frac{\mathrm{d}s}{\mathrm{d}t} = f(s, t, \theta),$$

其中s为状态,t为时间, θ 为控制参数.一个常微 分方程的数值求解器会把时间维度离散化,做n步 叠代求解,每步仅做从时刻 t_i 到时刻 $t_{i+1} = t_i + \Delta t$ 的演化,

$$s_n = \text{ODESolve}(f, s_0, \theta, t_0, t_n)$$

= $(s_{i+1} = \text{ODEStep}(f, s_i, \theta, t_i, \Delta t)$
for $i = 0, 1, \cdots, n-1$, (1)

其中 s_i 为完成第i步积分后的状态.为了简便,下 文将单步运算简记为ODEStep(s_i).最后我们还会 定义一个损失函数 $\mathcal{L} = loss(s_n)$.自动微分的目标 则是求解损失量对输入状态 $\frac{\partial \mathcal{L}}{\partial s_0}$ 和控制参数的导 数和 $\frac{\partial \mathcal{L}}{\partial \theta}$.

2.1 共轭态方法

共轭态方法[18,19] 是专门针对常微分方程的反

向传播方法.在研究中,人们发现积分过程的导数 的反向传播同样是一个积分过程,只不过方向相 反.于是人们通过构造一个可以同时回溯函数值和 回传导数的拓展函数,以对拓展函数的逆向积分的 形式完成导数的计算,如表1中算法1所示.该算 法的描述来自文献 [19],其中可以找到详细的推导 过程,这里对原算法中的符号做了替换以方便读者 理解.算法中的 $\frac{\partial q}{\partial s}, \frac{\partial q}{\partial \theta}$ 以及 $\frac{\partial L}{\partial s_n}$ 为局域导数,它们 可以通过手动推导或借助其他自动微分库来实现. 该方案在积分器严格可逆的时候梯度也严格,而当 积分器反向积分误差不可忽略时,则需要额外的处 理保证误差在可控范围,这一点我们会在随后的例 子中涉及.

Table 1. Continuous adjoint state method (Algorithm 1).

 输入: 动力学参数
$$\theta$$
, 开始时间 t_0 , 结束时间 t_n , 末态 s_n , 以及需要回传的导数 $\frac{\partial \mathcal{L}}{\partial s_n}$

 输出: $\frac{\partial \mathcal{L}}{\partial s_0}$, $\frac{\partial \mathcal{L}}{\partial \theta}$

 1 function aug_dynamics($(s, a, -), t, \theta$)

 2 $q = f(s, t, \theta)$
 #定义拓展动力学函数

 3 return $\left(q, -a^T \frac{\partial q}{\partial s}, -a^T \frac{\partial q}{\partial \theta}\right)$

 4 end

 5 $S_0 = \left(s_n, \frac{\partial \mathcal{L}}{\partial s_n}, 0\right)$
 #计算拓展动力学函数的初始状态

 6 $\left(s_0, \frac{\partial \mathcal{L}}{\partial s_0}, \frac{\partial \mathcal{L}}{\partial \theta}\right)$
 ODESolve(aug_dynamics, S_0, θ , t_n, t_0)

2.2 前向自动微分

顾名思义,前向自动微分是指向前(指与程序运行方向相同)传播导数.这里导数传播的规则和数学分析中的无穷小量的运算规则有关.数学中,在对一个输入变量p求导时,会让它携带一个无穷小量dp,并通过对这个无穷小量的运算完成对程序的求导.例如当作用函数f时,会有如下链式法则:

$$f\left(\vec{x} + \frac{\mathrm{d}\vec{x}}{\mathrm{d}p}\mathrm{d}p\right) = f(\vec{x}) + \left(\frac{\mathrm{d}f(\vec{x})}{\mathrm{d}\vec{x}}\frac{\mathrm{d}\vec{x}}{\mathrm{d}p}\right)\mathrm{d}p,\qquad(2)$$

其中 *x* 为函数 *f* 的输入参数的集合,可以包括 *p*本身. $\frac{df(\vec{x})}{d\vec{x}}$ 为局域雅可比矩阵,前向传播中我们将它与梯度矢量相乘得到新的梯度矢量. 实际程序实

现中,这个局域雅可比矩阵并不需要构造出来.考虑到任何程序都具有可拆分为基础指令的特点,人 们把程序拆解为基础标量指令,并在这些基础指令 上通过代码变换或算符重载的方式实现梯度矢量 的计算.以标量的乘法为例,变量会同时记录它的 数值和一阶小量的系数 (v, \dot{v}) ,其中 $\dot{v} = \frac{dv}{dp}$,人们重 新定义它的基本运算规则如下:

 $\dot{*}: ((a,\dot{a}), (b,\dot{b})) \mapsto (a*b, a\dot{b} + b\dot{a}).$

使得其在计算同时更新一阶小量的系数,而自动微分所要做的就是将所有的运算指令都引入小量的系数并做上述运算规则的替换.简单的运算规则的 替换对于人类来说尚可手动,但真实的程序可能会 包含数以亿计的这样的基础操作,虽然结果依然是 解析的,但是人们很难再通过人力得到具体的导数 表达式,而计算机恰恰很擅长这样繁琐但是规则简 单的任务.如图 1(a)所示,在求解常微分方程中, 单步前向自动微分可以形式化的表达为

$$ODEStep^{F} : \left(s_{i}, \frac{ds_{i}}{ds_{0}}, \frac{ds_{i}}{d\theta}\right) \mapsto \\ \left(ODEStep\left(s_{i}\right), \frac{\partial s_{i+1}}{\partial s_{i}} \frac{ds_{i}}{ds_{0}}, \frac{\partial s_{i+1}}{\partial s_{i}} \frac{ds_{i}}{d\theta} + \frac{\partial s_{i+1}}{\partial \theta}\right)$$

这里为了简洁略去了时间等常数参量.由于状态 s_i,s_{i+1}和控制参数 θ 均可包含多个变量,上述偏微 分均解释为雅可比行列式.一般前向自动微分在一 次前向运行中只对一个或若干个变量求导,因此计 算空间会随着一次求导的变量数目线性增加.无论 是一次求导多少个变量,前向自动微分求导的时间 都会随着需求导的变量的数目线性增长,这是限制 前向自动微分应用场景的最主要因素.

2.3 后向自动微分

后向自动微分与前向自动微分梯度传播方向 相反, 它解决了前向自动微分中计算开销随着需要 求导的变量数目线性增长的问题. 后向自动微分包 括正向计算和梯度反向传播两个过程. 正向计算过 程中, 程序进行普通的计算并获取所需的运行时信 息, 最后得到一个标量损失 *L*. 梯度回传的过程是 计算导数的过程, 可表示为更新变量的梯度 d*L*



图 1 物理模拟程序中对状态变量梯度的 (a) 前向传播和 (b) 后向传播过程. 其中线条上的箭头代表运算的方向, 圆 圈为后向自动微分中缓存的变量

Fig. 1. The (a) forward mode and (b) backward mode gradient propagation for state variables in a physics simulation. The arrows are computation directions, and circles are cached states in the backward mode automatic differentiation.

过程. 从 $\frac{d\mathcal{L}}{d\mathcal{L}} = 1$ 出发, 梯度反向传播对应如下链式 法则:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\vec{x}} = \sum_{i} \frac{\partial\mathcal{L}}{\partial\vec{y_i}} \frac{\mathrm{d}\vec{y_i}}{\mathrm{d}\vec{x}}.$$

为了实现该链式法则,软件包设计者们一般会在软件中规定一类可微分的函数集合,叫作原子函数 $\{f_p\}$,并定义了梯度在原子函数 $\vec{y} = f_p(\vec{x})$ 上的回传规则为伴随变量 $v = \frac{d\mathcal{L}}{d\vec{v}}$ 的更新规则,

$$\overline{f}_p:(\overline{x},\overline{y})\mapsto \left(\overline{x}+\overline{y}\frac{\partial \vec{y}}{\partial \vec{x}},\overline{y}\right).$$

其中, $\frac{\partial g}{\partial x}$ 为局域雅可比矩阵, 它不需要具体计算出 来, 而是通过后向传播函数实现. 根据实现不同, 这些原子函数可以是 Tapenade 和 NiLang^[24] 那样 仅包含有限个基础标和 Pytorch 中那样的可拓展 的常用函数集合. 用户将程序表示为原子函数的组 合, 在前向计算中, 计算机将遇到的原子函数的组 合, 在前向计算中, 计算机将遇到的原子函数的组 向传播函数连同所需的中间变量一起存放在堆栈 中, 并在反向传播中按照后进先出的顺序调用, 这 样程序便可以利用上述梯度回传规则更新伴随变 量. 具体到求解常微分方程的自动微分过程中, 其 反向传播过程如图 1(b)所示, 可以形式化的表达为

$$\overline{\text{ODEStep}}^{\text{B}}: (\overline{s_{i+1}}, \overline{\theta}, s_i) \mapsto \left(\overline{s_{i+1}} \frac{\partial s_{i+1}}{\partial s_i}, \overline{\theta} + \overline{s_{i+1}} \frac{\partial s_{i+1}}{\partial \theta}\right),$$

#正向计算(单步) $push(\Sigma, s_i),$ $s_{i+1} = \text{ODEStep}(s_i),$

这里也同样略去了时间等参量, push 和 pop分别 是对全局堆栈Σ的入栈和出栈操作.虽然导数回传 的计算复杂度与需要求导的变量数目无关,但后向 自动微分向堆栈中存储数据带来了正比于计算步 骤数的额外空间开销. 如何设计一个可以对计算状 态逆序的访问的算法,使得缓存带来的时间和空间 开销更小,是反向自动微分设计复杂性的来源,也 被称作"时间与空间的权衡"问题. 实际应用中, 人 们可以用检查点^[20]或者反向计算^[24]技巧来避免 缓存全部中间结果. 如图 2(a) 所示的检查点方案 中,正向计算中程序可以选择性的缓存(黑色圆 点)或者不缓存部分状态(空心圆点),这些被缓存 的状态被称作检查点. 在图中下方的反向传播过程 中,当需要的数据没有被缓存时,程序会从最近的 检查点出发重新计算该数据(下方步骤 2). 图 2(b) 所示的反计算方案是可逆计算中常用的避免缓存 的方案. 当运算在数学层面不可逆, 可逆计算要求 保留输入状态以使其在程序层面可逆,因此默认也 需要缓存每步的结果.而反计算通过反向运行(上 方步骤 5) 清零已分配的内存的方式并将空间资源 归还给系统控.在反向传播过程中,执行顺序和前 向镜像对称,每个指令变成原指令的逆,因此可以 自然地获得运行时的状态信息. 不论是检查点方案 还是可逆计算方案,节省内存都需要消耗更多的 时间,那么如何权衡两者的额外开销才是最好呢? 我们在附录中引入一个简单的理论模型——鹅 卵石游戏并详细讨论了如何在检查点方案中用 Treeverse 算法^[20](也叫作最优检查点算法)实现仅 需对数的额外时间和空间逆向遍历状态,以及如何 在可逆计算中用 Bennett 算法^[25] 实现对数的额外 空间和多项式的额外时间实现逆向遍历状态,这里 仅把基本结论列于表 2 中.

自动微分在物理模拟中的应用 3

本节有两个案例,其一是用前向自动微分和伴 随状态法求解参数较少的洛伦茨系统模拟的导数,

#反向传播 (单步)

$$s_i = pop(\Sigma),$$

 $(\overline{s_i}, \overline{\theta}) = \overline{ODEStep}^B(\overline{s_{i+1}}, \overline{\theta}, s_i).$

其二是用最基于最优检查点算法和可逆计算的后 向微分对步骤较多且内存消耗巨大的地震学模拟 的求导.这些例子的源代码可在 Github 仓库中找 到: https://github.com/GiggleLiu/WuLiXueBao.



图 2 使用 (a) 检查点方案和 (b) 反计算方案避免缓存全 部中间状态.图中黑箭头为正常计算过程,红箭头为梯度 回传过程,蓝箭头为梯度回传和反向计算的复合过程,箭 头上的数字代表了执行顺序.黑色和白色的圆点为被缓存 和未被缓存 (或通过反计算消除) 的状态

Fig. 2. Using (a) checkpointing scheme and (b) reverse computing scheme to avoid caching all intermediate states. The black arrows are regular forward computing, red arrows are gradient back propagation, and blue arrows are reverse computing. The numbers above the arrows are the execution order. Black and white circles represent cached states and not cached states (or those states deallocated in reverse computing) respecitively.

表 2 不同自动微分方案中,时间与空间复杂度的 关系. 这里伴随状态法没有考虑了缓存部分中间结 果以保证反向积分的正确性,考虑之后应为O(TS). 前向自动微分时间复杂度中的 N 代表了需要求导 的参数个数. 可逆计算中的多项式时间复杂度中的 $\epsilon > 0$

Table 2. The time and space complexities in different automatic differentiation schemes. Here, the adjoint state method does not consider the effect of round off error into consideration. The parameter Nin the forward mode autodiff represent the number of differentiable input parameters. The ϵ in the reverse computing is greater than 0.

方法	时间	空间	是否严格
伴随状态法	$\mathcal{O}(T)$	$\mathcal{O}(S)$	否
前向自动微分	$\mathcal{O}(NT)$	$\mathcal{O}(S)$	是
基于最优检查点的 后向自动微分	$\mathcal{O}(T\log T)$	$\mathcal{O}(S\log T)$	是
基于可逆计算的 后向自动微分	$\mathcal{O}(T^{1+\epsilon})$	$\mathcal{O}(S\log T)$	是

3.1 洛伦茨方程求解

洛伦茨系统^[27] 是人们研究混沌问题的经典模型, 它描述了一个定义在三维空间中的动力学

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \sigma(y-x),\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= x(\rho-z) - y,\\ \frac{\mathrm{d}z}{\mathrm{d}t} &= xy - \beta z. \end{aligned}$$

其中, σ , ρ 和 β 为三个控制参数.该系统的含时 演化的曲线如图 3(b) 所示. ρ>1时,系统有两 个吸引子^[28], 但仅当 $\rho < \sigma \frac{\sigma + \beta + 3}{\sigma - \beta - 1}$ 时才会出现 图 3(b) 中橘色曲线所示的粒子稳定地围绕其中一 个吸引子运动的情况,这时系统较为稳定并表现出 对初值较为不敏感的特点. 末了位置坐标对控制参 数和初始坐标的导数反映了末态对控制参数和初 始位置的敏感度,在一定程度上反映了系统出现了 混沌现象. 在数值模拟中, 我们用 4 阶龙格库塔方 法对时间积分并得到末了位置,模拟中固定初始位 置为 $(x_0, y_0, z_0) = (1, 0, 0)$, 控制参数为 $\beta = 8/3$, 积 分时间区间为[0, T = 30]以及积分步长为 3×10^{-3} . 由于该过程所含参数仅有6个,包括初始位置的三 个坐标 (x_0, y_0, z_0) 和三个控制参数 (σ, ρ, β) , 因此用 前向自动微分工具 ForwardDiff^[29] 求导比后向自 动微分有很大优势. 我们把导数的绝对值的平均与 初始 ρ,σ 的关系画在图 3 中. 可以看出只有在理论

预言的黑线下方才会有较小的导数,说明稳定吸 引子参数下系统动力学的确对初值依赖性较低. 表3对比了不同方法的时间和空间的消耗,可以看 到前向自动微分所用的时间仅为原代码的不到 2倍.这里的高效来自ForwardDiff中允许一次对 多个变量求导,代价是用了正比于参数数目倍数的 空间.该技术虽然没有改变随着输入变量数目增 加,计算复杂度线性增加的本质,但是减少了线性 部分的系数,对处理实际问题很有帮助.基于可逆 计算的后向自动微分库 NiLang^[24]需要的计算时 间为原计算的约 3.5 倍,其中包含了前向计算和反 向传播过程,因此这个速度并不算差.但是由于龙

表 3 不同方法对 10⁴步洛伦茨系统积分过程的微 分所需时间和空间,其中空间部分以状态数目为单 位,时间的估计通过 *BenchmarkTools.jl*软件包多 次统计得出.这里,NeuralODE 中单步计算的微分 由 NiLang 完成,检查点的步长为 200 步 Table 3. The time and space to simulate a 10⁴ step Lorenz integration, where the space is measured in unit of single state size, and the time is obtained by repeated sample using the Julia package *BenchmarkTools.jl.* Here, the single step back propagation in NeuralODE is obtained with NiLang, and the checkpoint step size is 200.

方法	Julia	ForwardDiff	NiLang	Neural ODE + NiLang
时间/ms	1.90	2.88	6.70	34.30
空间 (估计)	1	6	10^{4}	50



图 3 (a) 洛伦茨系统中固定参数 $\beta = 8/3$, 梯度大小与参数 $\rho \pi \sigma$ 的关系. 图中颜色代表了平均梯度大小, 黑线是理论上会不会存在稳定吸引子分界线. (b) 中的蓝色和黄色的线分别对应 (a) 图标识的蓝点处参数 ($\sigma = 10, \rho = 27$)和黄点处参数 ($\sigma = 10, \rho = 15$) 对应的动力学模拟

Fig. 3. (a) The magnitude of gradients as a function of parameters ρ and σ in a Lorenz system with β fixed to 8/3. The black line is a theoretical predicted stable attractor phase transition line. The blue and yellow curves in (b) are simulations at parameters corresponding to the blue ($\sigma = 10$, $\rho = 27$) and yellow ($\sigma = 10$, $\rho = 15$) dots in (a). 格库塔积分器不可逆,在不利用额外的计算时间交 换空间的情况下,需要缓存每一步的计算以保证可 逆性,因此要求有104倍于状态大小的缓存空间. 好在该问题的单个状态空间仅有三个维度,不做任 何缓存求导仍然很容易. 表中最后一列的伴随状态 法的单步求导利用了 NiLang 的自动微分, 虽然伴 随状态法在理论上可以做到无额外内存消耗,但是 会引入由于积分器不可逆带来的系统误差,这对于 研究混沌问题非常致命. 我们以前向自动微分的导 数是严格的导数作为基准,把伴随状态法求得的导 数的1²相对误差与积分步长的关系绘制于图 4(a) 中,可以看出相对误差随着积分步长指数增加.因 此,我们需要每隔一段积分,就设置一个检查点重 新加载正确的坐标. 图 4(b) 显示检查点越密, 误差 越小,消耗的额外空间也越多.在表格中的模拟中, 我们选择了检查点步长为 200, 对应检查点数目为

50, 相对误差约为10-5.

3.2 波的传播方程的微分

考虑由如下方程决定的波函数*u*(*x*₁, *x*₂, *t*)在非均匀二维介质中的传播过程:

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} - \nabla \cdot (c^2 \nabla u) = f, & t > 0, \\ u = u_0, & t = 0, \\ \frac{\partial u}{\partial t} = v_0, & t = 0. \end{cases}$$
(3)

其中 c 为波在介质中的传播速度.理想匹配层 (PML)^[30-32] 是模拟波在介质中运动的一种准确可 靠的方案,为了在有限尺寸模拟该动力学,PML 方 法引入了吸收层防止边界反射的影响.在引入辅助 场并对空间和时间进行离散化后,上述方程可变形 为如下数值计算过程:

$$\begin{cases} u_{i,j}^{n+1} \approx \frac{\Delta t^2}{1 + (\zeta_{1i} + \zeta_{2j})\Delta t/2} \Biggl((2 - \zeta_{1i}\zeta_{2j}) u_{i,j}^n - \frac{1 - (\zeta_{1i} + \zeta_{2j})\Delta t/2}{\Delta t^2} u_{i,j}^{n-1} + c_{i,j}^2 \frac{u_{i+1,j}^n - 2u_{i,j}^n + u_{i-1,j}^n}{\Delta x^2} \\ + c_{i,j}^2 \frac{u_{i,j+1}^n - 2u_{i,j}^n + u_{i,j-1}^n}{\Delta y^2} + \frac{(\phi_x)_{i+1,j} - (\phi_x)_{i-1,j}}{2\Delta x} + \frac{(\phi_y)_{i,j+1} - (\phi_y)_{i,j-1}}{2\Delta y} \Biggr),$$

$$(4)$$

$$(\phi_x)_{i,j}^{n+1} = (1 - \Delta t\zeta_{1i})(\phi_x)_{i,j}^n + \Delta t c_{i,j}^2(\zeta_{1i} - \zeta_{2j}) \frac{u_{i+1,j} - u_{i-1,j}}{2\Delta x},$$

$$(\phi_y)_{i,j}^{n+1} = (1 - \Delta t\zeta_{2j})(\phi_y)_{i,j}^n + \Delta t c_{i,j}^2(\zeta_{2j} - \zeta_{1i}) \frac{u_{i,j+1} - u_{i,j-1}}{2\Delta y},$$

这里的第一项为近似,因为它忽略了原式中介质传播速度c的空间梯度项的贡献. ζ_1 和 ζ_2 分别是x和

y方向的衰减系数, $\phi_x 和 \phi_y$ 分别为引入的辅助场的 $x \pi y$ 方向的分量, Δx , $\Delta y \pi \Delta t$ 分别是空间和时



图 4 (a) 利用共轭态方法对 $\rho = 27, \sigma = 10, \beta = 8/3$ 的洛伦茨系统求导时,相对 l^2 误差与积分步数的关系.其中一个点代表了 在该步长下,对 100 个随机初始点计算得到的中位数,缺失的数据代表该处出现数值溢出.(b) 每隔固定步长设置检查点后, 10⁴ 步积分的误差

Fig. 4. (a) Relative l^2 error as a function of the integration step in differentiating the Lorenz equation solver at parameters $\rho = 27$, $\sigma = 10$, $\beta = 8/3$ using the adjoint state method. Each data point in the graph represents a median of 100 random samples. Some data are missing due to the numerical overflow. (b) The same error curve of computing 10^4 steps with checkpointing to decrease the round off error.

间的离散化参数.该方程的详细推导可参考文 献 [26]. PML 模拟的自动微分在地震学中有着重 要的应用^[15],而且人们很早就认识到检查点方案 可以用于地震波模拟中^[14],从而使回溯中间状态 的内存需求大大减少.

在数值模拟中,我们用双精度浮点数模拟了 1000×1000的二维格点上的 PML 方程求解, 每个 状态要存储 4 个矩阵 $s_n = \{u^{n-1}, u^n, \phi_x^n, \phi_y^n\},$ 占用 存储空间为 32 MB. 虽然前向自动微分可以仅用 常数倍空间开销微分该程序,但由于仅传播速度c 就包含106个参数,前向自动微分带来的线性时间 复杂度的增加是不可接受的.同时,若不做任何内 存优化对该程序后向自动微分,积分104步需要存 储空间至少为 320 G, 远超出了普通 GPU 的存储 能力.此时,我们需要用附录中描述的 Bennett 算 法和 Treeverse 算法来节省后向自动微分的缓存. 图 5 中展示了这两种时间空间交换方案下,实际程 序在 GPU 上运行的时间与空间的关系, 计算设备 为 Nvidia Tesla V100. 纯可逆计算实现的 Bennett 算法中, 计算梯度的部分随着前向计算的步 骤数的增加而增加,额外开销和理论模型几乎一致. Bennett 算法在可逆计算的意义下是最优的时空 交换策略,但对普通硬件并不是最优的. Treeverse+ NiLang 的方案是指用可逆计算处理单步运算的微 分,同时用 Treeverse 算法处理步骤间的微分.这 里随着检查点数目的减少,计算时间的减少并不明 显. 这是因为增加的计算时间是前向计算, 而这里 单步后向计算梯度的时间是前向时间的 20 多倍, 因此即使仅用5个检查点,额外的增加的时间也不 到1倍.这里单步计算梯度的之所比前向计算慢很 多,是因为当使用 NiLang 微分并行运行的 GPU 核函数,必须要避免变量的共享读取以避免程序在 后向计算中同时并行更新同一块内存的梯度,这些 额外的设计带来了很多性能的损失. 与此对比, 在 单线程版本的 CPU 上, 前向和后向的单步运算时 间差距在 4 倍以内. 此外, 虽然 Treeverse算法在 可以做到高效的时间和空间的交换,但由于内存的 管理需要用到系统内存中的全局堆栈,无法直接用 于微分 GPU 的核函数. 而可逆计算则非常适合微 分这种非线性且具有一定可逆性的程序,这也是为 什么这里选择用可逆计算对单步运算求导来避免 手动求导 GPU 核函数的麻烦.



图 5 Bennett 算法和 Treeverse 算法应用于 PML 求解过 程的微分的时间与空间开销.其中图中标记的数字为函数 的实际前向运行步骤数与原始模拟步骤数 (10⁴)的比值. 纵轴的总时间是前向计算和后向传播时间的总和, 横轴的 空间的数值的实际含义为检查点的数目或可逆计算中的最 大状态数.在 Bennett 算法中, 后向运行次数和前向运行次 数一样, 而 Treeverse 算法中, 反向传播的次数固定为 10⁴

Fig. 5. The time and space cost of using Bennett algorithm and Treeverse algorithm to solve the PML equation. The annotated numbers are relative forward overhead defined as the ratio between actual forward computing steps and the original computing steps. The time in the y axis is the actual time that includes both forward computing and back propagation. The space in the x axis is defined as the maximum number of checkpoints in the checkpointing scheme or the maximum intermediate states in the reverse computing scheme. In the Bennett's algorithm, the number of forward and backward steps are the same, while in Treeverse algorithm, the number of back propagation steps is fixed to 10^4 .

感谢王磊老师的讨论以及计算设备的支持. 感谢彩云 天气 CTO 苑明理老师关于自动微分关于自动微分在天气 预报中应用的讨论. 感谢 QuEra 公司提供的研究赞助.

附录 A1 名词表

伴随变量 adjoint
核函数 kernel function
原子函数 primitive function
可逆计算 reversible computing
可逆编程 reversible programming
鹅卵石游戏 pebble game
伴随状态法 adjoint state method
自动微分 automatic differentiation (AD)
前向自动微分 forward mode AD
后向自动微分 reverse mode AD
地震学 seismic
检查点 checkpoint
拓展动力学 augmented dynamics
损失 loss

控制流 control flow 辛积分器 symplectic integrator 洛伦茨 Lorenz 龙格库塔 Runge-Kutta

附录 A2 时间与空间的交换,鹅卵石 游戏

鹅卵石游戏是一个定义在一维格子上的单人游戏.人 们最初提出这个模型是为了描述可逆计算中的时间与空间 的交换关系.游戏开始时,玩家拥有一堆鹅卵石以及一个一 维排布的*n*个格子,标记为0,1,2…*n*,并且在0号格子上有 一个预先布置的鹅卵石.游戏规则见表 A1.

表 A1 鹅卵石游戏-可逆计算版本

Гable A1.	${\rm Pebble}$	game	rules -	the	${\rm reversible}$	$\operatorname{computing}$
version.						

放置规则:	如果第 <i>i</i> 个格子上有鹅卵石,则可以从自己堆中 取一个鹅卵石放置于第 <i>i</i> +1个格子中,
回收规则:	仅当第 <i>i</i> 个格子上有鹅卵石,才可以把第 <i>i</i> +1 个格子上的鹅卵石取下放入自己的堆中,
结束条件:	第n个格子上有鹅卵石.
游戏目标:	在固定可使用鹅卵石数目为S (不包括初始鹅 卵石)的前提下,使用尽可能少的步骤数触发游 戏结束.

这里一个鹅卵石代表了一个单位的内存,而放置和取 回鹅卵石的过程分别代表了计算和反计算,因此均需要一 个步骤数,对应计算中的一个单位的运算时间.这里对应回 收规则中要求前一个格点中存在鹅卵石,对应可逆计算在 释放内存时,要求其前置状态存在以保证反计算可行.当鹅 卵石数目充足 (S≥n),我们用n个鹅卵石依次铺至终点 格子,此时时间复杂度和空间复杂度均为O(n).最少的鹅 卵石数目的玩法则需要用到可逆计算框架下时间和空间最 优交换方案 Bennett 算法. 如表 A2 中算法 (图 A1(b)) 所示, Bennett 算法将格 子均匀的分割为 $k \ge 2$ 等份, 先是像前执子k个区块使最 后一个区块的末尾存在鹅卵石, 然后从最后第k - 1个区块 开始反向执子收回中间k - 1个鹅卵石到自由堆中. 每个区 块又递归的均匀分割为k个子分块做同样的放置鹅卵石-保 留最后的鹅卵石-取回鹅卵石的操作, 直到格子无法再分割. 假设次过程的递归次数为l, 我们可以得到步骤数和鹅卵石 数为

$$T = (2k-1)^l, S = l(k-1) + 1.$$

其中, k = l满足总格子数 $n = k^l$.可以看出可逆计算的时间复杂度和原时间为多项式关系.同时可以看出k越小,

表 A2 Bennett 算法
Table A2. The Bennett's algorithm.
输入:初始状态集合 $S = \{0: s_0\}$,子分块数目 k ,分块 起点点 $i = 0$,分块长度 $L = n$
输出: 末态 S[n]
1 function bennett (S, k, i, L)
2 if $L = 1$ then
$3 S[i+1] \leftarrow 0$
$4 S[i+1] + = f_i S[i]$
5 else
$6 l \leftarrow \lceil L/k \rceil$
7 $k' \leftarrow \lceil L/l \rceil$
8 for $j = 1, 2, \cdots, k'$ do
9 bennett $(S, k, i + (j - 1)l), \min (l, L(j - 1)l))$
10 end
11 for $j = k' - 1, k' - 2, \dots 1$ do
12 bennett $(S, k, i + (j - 1)l, l)$
13 end
14 end
15 end



图 A1 (a) Treeverse 算法^[20] 示意图, 其中 $\eta(\tau, \delta) \equiv \begin{pmatrix} \tau + \delta \\ \delta \end{pmatrix} = \frac{(\tau + \delta)!}{\tau! \delta!}$; (b) Bennett 算法对应 k = 3的示意图^[33,25], 其中 P和 P^{-1} 分别代表了计算和反计算

Fig. A1. (a) An illustration of the Treeverse algorithm^[20], where $\eta(\tau, \delta) \equiv \begin{pmatrix} \tau + \delta \\ \delta \end{pmatrix} = \frac{(\tau + \delta)!}{\tau! \delta!}$. (b) An illustration of the Bennett's algorithm for $k = 3^{[33,25]}$, where *P* and *P*⁻¹ are forward computing and reverse computing respectively.

使用的总鹅卵石数目越小,因此最省空间的鹅卵石游戏解 法对应k = 2.作为例子,图 A2(b)展示了n = 16, k = 2(l = 4)时候的游戏解法,对应步骤数为(T + 1)/2 = 41, 这里的实际操作数少了大约一半,是因为最外层的 Bennett 过程不需要取回鹅卵石.



图 A2 (a) Treeverse 算法 ($\tau = 3$, $\delta = 3$) 和 (b) Bennett 算法 (k = 2, n = 4) 对应的最优时间空间交换策略下的 鹅卵石游戏解法, 横向是一维棋盘的格子, 纵向是步骤. 其 中 "0"为在这一步中收回的鹅卵石, "•"为在这一步中放 上的鹅卵石, 而颜色稍淡的 "•"则对应之前步骤中遗留在 棋盘上未收回的鹅卵石. (a) 中的红色格子代表已被涂鸦. (b) 中带旗帜的格点代表终点

Fig. A2. (a) The Treeverse algorithm ($\tau = 3$, $\delta = 3$) and (b) Bennett's algorithm (k = 2, n = 4) solutions to the pebble game, the *x* direction is the grid layout and the *y* direction is the number of steps. Here, a " \circ " represents a pebble returned to the free pool in current step, a " \bullet " represents the pebble added to the board in current step, and a " \bullet " represents a pebbles left on the board. In (a), red grids are painted grids. In (b), the grid with a flag sign is the goal. 我们稍微修改游戏设定可以得到检查点版本的鹅卵石 游戏.该版本中用户多了一支画笔可用于涂鸦格点,且初始 状态的*n*号格子已经被涂鸦,新的规则描述见表 A3.

检查点版本的鹅卵石游戏中,鹅卵石可以被随时取下, 代表不可逆的内存擦除.涂鸦过程则代表了梯度反向传播 的过程,它要求前置计算状态和后置梯度都存在.在鹅卵石 充足的情况下,最节省步骤数的解法和可逆计算版本一样, 即计算过程中不取下任何鹅卵石.而用最少鹅卵石的解法 则仅需要两枚鹅卵石交替使用,每当我们需要涂鸦一个格 子i,我们总是从初始鹅卵石开始扫描(依次放置一个鹅卵 石并取下前一个鹅卵石)i步至指定格子,因此总步骤 $k \leq n(n-1)/2$. 鹅卵石数目为2< $\delta < n$ 的最优解最 难, 需要用到如表 A4 中算法 (图 A1(a)) 所示的 Treeverse 算法. 该算法在完成第一遍从格子0开始的扫描会在格子上留 下 $\delta - 1 = S - 2$ 个鹅卵石 (不包括初始鹅卵石), 把格子 分割成δ个区块. 我们把这些没有被取下的鹅卵石称为检 查点,总可以从任意一个检查点出发扫描后方的格子.区块 的大小由二项分布函数 $\eta(\tau-1,\delta),\cdots,\eta(\tau-1,2),$ $\eta(\tau - 1, 1)$ 确定,其中 τ 的取值满足 $\eta(\tau, \delta) = n$.为了涂 鸦n-1号格点,我们从离n-1号格点最近的检查点 (最 后一个区块的开始处) 出发扫描至该点, 重复该过程直至涂 鸦至最后一个区块的开始处.由于最后一个区块尺寸最小, 仅为τ-1,因此我们并不担心这样的扫描会使得步骤数增 加太多. 当完成了最后一个区块的涂鸦, 便可把格子上用于 标记最后一个区块起点的鹅卵石取下以便重复利用. 为了 涂鸦倒数第二个区块,先扫描这个区块,并用刚回收的鹅卵 石将其分割为大小分别为 $\eta(\tau - 2, 2)$ 和 $\eta(\tau - 2, 1)$ 的两 个子区间.用同样的方式计算最后一个区间并递归分割前 一个子区间直至区块大小为1而无法继续分割. 整个算法 的步骤数和鹅卵石数目的关系是

$T \approx \tau n, \ S = (\delta + 1),$

由二项分布的性质可得, $\tau 和 \delta$ 的大小可以都正比于 log(n). 图 A2(a) 展示了 Treeverse 算法仅用 4 个鹅卵石, 46 个步骤数涂鸦完所有 20 个格子. ^①

	表 A3	鹅卵石游戏-检查点版本
Table A3.	Pebble	game rules—the checkpointing version.

放置规则:如果第i个格子上有鹅卵石,则可以从自己堆中取一个鹅卵石放置于第i+1个格子中.
回收规则:可以随意把格子上的鹅卵石取下放入自己的堆中,收回鹅卵石不计步骤数.
涂鸦规则:当第i个格子有鹅卵石,且第i+1个格子被涂鸦,可以涂鸦第i个格子,涂鸦不记入步骤数.
结束条件: 涂鸦完所有的格点.
游戏目标:在固定可使用鹅卵石数目为S (不包括初始鹅卵石)的前提下,使用尽可能少的步骤数触发游戏结束.

① 这里的 46 步并不是严格的最优解,因为图中扫描过程的最后一步并不需要立即释放内存从而可以减少步骤数.

算法 A2: Treeverse 算法 表 A4 Table A4. Algorithm A2: The Treeverse algorithm. 输入: 状态缓存集合 $S = \{0: s_0\}$, 需回传的梯度 $\overline{s_n} \equiv$,允许缓存的状态数 δ ,扫描次数 τ ,分块起点 $\beta = 0$, 分开终点 $\phi = n$ 内容以及把分块分割为两部分的分割点 $\sigma = 0$ 输出: 回传的梯度 $\overline{s_0} \equiv \frac{\partial \mathcal{L}}{\partial s_0}$ 1 function treeverse($S, \overline{s_{\phi}}, \delta, \tau, \beta, \sigma, \phi$) 2 if $\sigma > \beta$ then 3 $\delta = \delta - 1$ $\mathbf{4}$ $s = S[\beta]$ #加载初始状态 s_β 5 for $j = \beta, \beta + 1, \dots, \sigma - 1$ do 6 $s_{i+1} = f_i(s_i)$ #计算 s_σ 7end 8 $S[\sigma] = s_{\sigma}$ 9 \mathbf{end} 10#以κ为最优分割点(二项分布),递归调用 Treeverse算法 while $\tau > 0$ and $\kappa = \operatorname{mid}(\delta, \tau, \sigma, \phi) < \phi$ do 1112 $\overline{s_{\kappa}} = \text{treeverse}(S, \overline{s_{\phi}}, \delta, \tau, \sigma, \kappa, \phi)$ $\tau = \tau - 1$ 1314 $\phi = \kappa$ 15end $\overline{s_{\sigma}} = \overline{f_{\sigma}}(\overline{s_{\sigma+1}}, s_{\sigma})$ 16 #利用已有的 s_{σ} 和 $\overline{s_{\sigma}}$ 回 传导数 if $\sigma > \beta$ then 17 remove($S[\sigma]$) # 从缓存的状态集合中移除 s_{σ} 18 19 end 20return $\overline{s_{\sigma}}$ 21 end 22 function mid(δ , τ , σ , ϕ) #选取二项分布分割点 $\kappa = \left[(\delta \sigma + \tau \phi) / (\tau + \delta) \right]$ 23 24if $\kappa \geq \phi$ and $\delta > 0$ then $\kappa = \max(\sigma + 1, \phi - 1)$ 2526end 27 end

鹅卵石游戏是对程序的时间和空间的交换关系的非常 理想化的描述,它恰巧非常合适用于描述常微分方程求解 这样的不可逆线性程序.图 A3展示了在固定额外时间开 销下,程序应用 Bennett 算法和 Treeverse 算法得到的最 优的空间开销.可以看出,可逆计算整体上需要更多的空间 开销.尤其是当步骤数更多,或是允许的时间的额外开销更 大时,该差别愈加明显.当程序具有一定结构,可逆计算也 有不错的优点,例如可以利用可逆性节省内存.另外由于可 逆计算不需要对程序自动设置检查点,因此不需要借助全 局堆栈,对于微分运行在 GPU 设备上的核函数,避免全局 堆栈操作是必要的.



图 A3 为了回溯中间状态,时间和空间在两种最优时间-空间交换策略下的关系 (a)固定横轴为状态回溯的计算 时间与原函数计算时间的比值,对比再允许固定时间开销 下,内存的额外开销.其中 Bennett 算法代表了可逆计算下 的最优策略,而 Treeverse 则是传统计算允许的最优策略. (b)对比 Bennett 算法与 Treeverse 算法空间开销的比值

Fig. A3. The time and space to trace back states in two time-space tradeoff strategies. (a) Space overheads versus time overheads at different numbers of simulation steps. The Bennett's algorithm is the optimal time-space tradeoff scheme in reversible computing, while the Treeverse algorithm is the optimal time-space tradeoff scheme in regular irreversible computing. (b) The ratio between the space overheads in the Bennett's algorithm and that in the Treeverse algorithm.

参考文献

- Griewank A, Walther A 2008 SIAM, Evaluating derivatives: principles and techniques of algorithmic differentiation
- Rosset C 2019 Microsoft Blog, Turing-nlg: A 17-billionparameter language model by microsoft
- [3] Nolan J F 1953 Ph. D Dissertation (Cambridge: Massachusetts Institute of Technology)
- [4] Wengert R E 1964 Communications of the ACM 7 463
- [5] Linnainmaa S 1976 BIT Numerical Mathematics 16 146
- [6] Gutzwiller M C 1963 Phys. Rev. Lett. 10 159
- [7] Carleo G and Troyer M 2017 Science 355 602
- [8] Deng D L, Li X P, Sarma S D 2017 Phys. Rev. X 7 021021
- [9] Cai Z, Liu J G 2018 Phys. Rev. B 97 035116

- [10] Luo X Z, Liu J G, Zhang P, Wang L 2020 Quantum 4 341
- [11] Liao H J, Liu J G, Wang L, Xiang T 2019 Phys. Rev. X 9 031041
- [12] Liu J G, Wang L, Zhang P 2021 Phys. Rev. Lett. 126 090506
- [13] Heimbach P, Hill C, Giering R 2005 Future Generation Computer Systems 21 1356
- [14] Symes W W 2007 Geographics 72 SM213
- [15] Zhu W Q, Xu K L, Darve E, Beroza G C 2021 Computers & Geosciences 151
- [16] Bradbury J, Frostig R, Hawkins P, Johnson M J, Leary C, Maclaurin D, Necula G, Paszke A, VanderPlas J, Wanderman-Milne S, Zhang Q 2018 Software, JAX: Composable Transformations of Python+NumPy Programs
- [17] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z M, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J J, Chintala S 2019 arXiv: 1912.01703
- [18] Plessix R 2006 Geophysical Journal International 167 495
- [19] Chen R T, Rubanova Y, Bettencourt J, Duvenaud D K 2018 Advances in Neural Information Processing Systems Palais des Congrès de Montréal, Montréal, Canada
- [20] Griewank A 1992 Optimization Methods and Software 1 35

- [21] Forget G, Campin J M, Heimbach P, Hill C N, Ponte R M, Wunsch C 2015 Geoscientific Model Development 8 3071
- [22] Hascoet L, Pascual V 2013 ACM Transactions on Mathematical Software (TOMS) 39 20
- [23] Utke J, Naumann U, Fagan M, Tallent N, Strout M, Heimbach P, Hill G, Wunsch C 2008 ACM Transactions on Mathematical Software (TOMS) 34 1
- [24] Liu J G, Zhao T 2020 arXiv: 2003.04617
- [25] Levine R Y, Sherman A T 1990 SIAM Journal on Computing 19 673
- [26] Grote M J, Sim I 2010 arXiv 1001.0319
- [27] Lorenz E N 1963 J. Atmospheric Sci. 20 130
- [28] Hirsch M W, Smale S, Devaney R L 2012 Differential Equations, Dynamical Systems, and An Introduction to Chaos (Cambridge: Academic Press)
- [29] Revels J, Lubin M, Papamarkou R 2016 arXiv 1607.07892
- [30] Berenger J P 1994 Journal of computational physics 114 185
- [31] Roden J A, Gedney S D 2000 Microwave Opt. Technol. Lett. 27 334
- [32] Martin R, Komatitsch D, Ezziani A 2008 Geophysics 73 T51
- [33] Bennett C H 1973 IBM journal of Research and Development 17 525

SPECIAL TOPIC—Machine learning and physics

Automatic differentiation and its applications in physics simulation^{*}

Liu Jin-Guo^{1)†} Xu Kai-Lai²⁾

1) (Department of Physics, Harvard University, Cambridge 02138, USA)

2) (Stanford University, Stanford 94305, USA)

(Received 27 April 2021; revised manuscript received 17 June 2021)

Abstract

Automatic differentiation is a technology to differentiate a computer program automatically. It is known to many people for its use in machine learning in recent decades. Nowadays, researchers are becoming increasingly aware of its importance in scientific computing, especially in the physics simulation. Differentiating physics simulation can help us solve many important issues in chaos theory, electromagnetism, seismic and oceanographic. Meanwhile, it is also challenging because these applications often require a lot of computing time and space. This paper will review several automatic differentiation strategies for physics simulation, and compare their pros and cons. These methods include adjoint state methods, forward mode automatic differentiation, reverse mode automatic differentiation, and reversible programming automatic differentiation.

Keywords: automatic differentiation, scientific computing, reversible programming, optimal checkpointing, physics simulation

PACS: 02.60.Pn, 02.30.Jr, 91.30.-f

DOI: 10.7498/aps.70.20210813

 $[\]ast~$ Project suppored by the Research Program of QuEra Computing Inc.

[†] Corresponding author. E-mail: jinguoliu@g.harvard.edu