



机器学习在原子核物理中的应用

何万兵^{1*}, 何俊杰^{2,3}, 王睿², 马余刚^{1*}

1. 复旦大学现代物理研究所, 复旦大学核物理与离子束应用教育部重点实验室, 上海 200433;

2. 中国科学院上海应用物理研究所, 上海 201800;

3. 中国科学院大学, 北京 100049

*联系人, 何万兵, E-mail: hewanbing@fudan.edu.cn; 马余刚, E-mail: mayugang@fudan.edu.cn

收稿日期: 2021-11-06; 接受日期: 2021-12-27; 网络出版日期: 2022-04-20

国家自然科学基金(编号: 11890710, 11890714)、中国科学院战略性先导B类专项(编号: XDB34000000)、科技部国家重点研发计划(编号: 2016YFE0100900, 2018YFE0104600, 2020YFE0202001)和广东省基础与应用基础研究重大项目(编号: 2020B0301030008)资助

摘要 机器学习是人工智能研究领域最成功最活跃的研究方向,正在被广泛应用于大数据分析、自动驾驶、自然语言处理、互联网技术、医学诊断、市场分析等领域.同时也被广泛应用于粒子物理的实验和理论研究.本文针对核物理领域列举了一些值得被关注的机器学习技术.通过几个不同案例,介绍了机器学习在核物理研究的不同场景的应用.其中包括深度卷积神经网络结合理论模型用于提取原始数据中的团簇结构信息,无监督学习方法应用于实验数据分析提取原子核液气相变信息,用神经网络量子态(NQS)作为试探波函数求解薛定谔方程,贝叶斯神经网络应用于原子核半径的拟合与预测,图像识别技术应用于活性靶时间投影室探测器事件鉴别.展示了机器学习技术在核物理领域的广泛的应用价值.

关键词 机器学习, 原子核物理, 原子核团簇结构, 原子核液气相变

PACS: 21.10.Ft, 25.70.Pq, 29.40.Gx, 29.85.+c, 21.60.Kn

1 引言

机器学习是人工智能研究的子集,也是人工智能研究中最活跃的领域.作为计算科学和人工智能的分支,机器学习的起源可以追溯到20世纪40年代.早在电子计算机的发明之初,人们就对人工智能的实现充满了渴望,神经生理学家McCulloch和数学家Pitts^[1]合作提出了神经元模型,被认为是人工神经网络的开端.心理学家Hebb在自己的著作中提出了突触可塑性假说,为人工神经网络提供了可行的训练方案.

Turing用一个简单设计的图灵测试定义了人工智能.机器学习的概念是1959年由Samuel提出,并给出了具体的定义“不需要显式编程,使计算机具有学习能力的研究领域”.20世纪60年代,统计学方法开始被引入到机器学习领域,人们尝试构建机器学习的理论框架.直到1984年,Valiant提出了概率近似正确(Probably Approximately Correct, PAC)模型,开创性地将学习任务的难度用数学模型进行了量化,成为构建机器学习理论框架的转折点. Vapnik和Chervonenkis等人提出Vapnik-Chervonenkis理论,将学习算法的能力进行

引用格式: 何万兵, 何俊杰, 王睿, 等. 机器学习在原子核物理中的应用. 中国科学: 物理学 力学 天文学, 2022, 52: 252004

He W B, He J J, Wang R, et al. Machine learning applications in nuclear physics (in Chinese). Sci Sin-Phys Mech Astron, 2022, 52: 252004, doi: 10.1360/SSPMA-2021-0309

了量化, 完善了统计学习的理论框架. 统计学习理论目前也是机器学习的唯一理论框架, 以至于很多时候人们将机器学习与统计学习画上了等号. 伴随统计学习理论的建立, 有着严谨统计学习理论基础的支持向量机模型取得了巨大的成功, 成为90年代最为成功的机器学习模型. 21世纪开始, LeCun, Hinton和Bengio等人通过反向传播算法、卷积算法等重塑了深度神经网络. 深度神经网络表现出惊人的性能, 扭转了机器学习的格局, 同时对如何将深度神经网络纳入统计学习理论框架提出了挑战. 机器学习理论的发展与数学和统计物理学有着深刻的联系, 机器学习的应用不断被拓展到物理学研究中.

2 机器学习的基本概念

机器学习发展到现在形成了一个庞大的概念体系, 本文尝试简单梳理一下机器学习的一些基本概念.

如图 1所示, 可以把机器学习的概念分解为4个部分: 数据、模型、学习策略和任务. 数据是机器学习要分析和学习的对象, 常见的数据格式包括图片、文本语言、声音等. 模型是通过一定算法构建的具有可调参数的数学模型, 是机器学习的机器部分. 有了机器部分接下来就是训练机器, 也就是学习的概念, 通过不同的学习策略解决如何更新模型参数的问题. 任务就是要用训练好的机器解决什么样的问题. 根据这4个不同部分可以把机器学习的一些概念进行简单的分类归纳.

2.1 数据类型

在物理学研究领域, 机器学习常用的输入数据格

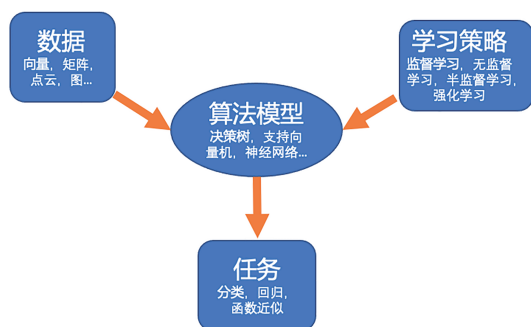


图 1 (网络版彩图)机器学习的基本概念

Figure 1 (Color online) Basic concepts in machine learning.

式包括多维空间矢量、二维或三维图像、点云和拓扑图等. 同样的数据内容可以根据具体情况表示为不同的数据结构. LHC对撞机实验中普遍使用了机器学习方法识别来自不同探测器的信号可以进行粒子的鉴别^[2], 每一个信号通道可以表示一个矢量维度, 不同通道的信号共同构成一个多维空间矢量作为原始数据. 利用粒子在电磁量能器中沉积的能量可以构建出不同快度区间的二维能量分布, 用机器学习中的图像识别技术来研究这种类似图像格式的二维谱数据是实验数据分析的常见用法, 通常用来做事件重建^[3]、事件堆积处理^[4]、目标检测^[5]、喷注研究^[6,7]等.

尽管通过填谱的方法很容易将探测器测得的原始数据构建二维或三维图像, 并应用于神经网络的输入. 但是这种基于图像的分析策略需要输入数据具有固定的大小和形状, 因此限制了数据的空间分辨率, 在表示粒子多重数具有很大涨落的事件时, 不得不引入归一化参数. 将数据表示为点云能够弥补这种缺点, 点云数据及其相关的点云网络能够更好地表示离散分布的数据结构. 机器学习领域针对点云数据开发了多种类型的神经网络模型, 可以统称为点云网络^[8,9].

图像和点云数据都是欧几里德空间的数据表示, 而实验数据中的高维特征通常具有不同的拓扑结构, 这种拓扑结构包含有丰富的物理信息, 包含模式识别所需要的重要信息. 如果数据具有典型的树状拓扑结构, 可以使用递归神经网络进行研究. 而对于更一般的拓扑结构通常使用图神经网络作为工具. 图神经网络通过图嵌入、关系矩阵等表示数据的拓扑结构, 能够将拓扑结构信息表示为可计算的矩阵数据^[10,11].

有一种特殊的机器学习任务不需要数据输入, 比如微分方程求解、量子态的神经网络表示^[12]等. 量子态的神经网络表示通过构建神经网络作为薛定谔方程的试探波函数, 根据变分算法迭代更新网络参数, 使其逼近真实的波函数形式. 其输入数据来自于高维空间的蒙特卡罗采样算法.

2.2 算法模型

机器学习的算法和模型具有多种不同的形式, 比如 k 均值、主成分分析、决策树、支持向量机(SVM)、神经网络. k 均值聚类算法和主成分分析方法都是无监督学习算法, 通常被用来对数据进行聚类、降维等预

处理. 提升决策树算法(BDT)具有简单的模型结构, 训练和预测计算比较快速, 在实验物理中得到了广泛的应用, BDT算法在希格斯粒子的寻找过程中起到重要的作用^[13], 更有效地从大量本底数据中检索出真实信号. 支持向量机(SVM)是一个优秀的分类算法模型, 具有更严格的统计学习理论基础, 不断被应用于实验数据的分析^[14,15].

其中神经网络由基本的神经元结构相互连接构成. 根据网络拓扑结构的不同又可以分为全连接神经网络、卷积神经网络、残差神经网络、循环神经网络等. 根据隐藏层深度又可分为浅层神经网络和深度神经网络. 浅层神经网络最早的实现是感知机, 但是感知机模型相比于通过巧妙设计的非神经网络算法(如支持向量机等)无法达到同等的性能. 但是自从2006年以来, 深度神经网络, 尤其是卷积神经网络得到了极大的发展和广泛的应用, 基于卷积结构的神经网络单元构建了各种复杂的高性能深度神经网络. 在物理学研究中, 深度神经网络逐渐替代了决策树算法成为最主要的机器学习模型.

神经网络中加入具有物理意义的参数被证明能够提高模型的性能^[16]. 在喷注结构研究中, 深度卷积神经网络用于处理由粒子能量分布给出的二维图像数据, 表现出比线性分类算法LDA、决策树等算法更好的性能^[6]. 循环神经网络和长短时记忆网络是自然语言分析的强力工具, 研究表明这类神经网络在喷注研究中能够很好地学习喷注的结构特征^[17]. 点云神经网络能够有效处理实验中由被探测粒子空间分布构成的点云数据^[18,19]. 处理拓扑图结构数据的神经网络称为图神经网络^[10]. 图神经网络可以通过图嵌入技术将数据表达为低维向量, 或者通过构建描述数据拓扑结构的拉普拉斯矩阵将拓扑结构信息表示为数值矩阵.

2.3 学习策略

根据学习策略的不同, 机器学习可以分为监督学习、无监督学习、弱监督学习和强化学习. 其中监督学习需要数据的标签, 也就是算法模型需要显式输入数据所对应的正确答案. 监督学习算法是实验物理中最主要的学习算法. 理论模型的计算是解析实验数据的主要工具, 理论模型计算出的样本数据对于机器学习是有标签的数据. 模拟样本和理论标签是构建监督

学习算法的必要因素, 因此监督学习算法广泛应用于各种实验数据分析当中^[20-22], 尤其是径迹重建、触发判选、稀有事例寻找等.

相反, 无监督学习不需要显式输入数据的标签, 而是从数据本身去归纳和分析, 实现所设计的特定任务. 无监督学习算法常见于数据的预处理, 如聚类^[23]和降维. 也常见用于隐变量提取^[24]和新物理寻找^[25,26].

还有一种概念称为半监督或者弱监督学习, 是指只有部分数据包含可用标签, 其余数据可能无标签, 也可能有错误标签^[27]. 或者数据集的单个样本没有标签, 但是数据集的某些特征是已知的, 也可以构建弱监督学习算法, 比如数据集的不同样本具有已知的比例, 通过这个比例信息构建的损失函数, 仍然能够使辨别模型得到收敛, 并得到与监督学习算法相当的精度水平^[28,29].

强化学习所要面对的数据样本是动态的, 甚至是不可穷举的, 这样的数据空间称为环境. 而且算法模型每运行一次, 环境会被改变, 也就是机器的行为会与环境进行交互, 我们所熟知的Alpha Go就是强化学习的实例. 在学习阶段, 机器的每次数据输入都会给出一次输出, 这次输出会改变环境, 但是并不会立刻收到对错或偏差的判断. 是否进行对错或偏差判断, 往往取决于环境的状态. 如果某种预设状态被触发, 就会进行一次对错或偏差的判断, 然后更新参数. 所以强化学习是明显区别于有监督和无监督学习的. 但是如果环境状态是可以被穷举的, 强化学习就可以退化为有监督学习. 强化学习算法通过迭代学习得到一个隐藏的状态与关联矩阵, 通常称为Q-Learning. Q值可以被扩展为神经网络, 即DQN模型. 强化学习在自动控制领域有很多的应用, 比如粒子加速器系统的智能控制^[30].

2.4 任务

由于机器学习的应用场景非常宽泛, 常被应用于不同领域不同类型的任务, 其中在物理学研究领域常被应用的场景主要包括数据预处理、分类、回归、函数近似等.

对样本数据进行无监督的聚类、降维、嵌入等是常见的数据预处理任务. 通过聚类算法能够将样本进行分类^[23], 对于数据的表示, 新物理的发现具有应用价值. 主成分分析是降维的主要手段, 通常将探测器

得到的高维数据表示为低维空间的向量^[31], 为进一步的数据分析提供更好的接口. 数据嵌入在处理拓扑结构数据时有重要的应用价值, 通过数据嵌入技术将拓扑结构信息映射到低维向量空间, 常被用于神经网络^[32].

机器学习通过模式判别能力对数据样本实现分类任务. 实验物理, 尤其是高能实验物理中常常会产生大量的数据, 从中筛选出有效的事例正是机器学习分类算法所擅长的. LHC对撞机上的CMS实验大量使用了BDT分类算法进行事例筛选. BDT算法通过学习大量的模拟数据, 使得其比传统方法对信号的灵敏度提高了大约23%, 这相当于数据量增加了50%^[33]. ATLAS实验同样大量使用BDT算法进行事例筛选, 对比传统算法, 灵敏度提高了50%^[34]. 新的实验分析开始使用深度神经网络, 在同样的分类任务中表现出了更好的性能. 深度神经网络在数据分析中可以直接使用更原始的数据, 并有效利用原始数据中的拓扑特征构建事件的分类标签^[3]. 现代高亮度对撞机实验在提高亮度的同时也造成了严重的事件堆积现象. 近年来卷积神经网络技术逐渐替代了传统的区域扣除等方法, 成为了事件堆积剔除的有效工具^[4].

机器学习应用于回归分析, 主要任务是得到数据的某个连续变量. 通过机器学习能够更好地刻度探测器响应规律^[35], 电磁量能器中更精确的能量回归计算能够帮助测量高能缪子的能量动量信息. 探测器的刻度方法通常是拟合一个固定的比例系数, 而实际的探测器响应会随着事件的不同特征而存在变化. 神经网络算法能够给出更准确的探测器刻度数据, 从而提高探测效率和降低统计误差. 逐事件的喷注介质修正是相对论重离子核反应实验的重要任务, 机器学习的回归分析算法能够更准确地估计喷注在介质中的能量损失. 卷积神经网络被用来约束宇宙学参数, 并给出了更加精确的结果^[36].

用机器学习算法近似未知函数包含了多种任务场景. 高能核物理与粒子物理的研究需要大量的模拟数据, 这种需求对计算资源提出了很高的要求, 通过机器学习方法解决快速数据模拟是个重要的研究方向^[37]. 生成对抗模型被广泛应用于数据生成模型研究^[38–40], 生成模型的本质是通过对抗训练使神经网络模型学习到了一个近似的样本分布函数. 标准化流模型能够通过一系列可逆算法生成任意形式的复杂分布函数, 研

究表明基于标准化流的模型适用于构建生成模型^[41].

用理论模型解释观测数据是实验数据处理分析过程中的重要环节, 理论模型的参数往往需要通过实验数据的统计推断来确定. 对于许多理论模型在对数据进行生成和模拟的过程中无法对假设参数给出解析的似然函数, 这给进一步的统计分析带来了困难. 通过机器学习方法构建的广义似然比检验统计可以实现无似然推断.

3 机器学习在核物理中的应用举例

分别以几个简单的机器学习实例详细介绍几种不同的机器学习方法或应用场景, 其中涵盖了监督学习和无监督学习, 以及从模型中提取观测量和直接将模型应用于实验数据分析, 还有机器学习应用于量子理论的研究, 以及考虑数据误差的贝叶斯模型, 最后介绍了将机器学习应用于一种中低能核物理实验特别重要的探测器——活性靶时间投影室.

3.1 基于机器学习方法鉴别相对论重离子碰撞实验中实验初始原子核团簇结构

在中高能核物理领域的两个重要的对撞机LHC和RHIC, 除p-p系统以外, 目前所运行的重离子束流以Pb-Pb和Au-Au等系统为主. 在重离子对撞的数据分析过程中, 所有的理论模型都将Pb和Au等大质量原子核近似为球形或椭球形的结构. 而对于¹²C和¹⁶O等轻核系统, 在原子核结构的理论计算表明¹²C和¹⁶O可能存在多种构型的 α 团簇结构, 也就是说不能把轻质量原子核当作简单的球形结构去构建模型. 在RHIC能区和LHC能区研究¹²C和¹⁶O等轻核的对撞数据, 必须认真考虑 α 团簇结构的影响. 采用机器学习的方法从末态带电粒子的横动量谱中尝试识别 α 团簇结构的影响, 论证 α 团簇结构是否对相对论能区的原子核对撞实验具有可观测的影响.

本文采用了一种混合式的贝叶斯卷积神经网络, 具体的网络结构如表1所示.

模型具有典型的深度神经网络结构, 采用了4个卷积层, 每个卷积层的卷积核大小为 3×3 结构, 数量呈递增结构, 与常规的卷积神经网络不同, 这里的最后三个全连接层采用了贝叶斯模型. 深度卷积神经网络是目前图像识别领域性能最好的模型, 我们构建的输入数

据正是将粒子的横动量分布填谱生成二维图像, 全连接层采用贝叶斯模型可以利用贝叶斯神经网络良好的泛化能力, 很好地防止过拟合现象的发生.

样本数据由AMPT模型模拟产生, AMPT模型是被广泛使用的相对论能区核反应模拟工具. 首先在AMPT模型中构建了初始状态具有特定 α 团簇结构的 ^{12}C 和 ^{16}O 原子核, 每组不同构型的 $^{12}\text{C}/^{16}\text{O}+^{197}\text{Au}$ 系统以每核子对能量200 GeV分别模拟了 8×10^5 个事件. 选择快度为-1到1、横向动量为0到2 GeV/c的带电 π 介子. 如图2所示, 其中横坐标为粒子发射方位角, 纵坐标为横动量, 图片尺寸为 64×64 像素. 为了减小事件的涨落, 每个图片由4000个事件叠加产生.

图3为不同数量的事件叠加对神经网络学习精确度影响, 横坐标是数据集训练轮数, 纵坐标为模型验证准确度, 我们发现4000个事件的叠加, 使 ^{12}C 和 ^{16}O 可以给出的精确度分别达到95%和99%. 这个结果表明 α 团簇结构能够对相对论能区产生显著的可观测效应, 对训练好的模型进行解析有助于理解神经网络模型捕捉到了什么样的特征来识别 α 团簇结构的影响.

考虑到原子核的基态可能同时具有 α 团簇结构和类液滴结构特征, 需要验证在一个图像中按一定比例混合 α 团簇结构和Woods-Saxon分布事件后, 神经网络模型是否能够很好量化这种混合比例. 这里

表1 贝叶斯卷积神经网络结构

Table 1 Architecture of Bayesian convolutional neural network

顺序	神经网络层类型	网络层参数
0	输入层	64×64
1	卷积层	6个卷积核
2	池化层	池化单元 2×2
3	卷积层	16个卷积核
4	池化层	池化单元 2×2
5	卷积层	32个卷积核
6	池化层	池化单元 2×2
7	卷积层	64个卷积核
8	池化层	池化单元 2×2
9	Flatten	1024节点
10	贝叶斯全连接层	256节点
11	贝叶斯全连接层	64节点
12	贝叶斯全连接层+Softmax	2节点输出

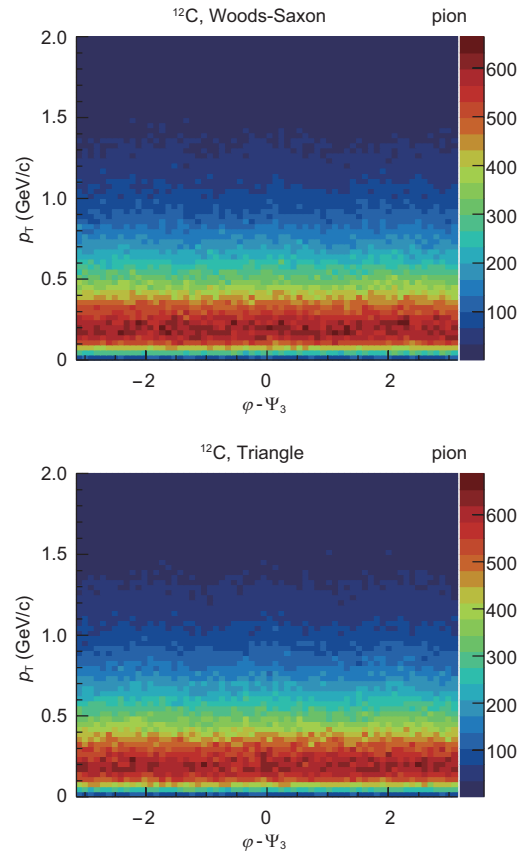


图2 (网络版彩图) ^{12}C 初始状态核子的末态横动量谱. 横坐标为带电 π 介子相对事件平面的发射角度, 纵坐标为横动量. 上图为 ^{12}C 初始状态核子满足Woods-Saxon分布情况时的末态横动量谱, 下图是 ^{12}C 初始状态核子具有 3α 结构时得到的末态横动量谱^[42]

Figure 2 (Color online) The final transverse momentum spectrum of the initial state nucleon. The horizontal axis is the azimuthal angle of emitted charged pions, and the vertical axis is the transverse momentum. The upper panel is for Woods-Saxon distributed ^{12}C and lower panel is for clustered ^{12}C [42].

训练神经网络来处理这个回归问题. 对于回归任务, BCNN的结构与之前的相同, 只是Softmax层和交叉熵损失分别被Sigmoid函数和均方误差代替. 将Woods-Saxon分布结构的比例作为图像的标签, 我们准备了6个不同混合比例作为回归问题的标签, 分别为0%, 20%, 40%, 60%, 80%和100%. 得到的结果如图4所示, 以 $N_{\text{Event}}=4000$ 的 ^{12}C 为例, 显示了验证集的预测结果. 虽然峰值与真实比例有关, 但预测的涨落非常大. 从相对论对撞事件的发射粒子信息中, 机器学习可以一定程度地量化 α 团簇结构, 但是回归分析的性能远不如分类问题精确.

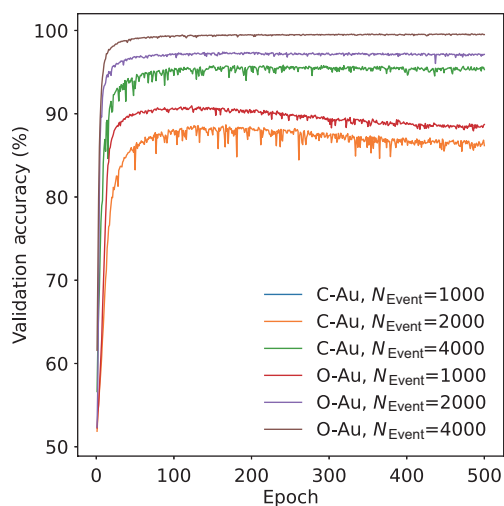


图 3 (网络版彩图)不同数量事件叠加训练得到的模型预测精确度 [42]

Figure 3 (Color online) Evolution of the validation accuracy with different amount of mixed events during the training process [42].

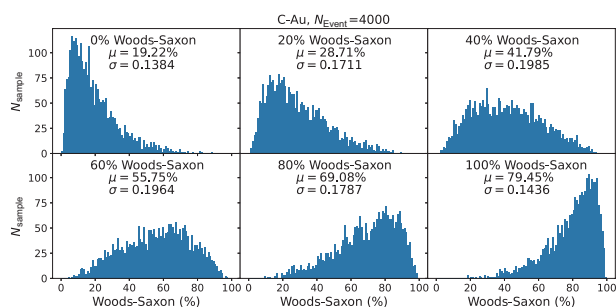


图 4 (网络版彩图)回归训练得到的模型用于给出不同初态混合比例的预测, 每幅图分别对应不同的初态混合比例 [42]

Figure 4 (Color online) Histograms of predictions on the validation set from a regression model for ^{12}C [42].

3.2 基于无监督学习的原子核液气相变研究

在核物理中的一个重要任务是研究夸克胶子等离子体(QGP)的性质和相变过程, 近年来在实验和理论研究中都取得了大量的成果 [43]. 机器学习方法也被证明可以应用于QCD相变研究 [44]. 在中低能区, 重离子反应实验是激发原子核发生液气相变的实验手段. 处于基态的原子核, 近似为液体状态, 核子密度在饱和密度 ρ_0 附近, 约为 0.16 核子/ fm^3 . 在炮弹原子核被加速到每核子几十个MeV的能量后, 与靶核发生碰撞而被激发, 此时弹靶核可视为加热液体状态. 激发能量较低时, 压缩的原子核将膨胀, 然后表现出阻尼单极振荡,

伴随着一些轻粒子的发射, 并保持液态. 随着激发能的增加, 原子核的膨胀变得更为剧烈, 并将原子核推向旋节区. 在旋节区, 由于核子-核子相互作用中的吸引部分, 高密度区将吸引低密度区的核子. 这导致中间质量碎片的形成. 如果激发能继续增加, 被激发的原子核将迅速膨胀, 足以穿过旋节区. 中等质量碎片的形成变得不那么重要, 而以轻碎片为主, 这对应于气相. 因此, 中等质量碎片或碎片多重数分布是提取原子核液气相变的关键信息.

将机器学习方法直接用于重离子反应的实验数据分析, 尝试提取与原子核液气相变相关的物理量. 实验数据是TAMU K500超导回旋加速器装置产生的能量 47 MeV/u的 ^{40}Ar 与 ^{27}Al 和 ^{48}Ti 的反应. 用 4π 探测器记录了带电碎片的电荷与动量信息.

不使用理论模型给出的计算作为标签而直接把机器学习方法应用于实验数据分析通常只能选择无监督学习训练策略. 这里我们首先用神经网络构建了一个自编码器, 结构如图 5 所示, 自编码器分为两个部分, 输入数据端连接的是编码器部分, 输出数据端连接的是解码器部分. 编码器和解码器采用对称结构, 分别具有两个隐藏层, 每层由多个神经元组成, 层与层之间采用全连接模式. 其中编码器将输入数据编码为一个隐变量, 解码器再将隐变量还原为原始数据格式.

其中输入的数据是通过类弹Quasi-Projectile (QP) 碎片逐事件构建的带电粒子多重数谱 $ZM_c(Z)$, 选取了40081个有效事件, 其中把2/3的数据作为模型的训练样本, 1/3作为测试样本. 其中把测试样本的平均多重数填谱可以得到图 6. 图中的虚线谱为实验数据, 可以比较直观地看出不同激发能量条件下的带电粒子多重数分布的差异, 圆圈为解码器通过隐变量还原后的结果, 与实验值吻合较好.

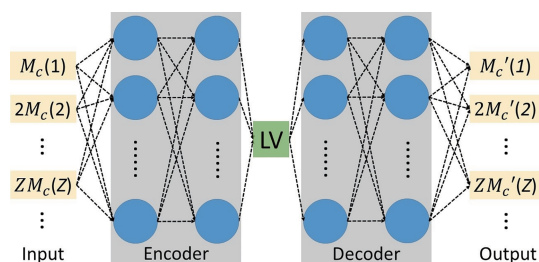


图 5 (网络版彩图)神经网络自编码器模型

Figure 5 (Color online) The construction of the autoencoder network.

解码器能够有效还原原始数据, 说明训练后的编码器给出了比较收敛的隐变量, 并且该隐变量很好地量化了不同能量下的带电碎片多重数分布的变化情况. 可以通过隐变量与温度和激发能的相关性尝试解读隐变量的物理含义, 如图 7 所示.

其中图 7 中 T_{ap} 由式(1)给出.

$$\langle T_{\text{ap}} \rangle = \sqrt{\frac{\langle Q_{xy}^2 \rangle - \langle Q_{xy} \rangle^2}{4m^2}}. \quad (1)$$

E_{ex} 由式(2)给出.

$$E_{\text{ex}} = \sum_{i=1}^{M_{\text{QP}}} E_i^{\text{kin}} + \frac{3}{2} M_n T - Q. \quad (2)$$

两幅图都表现出了 Sigmoid 函数的特征, 说明编码器模型将低温低激发能区和高温高激发能区识别为两种不同的相, 而中间的过渡区可以理解为两种相的混合态. 图中的误差来自于逐事件的涨落, 虽然涨落较大, 但是

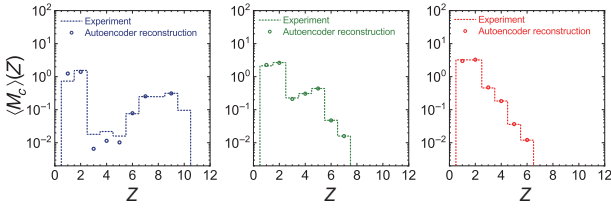


图 6 (网络版彩图)带电碎片的平均分布谱. (左) 低激发能(0.9–2.8 MeV); (中) 中等激发能(5.3–5.4 MeV); (右) 高激发能(8.1–13.0 MeV) [45]

Figure 6 (Color online) The averaged charge multiplicity distribution $\langle M_c(Z) \rangle$. Panel (left) for low excitation (0.9–2.8 MeV), panel (middle) for intermediate excitation (5.3–5.4 MeV), and panel (right) for high excitation (8.1–13.0 MeV) [45].

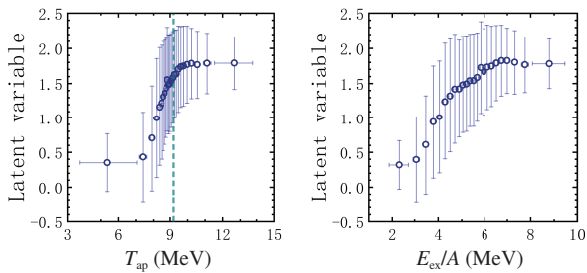


图 7 (网络版彩图)编码器网络提取的隐变量随着温度和激发能的变化关系 [45]

Figure 7 (Color online) The mean and standard deviation of the latent variable in different T_{ap} and E_{ex}/A bins, respectively [45].

变化规律也比较显著.

采用一种假设标签方法使神经网络的监督学习训练策略实现了对实验数据的直接分析, 这等价于无监督学习. 我们构建了一个简单的贝叶斯神经网络. 暂时忽略贝叶斯神经网络的细节, 而只关注训练策略. 这个网络由两层隐藏层组成, 输入数据和自编码器相同, 来自带电粒子多重数分布, 而输出为两维矢量, 其中一个分量表示液相, 另一个分量为气相. 以试探的方式去选择相变的临界温度 T_{ap} , 然后根据假设的 T_{ap} 值去标注逐事件的气相或者液相标签. 进而用这种假设标签去训练模型, 根据最终收敛到的结果, 可以画出模型的准确率随着试探临界温度 T_{ap} 和激发能的变化关系, 如图 8 所示.

横坐标分别为试探临界温度和激发能, 纵坐标为神经网络模型对假设相标签的识别准确度, 这里的误差无法根据逐事件的涨落显示, 所以选取贝叶斯神经网络来给出当前训练模型的误差大小. 以上结果中, 自然地可以将神经网络模型的准确度与液气相变的趋势关联起来, 也就是把性能的突变看作是相变临界点. 这种机器学习方法最后拟合给出的临界温度为 (9.24 ± 0.04) MeV, 与量热曲线法给出的 (9.0 ± 0.4) MeV 一致.

3.3 神经网络量子态

神经网络量子态(NQS)研究逐渐成为了一个重要的研究方向, NQS用人工神经网络(ANN)表示量子系统的量子态或波函数, 通过一定的学习策略调整神经网络的参数, 使得神经网络能够近似描述量子系统的

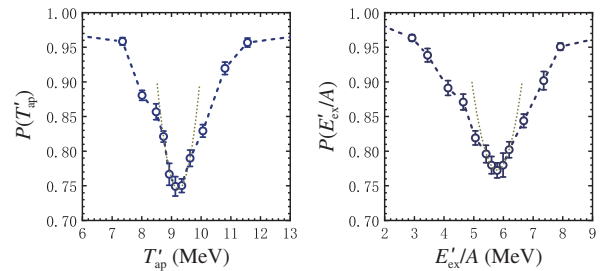


图 8 (网络版彩图)模型预测精确度 $P(T'_{\text{ap}})$ 和 $P(E'_{\text{ex}}/A)$ 随温度和激发能的变化趋势 [45]

Figure 8 (Color online) The performance curve $P(T'_{\text{ap}})$ and $P(E'_{\text{ex}}/A)$, i.e., the testing accuracy as a function of the proposed temperature T'_{ap} and transition excitation energy E'_{ex} , respectively [45].

状态或波函数. Carleo和Troyer展示了如何用受限玻尔兹曼机(RBM)解横场伊辛模型(Transverse-Field Ising, TFI)和反铁磁海森堡模型(Anti-Ferromagnetic Heisenberg, AFH)的基态. 研究思路是先构建一个RBM网络, 作为量子系统的试探波函数, 然后通过量子系统的哈密顿量写出用试探波函数表示的系统能量期望值, 也就是用RBM网络计算系统能量的期望值, 通过梯度下降算法最小化系统能量, 从而不断优化RBM网络参数, 系统能量收敛到最小值时RBM网络就被优化为量子系统的基态波函数, 研究表明该方法具有很好的精度.

这里展示一个通过深度神经网络计算超核系统结合能的例子, 我们构建一个具有多隐藏层的深度神经网络模型, 模型中为了计算高阶微分不能采用Relu作为隐藏层, 可以选取Tanh(x), Softplus(x), Sigmoid(x)等这些具有二阶导数的函数作为激活函数.

任一算符 \hat{A} 的泛函可以采用马尔科夫链采样方法进行计算, 在该采样算法中, 坐标空间数据点的采样分布服从波函数给出的概率分布, 即 $p(x) \propto \varphi^2(x)$, 样本集合 $X_1, \dots, X_{N_{\text{sample}}}$ 包括 N_{sample} 个样本, 则算符 \hat{A} 的泛函为

$$\begin{aligned} \langle \hat{A} \rangle &= \frac{\int \varphi(x) \hat{A} \varphi(x) dx}{\int \varphi^2(x) dx} = \int p(x) \varphi^{-1}(x) \hat{A} \varphi(x) dx \\ &\cong \frac{1}{N_{\text{sample}}} \sum_{i=1}^{N_{\text{sample}}} \varphi^{-1}(x_i) \hat{A} \varphi(x_i) \equiv \langle \varphi \hat{A} \varphi \rangle_{\text{sample}}. \end{aligned} \quad (3)$$

计算神经网络参数的梯度时, 反向传播算法可以给出每个参数的梯度表达式为(公式中的下标s意为蒙特卡罗采样算法):

$$\begin{aligned} \frac{\partial \hat{A}}{\partial p} &= \frac{\partial}{\partial p} \frac{\int \varphi(x) \hat{A} \varphi(x) dx}{\int \varphi^2(x) dx} \\ &\cong 2 \langle \varphi^{-1}(x) O_p \hat{A} \varphi(x) \rangle_s - 2 \langle \varphi^{-1}(x) \hat{A} \varphi(x) \rangle_s \langle O_p \rangle_s, \end{aligned} \quad (4)$$

其中,

$$O_p = \varphi^{-1}(x) \frac{\partial \varphi(x)}{\partial p}.$$

将模型用于计算氦核系统时可以得到系统的结合能随着计算迭代次数的收敛过程, 如图9所示,

最后收敛到-2.23 MeV, 与实验值一致.

在 Ξ 超核的计算中, 选取如下形式的唯象势形式:

$$V_{\Xi\text{-core}} = V_{\text{rep}} \exp\left(-\left(\frac{r}{\beta_{\text{rep}}}\right)^2\right) - V_{\text{att}} \exp\left(-\left(\frac{r}{\beta_{\text{att}}}\right)^2\right), \quad (5)$$

$$0 \leq r \leq \infty,$$

Ξ -中心核相互作用势可以用两项高斯势之和近似, 其中势参数如表2所示.

最后计算得到的势能结果及与超球谐展开(HHE)计算方法得到的结果^[46]如表3所示.

神经网络计算结果与其他数值计算结果相一致. 这种机器学习方法是求解量子系统的一种新的尝试, 展现出了很好的计算精度. 吸引着人们去继续探索如何将NQS应用到更多的量子力学问题求解中去.

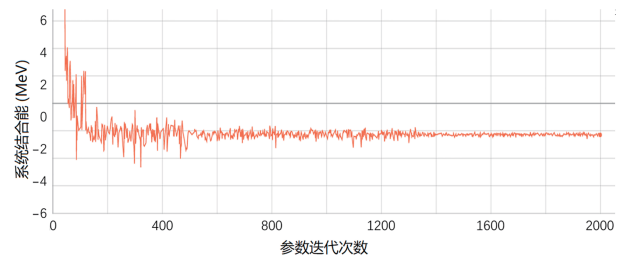


图9 (网络版彩图)神经网络量子态描述的系统结合能随迭代次数的变化

Figure 9 (Color online) Evolution of the binding energy during the training process.

表2 Ξ 超核势能参数

Table 2 Potential parameters of hyper Ξ

超核系统	V_{att} (MeV)	β_{att} (fm)
${}^5_{\Xi}\text{He}$	404.900	1.41
${}^{13}_{\Xi}\text{C}$	381.450	1.41
${}^{15}_{\Xi}\text{C}$	387.961	1.41
${}^{17}_{\Xi}\text{O}$	442.100	1.41
${}^{89}_{\Xi}\text{Sr}$	417.670	1.41

表3 Ξ 超核结合能

Table 3 Calculated binding energy of hyper Ξ

超核系统	ANN (MeV)	HHE计算 ^[46] (MeV)
${}^5_{\Xi}\text{He}$	3.30	3.3396
${}^{13}_{\Xi}\text{C}$	3.68	3.7030
${}^{15}_{\Xi}\text{C}$	4.30	4.3894
${}^{17}_{\Xi}\text{O}$	16.0	16.0320
${}^{89}_{\Xi}\text{Sr}$	21.3	21.3030

3.4 基于贝叶斯神经网络的原子核半径拟合

贝叶斯神经网络通常将传统神经网络的参数替换为某种分布, 每个神经元的偏置量和连接权重值都不是一个固定值. 如图 10 所示, 左图是传统的神经网络, 右图是贝叶斯神经网络. 贝叶斯神经网络在从输入端向输出端逐步计算时, 每个神经元的参数由该参数所服从的分布随机产生. 从训练的角度来理解, 传统神经网络的优化过程是优化每个神经元的连接权重和偏置量, 而贝叶斯神经网络是优化每个神经元连接权重和偏置量所符合的分布函数的参数, 比如选取高斯函数作为神经元参数的分布函数, 在训练过程中就需要优化每个神经元参数所对应的高斯分布的均值和标准差. 贝叶斯神经网络具有很好的泛化能力, 不容易出现过拟合, 在很小的数据集也可以准确收敛, 同时在实验数据分析中能够自然地考虑数据涨落并给出误差.

贝叶斯神经网络的训练与传统神经网络有很大不同, 需要借助于贝叶斯公式, 通过数据的训练, 将神经网络参数从先验分布计算出后验分布. 用 $q(w|\theta)$ 表示神经网络权重的变分后验分布, θ 表示描述权重分布的参数, $p(w|D)$ 表示神经网络权重的真实后验分布, 或者是目标分布, 通过 Kullback-Leibler (KL) 散度衡量这两个分布的差异, 就可以通过最小化 KL 散度使神经网络优化到接近真实的后验分布. 优化算法可以表示为如下公式:

$$\begin{aligned}\theta &= \arg \min_{\theta} D_{\text{KL}}[q(w|\theta)||p(w|D)] \\ &= \arg \min_{\theta} \int q(w|\theta) \ln \frac{q(w|\theta)}{p(w)p(D|w)} dw \\ &= \arg \min_{\theta} D_{\text{KL}}[q(w|\theta)||p(w)] - E_{q(w|\theta)}[\log p(D|w)].\end{aligned}\quad (6)$$

通过贝叶斯公式, 将后验分布 $p(w|D)$ 写为先验分布 $p(w)$ 和数据似然 $p(D|w)$ 的乘积. 其中变分后验 $q(w|\theta)$ 如果无法得到解析形式, 在具体计算中可以通过蒙特卡罗采样方法近似计算. 这种贝叶斯方法是一种基于变分推断的近似贝叶斯方法. 贝叶斯神经网络在散裂反应的同位素截面研究 [47]、原子核质量预测 [48] 等方面都有很好的应用, 同样决策树方法也在原子核质量等参数的预测 [49] 上给出了很好的结果.

这里介绍用贝叶斯神经网络去学习不同原子核的半径大小, 尝试拟合出一个可以统一描述原子核大小的近似模型, 并期望它有一定的预测能力. 使用核

素的质子数 Z 、中子数 N 、平均结合能 B 、单中子分离能 S_n 、单质子分离能 S_p 作为模型的输入, 半径值假设为输出, 两个隐藏层的神经元数为 128 个, 选用 ReLU 函数作为激活函数; 以是否加入双中子、双质子分离能作为标准, 设置输入层神经元为 5 个或 7 个. 图 11 是对核素图上 894 个原子核的数据进行学习得到的模型收敛情况, 达到了比较好的精度.

图 12 给出的是贝叶斯神经网络预测原子核半径训练集和测试集的偏差对比. 测试集的偏差分布宽度约为 0.0265 fm, 稍大于训练集的 0.023 fm. 这个结果好于传统经验公式的计算结果.

如果关注壳层结构附近的同位素链, 比如 Sr 和 Ba, 可以发现贝叶斯神经网络模型能够学习到壳效应对原子核半径的影响规律. 如图 13 所示.

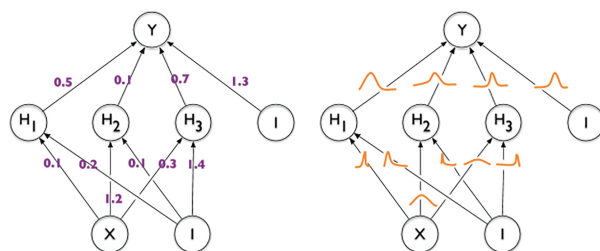


图 10 (网络版彩图)普通神经网络(左)与贝叶斯神经网络(右)

Figure 10 (Color online) Neural network (left) and BNN (right).

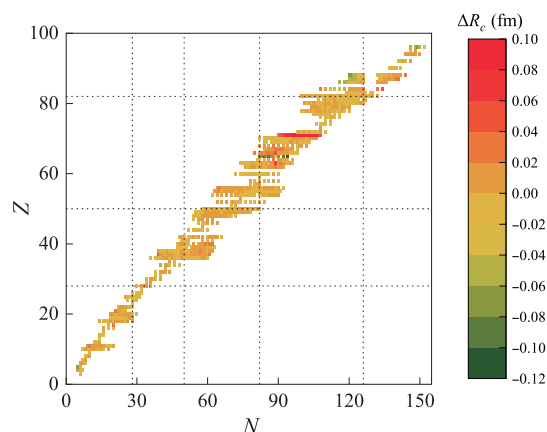


图 11 (网络版彩图)贝叶斯神经网络预测原子核半径与实验值之差

Figure 11 (Color online) Bias between predictions and experiments of nuclear radii with BNN.

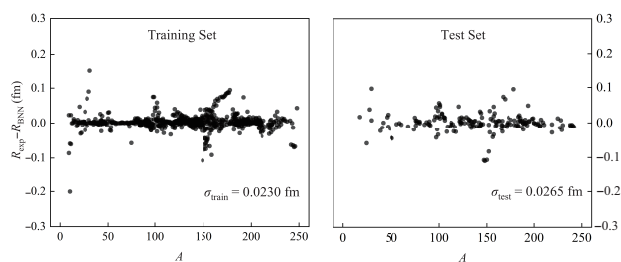


图 12 贝叶斯神经网络预测原子核半径训练集和测试集的偏差对比

Figure 12 Bias of nuclear radii with BNN on (left) training set, and (right) test set.

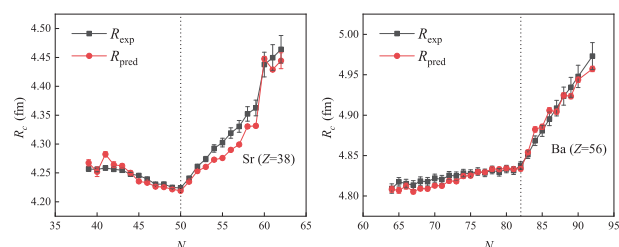


图 13 (网络版彩图)贝叶斯神经网络在幻数附近的原子核同位素链的预测. 左图为Sr同位素链, 右图为Ba同位素链

Figure 13 (Color online) Nuclear radii of Sr isotopes (left) and Ba isotopes (right) with BNN.

3.5 机器学习方法应用于活性靶时间投影室中的事件鉴别

活性靶时间投影室(AT-TPC)正在被广泛应用于中低能核物理实验. AT-TPC结合了活性靶和时间投影室的优点, 能够使中低能实验的探测能力大大提高. 相比固定靶实验, AT-TPC可以准确测量反冲粒子能量, 尤其是低能量反冲粒子, 也避免了固定靶实验中靶厚度对实验的影响, 能够提高靶材料有效厚度. 通过外加磁场和 4π 探测器, 可以精确重建粒子径迹, 实现高精度的粒子鉴别. 探测能力提高的同时, 也给数据分析带来了很大的挑战, 首先是探测器通道成倍增长, 径迹重建需要消耗更多的计算资源, 有效事件判选算法更加复杂. 这些困难在核物理实验中更加突出, 高能领域有着更丰富的处理经验. 机器学习被广泛使用在核物理实验中的各个领域, 包括加速器控制、事件重建与判选、数据分析等. 机器学习技术对于AT-TPC实验同样会有重要的应用价值.

AT-TPC实验中, 束流粒子击中不同的靶元素所产生的事例具有不同的信号特征, 通常可以通过重构

系统的能量和动量等信息鉴别靶核元素. 机器学习方法相比传统方法是一个更好的选择. 通过GEANT4模拟了5 MeV/u的 ^{12}C 束流与AT-TPC $30\text{ cm} \times 30\text{ cm} \times 30\text{ cm}$ 灵敏体积内的甲烷气体的反应, 用得到的带电粒子径迹构建图像作为机器学习的输入. 机器学习模型采用了在ImageNet数据库预训练好的VGG16神经网络作为训练对象. 如图14所示, VGG16网络是一个典型的深度卷积神经网络, 具有很好的泛化能力, 在图像识别领域被广泛使用.

ImageNet是一个按照WordNet层次结构组织的大型图像数据库, 其中层次结构的每个节点都由成百上千的图像描述. ImageNet数据库预训练模型, 能够使模型在应用于小规模的数据集研究时很快获得收敛. 将预训练的VGG16模型再经过AT-TPC模拟数据的进一步训练, 模型在测试数据集中辨别氢核与碳核的精确度几乎达到100%. 实际的实验数据中, 信号带有较大的噪声, 神经网络模型的性能相比模拟数据集会有所降低.

4 总结与展望

机器学习算法不断被应用到物理学研究的各个领域, 尤其是实验物理领域的数据分析中, 在径迹重建和事件选择等多个环节表现出了比传统的物理特征分析方法更好的性能. 机器学习技术也在更多领域如探测器电子学、无监督数据分析、量子多体系统新解法等领域带来新的提升.

实验数据中的涨落, 以及神经网络带来的感知误差似乎没有受到足够的重视. 数据涨落和感知误差的量化是实验研究的重要环节. 对于误差和涨落的考虑目前最成功的是贝叶斯神经网络, 通过数据似然近似推断后验概率分布. 这个方法的一个缺点是计算量大, 模型训练困难, 难以用来构建深度神经网络. 如何构建有效的替代算法是一个待解决的问题. 一个值得关注的尝试是被称为Edward的概率建模框架^[50], 能够比较

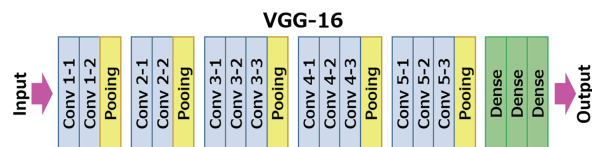


图 14 (网络版彩图) VGG16网络模型的层级构成

Figure 14 (Color online) The construction of VGG16 network.

灵活地构建深度概率模型.

随着高亮度对撞机实验的技术升级, 研制探测器快电子学是重要的任务. 其中机器学习技术对于电子学计算单元的应用场景被看好. 一些预研和尝试表明, 机器学习技术在探测器电子学领域有着重要的应用价值. 机器学习被用于探测器硬件级别的首次尝试是将BDT算法应用于一级触发电子学的FPGA单元^[51], 深度神经网络算法也被逐渐应用于电子学的FPGA算法单元^[52,53]. 图神经网络有望被应用于LHC上的下一代高粒度探测器的读出电子学^[54].

通过机器学习的几个基本概念(包括数据格式、模型、学习策略和任务), 将机器学习进行了笼统的分

类和举例, 其中主要是机器学习在物理学研究中的应用. 机器学习在高能粒子物理的实验研究中有着丰富的应用. 而在核物理, 尤其是中低能核物理实验中的应用还相对较少. 这与中低能核物理实验中数据量和探测器规模有一定关系. 随着机器学习技术的发展, 机器学习技术正在迅速扩展到核物理的实验和理论研究中. 通过几个实例分别介绍了基于监督学习算法的深度神经网络如何学习高能核物理中原始数据的物理特征、基于无监督学习算法对原子核液气相变实验数据的分析、贝叶斯神经网络对于有误差数据的函数近似, 以及机器学习方法应用于活性靶TPC实验中进行事件堆叠判断和处理.

参考文献

- McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Mathemat Biophys*, 1943, 5: 115–133
- Derkach D, Hushchyn M, Likhomanenko T. Machine-learning-based global particle-identification algorithms at the LHCb experiment. *J Phys-Conf Ser*, 2018, 1085: 042038
- Nguyen T Q, Weitekamp III D, Anderson D. Topology classification with deep learning to improve real-time event selection at the LHC. *Comput Softw Big Sci*, 2019, 3: 12
- Komiske P T, Metodieff E M, Nachman B. Pileup Mitigation with Machine Learning (PUMML). *J High Energy Phys*, 2017, 2017: 51
- Pol A A, Aarstad T, Govorkova K. Jet single shot detection. *EPJ Web Conf*, 2021, 251: 04027
- Cogan J, Kagan M, Strauss E. Jet-images: Computer vision inspired techniques for jet tagging. *J High Energy Phys*, 2015, 2015: 118
- Lin J, Freytsis M, Moulton I. Boosting $H \rightarrow b\bar{b}$ with machine learning. *J High Energy Phys*, 2018, 2018: 101
- Qu H, Gouskos L. Jet tagging via particle clouds. *Phys Rev D*, 2015, 92: 056019
- Shimmin C. Particle convolution for high energy physics. arXiv: 2107.02908
- Shlomi J, Battaglia P, Vlimant J R. Graph neural networks in particle physics. *Mach Learn-Sci Technol*, 2021, 2: 021001
- Qasim S R, Kieseler J, Iiyama Y. Learning representations of irregular particle-detector geometry with distance-weighted graph networks. *Eur Phys J C*, 2019, 79: 608
- Carleo G, Troyer M. Solving the quantum many-body problem with artificial neural networks. *Science*, 2017, 355: 602–606
- Aad G, et al. (The ATLAS Collaboration). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys Lett B*, 2012, 716: 1–29
- Aaltonen T, et al. (CDF Collaboration). Search for the standard model Higgs boson produced in association with a W^\pm boson with 7.5 fb⁻¹ integrated luminosity at CDF. *Phys Rev D*, 2012, 86: 032011
- Vaiciulis A. Support vector machines in analysis of top quark production. *Nucl Instrum Meth Phys Res Sect A-Acc Spectrom Detect Assoc Equipm*, 2003, 502: 492–494
- Baldi P, Cranmer K, Faucett T. Parameterized neural networks for high-energy physics. *Eur Phys J C*, 2016, 76: 235
- Cheng T. Recursive neural networks in quark/gluon tagging. *Comput Softw Big Sci*, 2018, 2: 3
- Komiske P T, Metodieff E M, Thaler J. Energy flow networks: Deep sets for particle jets. *J High Energy Phys*, 2019, 2019: 121
- Mikuni V, Canelli F. Point cloud transformers applied to collider physics. *Mach Learn-Sci Technol*, 2021, 2: 035027
- Carleo G, Cirac I, Cranmer K. Machine learning and the physical sciences. *Rev Mod Phys*, 2019, 91: 045002
- Guest D, Cranmer K, Whiteson D. Deep learning and its application to LHC physics. *Annu Rev Nucl Part Sci*, 2018, 68: 161–181
- Radovic A, Williams M, Rousseau D. Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 2018, 560: 41–48
- Mikuni V, Canelli F. Unsupervised clustering for collider physics. *Phys Rev D*, 2021, 103: 092007
- Dillon B M, Faroughy D A, Kamenik J F. Learning the latent structure of collider events. *J High Energy Phys*, 2020, 2020: 206
- De Simone A, Jacques T. Guiding new physics searches with unsupervised learning. *Eur Phys J C*, 2019, 79: 289

- 26 Blance A, Spannowsky M, Waite P. Adversarially-trained autoencoders for robust unsupervised new physics searches. *J High Energy Phys*, 2019, 2019: 47
- 27 Komiske P T, Metodiev E M, Nachman B. Learning to classify from impure samples with high-dimensional data. *Phys Rev D*, 2018, 98: 011502
- 28 Metodiev E M, Nachman B, Thaler J. Classification without labels: Learning from mixed samples in high energy physics. *J High Energy Phys*, 2017, 2017: 174
- 29 Dery L M, Nachman B, Rubbo F. Weakly supervised classification in high energy physics. *J High Energy Phys*, 2017, 2017: 145
- 30 St. John J, Herwig C, Kafkes D. Real-time artificial intelligence for accelerator control: A study at the Fermilab Booster. *Phys Rev Accel Beams*, 2021, 24: 104601
- 31 Liu Z, Zhao W, Song H. Principal component analysis of collective flow in relativistic heavy-ion collisions. *Eur Phys J C*, 2019, 79: 870
- 32 Alonso-Monsalve S, Douqa D, Jesús-Valls C. Graph neural network for 3D classification of ambiguities and optical crosstalk in scintillator-based neutrino detectors. *Phys Rev D*, 2021, 103: 032005
- 33 Bourilkov D, Acosta D, Bortignon P. Machine learning techniques in the CMS search for Higgs Decays to Dimuons. *EPJ Web Conf*, 2019, 214: 06002
- 34 Aad G, et al. (The ATLAS Collaboration). A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector. *Phys Lett B*, 2021, 812: 135980
- 35 Cheong S, Cukierman A, Nachman B. Parametrizing the detector response with neural networks. *J Inst*, 2020, 15: P01030
- 36 Pan S Y, Liu M X, Forero-Romero J, et al. Cosmological parameter estimation from large-scale structure deep learning. *Sci China-Phys Mech Astron*, 2020, 63: 110412
- 37 Huang H, Xiao B, Liu Z. Applications of deep learning to relativistic hydrodynamics. *Phys Rev Res*, 2021, 3: 023256
- 38 Paganini M, de Oliveira L, Nachman B. CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys Rev D*, 2018, 97: 014021
- 39 Ahdida C, Albanese R, Alexandrov A. Fast simulation of muons produced at the SHiP experiment using Generative Adversarial Networks. *J Inst*, 2019, 14: P11028
- 40 Paganini M, de Oliveira L, Nachman B. Accelerating science with generative adversarial networks: An application to 3D particle showers in multilayer calorimeters. *Phys Rev Lett*, 2018, 120: 042003
- 41 Gao C, Höche S, Isaacson J. Event generation with normalizing flows. *Phys Rev D*, 2020, 101: 076002
- 42 He J, He W B, Ma Y G. Machine-learning-based identification for initial clustering structure in relativistic heavy-ion collisions. *Phys Rev C*, 2021, 104: 044902
- 43 Niida T, Miake Y. Signatures of QGP at RHIC and the LHC. *AAPPs Bull*, 2021, 31: 12
- 44 Pang L G, Zhou K, Su N. An equation-of-state-meter of quantum chromodynamics transition from deep learning. *Nat Commun*, 2018, 9: 210
- 45 Wang R, Ma Y G, Wada R. Nuclear liquid-gas phase transition with machine learning. *Phys Rev Res*, 2020, 2: 043202
- 46 Khan Md A, Alam M, Hasan M, et al. Hyperspherical few-body model calculation for cascade Ξ exotic hypernuclei with reference to nuclear driplines. *Bulgarian J Phys*, 2020, 47: 87–101
- 47 Ma C W, Peng D, Wei H L. Isotopic cross-sections in proton induced spallation reactions based on the Bayesian neural network method. *Chin Phys C*, 2020, 44: 014104
- 48 Niu Z M, Liang H Z. Nuclear mass predictions based on Bayesian neural network approach with pairing and shell effects. *Phys Lett B*, 2018, 778: 48–53
- 49 Gao Z P, Wang Y J, Lü H L. Machine learning the nuclear mass. *Nucl Sci Tech*, 2021, 32: 109
- 50 Tran D, Hoffman M D, Saurous R A, et al. Deep probabilistic programming. arXiv: 1701.03757
- 51 Acosta D, Brinkerhoff A, Busch E. Boosted decision trees in the level-1 muon endcap trigger at CMS. *J Phys-Conf Ser*, 2018, 1085: 042042
- 52 Duarte J, Han S, Harris P. Fast inference of deep neural networks in FPGAs for particle physics. *J Inst*, 2018, 13: P07027
- 53 Aarrestad T, Loncar V, Ghilmetti N. Fast convolutional neural networks on FPGAs with hls4ml. *Mach Learn-Sci Technol*, 2021, 2: 045015
- 54 Iiyama Y, Cerminara G, Gupta A. Distance-weighted graph neural networks on FPGAs for real-time particle reconstruction in high energy physics. *Front Big Data*, 2021, 3: 44

Machine learning applications in nuclear physics

HE WanBing^{1*}, HE JunJie^{2,3}, Wang Rui² & MA YuGang^{1*}

¹*Key Laboratory of Nuclear Physics and Ion-beam Application (MOE), Institute of Modern Physics, Fudan University, Shanghai 200433, China;*

²*Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China;*

³*University of Chinese Academy of Sciences, Beijing 100049, China*

Machine learning is the most successful and active research direction in the field of artificial intelligence. It is being applied in various fields, such as big data analysis, autonomous driving, natural language processing, Internet technology, medical diagnosis, and market analysis. Accordingly, it is also widely used in the experimental and theoretical research of particle physics. This article discusses several machine learning techniques that are particularly relevant to nuclear physics. Additionally, the application of machine learning in various contexts of nuclear physics research is demonstrated through several different cases, including the deep convolutional neural networks combined with theoretical models to extract information about the nuclear cluster structure from low level data, unsupervised learning methods applied to experimental data to extract information about nuclear liquid-gas phase transitions, and neural network quantum state (NQS) as a trial wave function to solve the Schrodinger equation. Bayesian neural networks are applied to fit and predict the nucleus radius, and image recognition is applied to identify events in the active target time projection chamber detector. It shows the wide application of machine learning technology in the field of nuclear physics.

machine learning, nuclear physics, nuclear cluster, nuclear liquid-gas phase transition

PACS: 21.10.Ft, 25.70.Pq, 29.40.Gx, 29.85.+c, 21.60.Kn

doi: [10.1360/SSPMA-2021-0309](https://doi.org/10.1360/SSPMA-2021-0309)