

目录

一、会务指南	3
二、会议组织	4
三、会议日程	6
四、邀请报告	16
五、分会报告 A1 专题：机器学习理论	18
六、分会报告 A3 专题：大模型的理解与应用	19
七、分会报告 A4 专题：脑启发的神经网络模型、算法与应用（I）	21
八、分会报告 A5 专题：机器学习与微分方程	22
九、分会报告 A6 专题：动理学方程与机器学习	24
十、分会报告 A7 专题：GENERATIVE AI FOR SCIENCE（I）	25
十一、分会报告 A8 专题：机器学习与化学体系计算（I）	27
十二、分会报告 A9 专题：机器学习在空天动力中的应用（I）	29
十三、分会报告 B1 专题：人工智能的数学理论	30
十四、分会报告 B3 专题：机器学习与交叉（I）	31
十五、分会报告 B4 专题：脑启发的神经网络模型、算法与应用（II）	34
十六、分会报告 B5 专题：反问题中的机器学习方法	36
十七、分会报告 B6 专题：科学机器学习方法及其在计算力学中的应用	38
十八、分会报告 B7 专题：GENERATIVE AI FOR SCIENCE（II）	39
十九、分会报告 B8 专题：机器学习与化学体系计算（II）	41
二十、分会报告 B9 专题：机器学习在空天动力中的应用（II）	42
二十一、分会报告 C1 专题：深度学习理论与进展（I）	43

二十二、分会报告 C2 专题：机器学习与统计	44
二十三、分会报告 C3 专题：科学机器学习	46
二十四、分会报告 C4 专题：统计机器学习与 AI4MATH	47
二十五、分会报告 C5 专题：数据科学中的随机模型与算法（I）	49
二十六、分会报告 C6 专题：AI FOR IMAGING.....	51
二十七、分会报告 C7 专题：深度学习求解复杂流体的高效算法及分析（I）	53
二十八、分会报告 C8 专题：科学智能中的高性能计算	56
二十九、分会报告 C9 专题：机器学习与最优化方法（I）	57
三十、分会报告 D1 专题：深度学习理论与进展（II）	59
三十一、分会报告 D2 专题：RECENT ADVANCES ON LEARNING TO OPTIMIZE.....	61
三十二、分会报告 D3 专题：LEARNING DYNAMICAL MODELS FROM DATA: ALGORITHMS, ANALYSES AND APPLICATIONS	64
三十三、分会报告 D4 专题：机器学习与交叉（II）	65
三十四、分会报告 D5 专题：数据科学中的随机模型与算法（II）	67
三十五、分会报告 D6 专题：机理驱动的深度学习算法和理论	68
三十六、分会报告 D7 专题：深度学习求解复杂流体的高效算法及分析（II）	70
三十七、分会报告 D9 专题：机器学习与最优化方法（II）	72
三十八、分会报告 E1 专题：学生分会（I）	73
三十九、分会报告 E2 专题：学生分会（II）	75

一、会务指南

2024 年中国机器学习与科学应用大会（CSML2024）由上海交通大学自然科学研究院、上海交通大学数学科学学院、上海国家应用数学中心（上海交通大学分中心）、科学工程计算教育部重点实验室、上海市现代分析前沿科学研究基地主办，北京大学国际机器学习研究中心、北京国际数学研究中心协办。

大会将聚焦机器学习的三大关键领域。在机器学习的数学理论方面，我们将深入探讨机器学习的理论基础和模型建立，从数学视角理解其深层次结构。在机器学习的科学应用领域，我们将讨论机器学习如何解决复杂的科学问题，并借鉴科学计算的方法开发新算法。在机器学习的工程应用方面，我们将关注如何将机器学习的研究成果转化为实际工程实践，解决技术难题，推动技术革新和效率提升。

The conference will focus on three key areas of machine learning. In the area of mathematical theory, we will delve into the theoretical foundations and model building of machine learning, understanding its deep structures from a mathematical perspective. In the area of scientific applications, we will discuss how machine learning can solve complex scientific problems and how methods from scientific computing can be used to develop new algorithms. In the area of engineering applications, we will focus on how to translate machine learning research outcomes into practical engineering practices to solve technical challenges, thereby driving technological innovation and efficiency improvements.

会议时间：2024 年 8 月 9 日（周五）-8 月 11 日（周日），周五注册，周末两天开会

会议地点：上海交通大学（上海市闵行区东川路 800 号）

会议网站：<https://c2sml.cn/conference.html>

会务组联系方式：张耀宇：zhyy.sjtu@sjtu.edu.cn

二、会议组织

大会主席：

鄂维南（中国科学院院士，北京大学，国际机器学习研究中心）

金石（欧洲人文和自然科学院外籍院士、欧洲科学院院士，上海交通大学，自然科学研究院）

会议学术委员会（按姓氏拼音排序）：

包刚（院士，浙江大学）

鄂维南（院士，北京大学）

高小山（研究员，中国科学院数学与系统科学研究院）

江松（院士，北京应用物理与计算数学研究所）

金石（教授，上海交通大学）

刘卫东（教授，上海交通大学）

马志明（院士，中国科学院数学与系统科学研究院）

王立威（教授，北京大学）

杨志坚（教授，武汉大学）

印卧涛（阿里巴巴（美国）达摩院决策智能实验室）

张钹（院士，清华大学）

张平文（院士，北京大学）

张志华（教授，北京大学）

周志华（教授，南京大学）

圆桌论坛：

鄂维南（中国科学院院士、北京大学讲席教授，北京大学国际机器学习研究中心）

金石（欧洲人文和自然科学院外籍院士、欧洲科学院院士，上海交通大学讲席教授，上海交通大学自然科学研究院/数学科学学院）

董 彬（北京大学博雅特聘教授，北京大学北京国际数学研究中心，北京大学国际机器学习研究中心）

周栋焯（特聘教授，上海交通大学，自然科学研究院/数学科学学院）

张耀宇（长聘教轨副教授，上海交通大学，自然科学研究院/数学科学学院）

林洲汉（长聘教轨副教授，上海交通大学，约翰·霍普克罗夫特计算机科学中心）

陈思衡（长聘教轨副教授，上海交通大学，上海人工智能实验室）

邀请大会报告人（按姓氏拼音排序）：

陈景润（教授，中国科学技术大学）

李千骁（助理教授，新加坡国立大学）

王 磊（研究员，中国科学院物理研究所）

张耀宇（长聘教轨副教授，上海交通大学自然科学研究院/数学科学学院）

会议组织委员会（按姓氏拼音排序）：

组织委员会主席：

张耀宇（上海交通大学）

组织委员会成员：

董 彬（北京大学）

焦雨领（武汉大学）

罗 涛（上海交通大学）

马 征（上海交通大学）

许志钦（上海交通大学）

三、会议日程

2024年8月10日 - 11日, 上海交通大学闵行校区陈瑞球楼

时间	活动内容	主持人	地点
8月10日			
7:30-8:20	签到		陈瑞球楼 100 会议室
8:20-8:40	致辞	张耀宇	
8:40-9:25	邀请报告 张耀宇 (上海交通大学) 神经网络样本效率的乐观估计	董彬	
9:25-10:10	邀请报告 李千骁 (新加坡国立大学) 从数据中构建宏观动力学模型		
10:10-10:30	茶歇		
10:30-12:00	圆桌论坛	张耀宇	
12:00-13:30	午餐		闵行校区各餐厅
13:30-15:30	A1 专题：机器学习理论 李建：梯度算法的隐式偏差与对抗鲁棒性 魏哲巍：谱域图机器学习理论与方法 吴磊：Understanding the implicit bias of stochastic gradient descent: A dynamical stability perspective 李彤阳：量子机器学习算法	王立威	会场 1 陈瑞球楼 310
	A3 专题：大模型的理解与应用	黄维然	会场 3

	<p>桂韬：大模型智能体能力对齐</p> <p>袁洋：定位即智能</p> <p>张辉帅：文生图大模型的记忆问题</p> <p>易鸣洋：Towards Understanding the Working Mechanism of Text-to-Image Diffusion Model</p>		<p>陈瑞球 楼 311</p>
	<p>A4 专题：脑启发的神经网络模型、算法与应用 (I)</p> <p>李国齐：基于脉冲神经网络的类脑大模型</p> <p>邓磊：脉冲神经动力学网络及硬件实现</p> <p>张铁林：类脑智能算法的思考及未来发展</p> <p>余肇飞：脉冲视觉：建立机器视觉与生物视觉之间的桥梁</p>	<p>周栋焯 李松挺</p>	<p>会场 4 陈瑞球 楼 312</p>
	<p>A5 专题：机器学习与微分方程</p> <p>周涛：Deep adaptive sampling for numerical PDEs</p> <p>黄建国：Friedrichs 学习：通过深度学习求解偏微分方程弱解</p> <p>廖奇峰：KRnet 流模型密度函数估计方法及高维偏微分方程求解</p> <p>曹语：探索扩散模型生成过程中的最佳动力学方程</p>	<p>黄建国</p>	<p>会场 5 陈瑞球 楼 313</p>
	<p>A6 专题：动理学方程与机器学习</p> <p>熊涛：动理学方程基于神经网络的动态区域分解方法</p> <p>王艳莉：Solving Boltzmann equation using neural sparse representation</p> <p>吴科科：基于渐近保持格式的深度学习方法求解多尺度动理学方程</p> <p>马征：DeepRTE：辐射输运方程的预训练神经算子</p>	<p>陈景润 马征</p>	<p>会场 6 陈瑞球 楼 429</p>
	<p>A7 专题：Generative AI for Science (I)</p> <p>黄文炳：几何深度学习在物质科学中的应用</p> <p>许洪腾：数值算法驱动的生成式模型架构设计与参数微调</p>	<p>王涵 王磊</p>	<p>会场 7 陈瑞球 楼 430</p>

	张潘: Generative decoding for quantum error-correcting codes 吴泰霖: 基于扩散生成模型的物理系统设计和控制		
	A8 专题: 机器学习与化学体系计算 (I) 陈基: 机器学习量子蒙特卡洛在化学物理中的应用 王笑楠: AI 集成指导高精度化学合成与低碳化工材料设计 褚维斌: AI 赋能激发态载流子动力学 陆子恒: 材料基础模型: 从模拟到生成 吕定顺: GPU-accelerated Auxiliary-field quantum Monte Carlo with multi-Slater determinant trial states	胡伟 李颖洲	会场 8 陈瑞球 楼 432
	A9 专题: 机器学习在空天动力中的应用 (I) 温新: 基于数据融合的流场测量与控制 韩旺: 人工智能赋能的发动机流场测量系统 安健: 深度学习技术与空天动力 蔡尊: 宽范围超声速燃烧基础研究	张天汉	会场 9 陈瑞球 楼 507
15:30-16:00	茶歇		
16:00-18:00	B1 专题: 人工智能的数学理论 孙若愚: 大模型训练和 Adam: 原理探究和新方法 Adam-mini 邹荻凡: Transformer 架构的运行机制研究: 模型深度和多头注意力的作用 许志钦: 大模型的推理与记忆 王宇光: 生成式 AI 的数学基础与分子设计应用	张耀宇	会场 1 陈瑞球 楼 310
	B3 专题: 机器学习与交叉 (I) 戴书洋: 基于优化的位错结构与演化的预测方法	胡丹	会场 3 陈瑞球

	<p>胡丹: 基于非均匀随机游走的自适应采样</p> <p>邱越: Sparse discovery of differential equations based on multi-fidelity Gaussian process</p> <p>赵进: pETNNs: Partial Evolutionary Tensor Neural Networks for Solving Time-dependent Partial Differential Equations</p> <p>焦雨领: Unsupervised transfer learning via adversarial contrastive training</p>		楼 311
	<p>B4 专题: 脑启发的神经网络模型、算法与应用 (II)</p> <p>余山: 神经网络中的具身概念形成、理解和交流</p> <p>弭元元: A Brain-inspired Computational Model for Spatio-temporal Sequence Recognition</p> <p>杜凯: 大脑精细模拟: 神经科学与人工智能的新前沿</p> <p>刘翀鸣: Dendritic Integration-Inspired Artificial Neural Networks Enhance Data Correlation</p>	周栋焯 李松挺	会场 4 陈瑞球 楼 312
	<p>B5 专题: 反问题中的机器学习方法</p> <p>贾骏雄: 学习统计逆问题中先验分布的预测函数</p> <p>王玉亮: 数据辅助的两阶段方法解决反随机源散射问题</p> <p>尹运文: Physics-aware deep learning framework for the limited aperture inverse obstacle scattering problem</p> <p>张一萱: 最优运输理论的反问题应用</p>	闫亮	会场 5 陈瑞球 楼 313
	<p>B6 专题: 科学机器学习方法及其在计算力学中的应用</p> <p>史作强: NSNO: 基于 Neumann 级数的 Helmholtz 方程神经算子方法</p> <p>孙赫: 人工智能加速超声 CT 影像重建</p> <p>孟旭辉: 物理机理神经网络中的物理模型校准方法</p>	黄政宇	会场 6 陈瑞球 楼 429

	张伟伟：计算力学的两种范式对比分析、融合与展望		
	B7 专题：Generative AI for Science (II) 汪鸿帅：通用、准确、生成式预训练分子大模型 Uni-Mol 陈品：基于 AI 的晶体结构生成：算法及应用 孙剑：人工智能生成与泛化的最优传输理论与方法 叶财渊：Con-CDVAE：晶体结构的条件生成方法	王涵 王磊	会场 7 陈瑞球 楼 430
	B8 专题：机器学习与化学体系计算 (II) 兰峥岗：机器学习在非绝热动力学中的应用 陈林江：科研智能体助力理实交融的机器化学家探索 胡素磊：解析、数值、回归和智能融合驱动的动力学方程标度和分类理论 胡伟：算法和算力驱动的第一性原理计算材料模拟	胡伟 李颖洲	会场 8 陈瑞球 楼 432
	B9 专题：机器学习在空天动力中的应用 (II) 徐辉：结合第一性原理的 AI 技术及其在航空航天领域的创新应用 陈东平：固体推进剂分子模拟及其燃烧反应动力学研究 韩啸：机器学习在燃烧振荡中的应用	张天汉	会场 9 陈瑞球 楼 507
8 月 11 日			
8:30-9:15	邀请报告 陈景润(中国科学技术大学) CAX 设计分析一体化	罗涛	陈瑞球 楼 100
9:15-10:00	邀请报告 王磊(中国科学院物理研究所) 物质科学研究中的生成模型		会议室
10:00-10:15	茶歇		

10:15-12:15	<p>C1 专题：深度学习理论与进展（I）</p> <p>方聪：实例更优的加速梯度下降算法</p> <p>刘方辉：Is double descent well-defined? From kernel methods to neural networks</p> <p>廖振宇：深度学习中的随机矩阵最新进展</p> <p>曹原：复杂学习系统中的多重下降现象</p>	吴磊	会场 1 陈瑞球 楼 310
	<p>C2 专题：机器学习与统计</p> <p>林乾：高维核回归的若干新现象</p> <p>林伟：A Statistical Theory of Regularization-Based Continual Learning</p> <p>杨朋昆：尺度率的多阶段性质与极限</p> <p>毛晓军：Ternary Quantization for Distributed Mean Estimation</p>	毛晓军	会场 2 陈瑞球 楼 309
	<p>C3 专题：科学机器学习</p> <p>白磊：地球计算</p> <p>刘扶芮：之江实验室生物智能计算平台的特色应用</p> <p>张拳石：神经网络是否可以被严谨地解释清楚？以及可解释性技术在大模型上的应用落地</p> <p>刘鹏飞：大模型中的复杂推理</p>	林洲汉	会场 3 陈瑞球 楼 311
	<p>C4 专题：统计机器学习与 AI4Math</p> <p>邱怡轩：AI for Permutation: Trainable Sparse Optimal Transport</p> <p>王文佳：Constrained Policy Optimization with Explicit Behavior Density (CPED) for Offline Reinforcement Learning</p> <p>刘勇：图神经网络的泛化理论分析</p> <p>刘征瀛：大语言模型的数学推理能力</p>	胡天阳	会场 4 陈瑞球 楼 312

	<p>C5 专题：数据科学中的随机模型与算法 (I)</p> <p>柴利慧：基于随机梯度重构的地震波成像方法</p> <p>成诚：Graph Fourier Transforms on Directed Graphs</p> <p>郭玲：Uncertainty Quantification in Scientific Machine Learning</p> <p>黄忠亿：针对奇异摄动问题的算子学习</p>	<p>李磊</p> <p>凌舒扬</p> <p>翟佳羽</p>	<p>会场 5</p> <p>陈瑞球</p> <p>楼 313</p>
	<p>C6 专题：AI for Imaging</p> <p>胡战利：医学 PET 成像与仪器研发</p> <p>牛田野：智能锥束 CT 成像</p> <p>庞彤瑶：图像恢复的无监督学习算法</p> <p>Oscar Leong: Generative Networks for Inverse Problems without Ground-Truth Data</p> <p>张法：基于 AI 的冷冻电镜图像数据处理技术</p>	<p>包承龙</p> <p>梁经纬</p>	<p>会场 6</p> <p>陈瑞球</p> <p>楼 429</p>
	<p>C7 专题：深度学习求解复杂流体的高效算法及分析 (I)</p> <p>王飞：求解偏微分方程的随机神经网络方法</p> <p>洪启臻：稀疏观测数据下深度神经算子学习重建高温热化学非平衡流场</p> <p>顾亦奇：Deep adaptive basis Galerkin method for evolution equations</p> <p>谢芳芳：基于机器学习的流动建模和控制</p>	<p>毛志平</p> <p>孟旭辉</p>	<p>会场 7</p> <p>陈瑞球</p> <p>楼 430</p>
	<p>C8 专题：科学智能中的高性能计算</p> <p>贾伟乐：HPC+AI 驱动的微观尺度模拟初探</p> <p>李士刚：大模型高效可扩展并行策略</p> <p>杨海龙：精度可复现模型弹性训练技术</p> <p>赵世振：光电混合的高性能网络互连</p> <p>琚锡廷：融合气象大模型的高性能资料同化系统</p>	<p>贾伟乐</p>	<p>会场 8</p> <p>陈瑞球</p> <p>楼 432</p>

	<p>C9 专题：机器学习与最优化方法 (I)</p> <p>严骏驰：组合优化问题的机器学习求解方法</p> <p>王祥丰：基于大语言模型的云计算调度方法设计</p> <p>陈子昂：On Representing (Mixed-Integer) Linear Programs by Graph Neural Networks</p> <p>袁坤：An Effective AI-Driven Algorithm for Decentralized Optimization</p>	袁坤 文再文	会场 9 陈瑞球 楼 507
12:15-13:30	午餐		闵行校区 各餐厅
13:30-15:30	<p>D1 专题：深度学习理论与进展 (II)</p> <p>龙吉昊：分析随机特征模型和两层神经网络的对偶框架</p> <p>袁洋：矩阵信息论在人工智能中的应用</p> <p>张景昭：Understanding Implicit Bias in Input Spaces</p> <p>罗涛：两层神经网络全局最小值的几何性质与局部恢复</p>	吴磊	会场 1 陈瑞球 楼 310
	<p>D2 专题：Recent Advances on Learning to Optimize</p> <p>戴文睿：可靠成像的深度学习逆问题求解</p> <p>顾心悦：可信预测与时序可解释方法</p> <p>李雨晴：凝聚现象在多种算法训练的神经网络中的普遍性</p> <p>谢中林：ODE-based Learning to Optimize</p>	梁经纬	会场 2 陈瑞球 楼 309
	<p>D3 专题：Learning dynamical models from data: algorithms, analyses and applications</p> <p>林伟：Predicting and modulating complex dynamics using data-driven and machine learning techniques</p> <p>吴开亮：From Data to Dynamics: Deep Modeling of Unknown Differential Equations</p> <p>熊诗颖：RoeNet：基于数据预测双曲系统的间断解</p>	祝爱卿 李千骁	会场 3 陈瑞球 楼 311

	祝爱卿：动态高斯混合近似：从数据中学习随机微分方程		
	<p>D4 专题：机器学习与交叉（II）</p> <p>高卫国：GAN 训练的三个阶段</p> <p>干则成：准二维库伦体系中核函数的指数函数求和估计及其 $O(N)$ 复杂度的随机分批分子动力学模拟</p> <p>蒋诗晓：传统无网格方法求解黎曼流形上的向量场方程</p> <p>张继伟：非局部模型的理论 and 计算</p>	戴书洋	会场 4 陈瑞球 楼 312
	<p>D5 专题：数据科学中的随机模型与算法（II）</p> <p>魏朝祯：Wasserstein 梯度流的保结构原始对偶方法</p> <p>于海军：微分方程高效神经网络方法及自适应数据采样技巧</p> <p>周翔：抽样不变测度的弱生成网络抽样方法</p> <p>邹东勉：基于 Gromov-Monge 正则化的单调生成器</p>	李磊 凌舒扬 翟佳羽	会场 5 陈瑞球 楼 313
	<p>D6 专题：机理驱动深度学习算法和理论</p> <p>刘皓：Exploring low-dimensional data structures by deep neural networks with applications on operator learning</p> <p>王亚婷：基于学习的多尺度流动多连续体模型</p> <p>徐新鹏：关于“复杂微环境中细胞迁移随机动力学”的深度学习学习方法</p> <p>冯寒：卷积神经网络的特征稀疏性及其驱动的学习策略研究</p>	王东 刘皓	会场 6 陈瑞球 楼 429
	<p>D7 专题：深度学习求解复杂流体的高效算法及分析（II）</p> <p>蔡声泽：基于人工智能的流场可视化测量与计算</p> <p>林晨森：使用算子神经学习弥合多尺度气泡动力学中的尺度差异</p>	毛志平 孟旭辉	会场 7 陈瑞球 楼 430

	<p>金鹏展：基于 MIONet 的 PDE 混合迭代法</p> <p>金晓威：物理增强流场智能计算方法</p>		
	<p>D9 专题：机器学习与最优化方法 (II)</p> <p>高斌：Hyper-gradient in bilevel optimization: efficient computation by Krylov Subspace and enhanced investigation in reinforcement learning</p> <p>林涛：面向海量数据和异构数据的高效深度学习</p> <p>严明：分布式优化算法</p> <p>朱桢源：一类复合优化问题的最优复杂度上界和下界</p>	<p>袁坤</p> <p>文再文</p>	<p>会场 9</p> <p>陈瑞球</p> <p>楼 507</p>
15:40-17:40	<p>E1 专题：学生分会 (I)</p> <p>梅子健：基于机器学习的结构优化算法</p> <p>白志威：神经网络损失景观的深度嵌入原则</p> <p>严菁：泛化误差的相变行为</p> <p>李晓龙：Loss Spike in Training Neural Networks</p>	<p>白志威</p>	<p>会场 1</p> <p>陈瑞球</p> <p>楼 310</p>
	<p>E2 专题：学生分会 (II)</p> <p>尹树雨：通过 Fokker-Planck 方程探究半梯度 Q 学习的隐式偏好</p> <p>杭良慨：Input gradient annealing neural network for solving low-temperature Fokker-Planck equations</p> <p>陈楚淇：Quantifying Training Difficulty and Accelerating Convergence in Neural Network-Based PDE Solvers</p> <p>王冰鑫：大型 Transformer 是更好的 EEG 学习者</p>	<p>尹树雨</p>	<p>会场 2</p> <p>陈瑞球</p> <p>楼 309</p>

四、邀请报告

1. 张耀宇

简介: 张耀宇 2012 年于上海交通大学致远学院获物理学学士学位。2016 年于上海交通大学获数学博士学位。2016 年至 2020 年，分别在纽约大学阿布扎比分校&柯朗研究所、普林斯顿高等研究院做博士后研究。2020 年加入上海交通大学自然科学研究院/数学科学学院任长聘教轨副教授。他的研究聚焦于深度学习的基础理论，亮点研究包括建立神经网络动力学态的相图，发现频率原则和嵌入原则两种基本规律。

报告题目: 神经网络样本效率的乐观估计

摘要: 估计神经网络的样本效率，即拟合目标函数所需的样本量，是深度学习泛化理论中的一个关键问题。我们提出了一种基于凝聚现象的新样本量估计方法——乐观估计，定量描述了神经网络通过凝聚现象所能实现的极限样本效率。研究结果表明，增加网络的宽度以提升参数量并不会损害其样本效率，而增加（不必要的）连接则会显著降低样本效率。我们的分析为实践中普遍采用的增宽而非增加连接的扩展策略提供了理论支持，并为下一代神经网络架构的研发提供了理论基础。

2. 李千骁

简介: Qianxiao Li is an assistant professor in the Department of Mathematics, and a principal investigator in the Institute for Functional Intelligent Materials, National University of Singapore. He graduated with a BA in mathematics from the University of Cambridge and a PhD in applied mathematics from Princeton University. His research interests include the interplay of machine learning and dynamical systems, control theory and data-driven methods for science and engineering.

报告题目: 从数据中构建宏观动力学模型

摘要: We discuss some recent work on constructing stable and interpretable macroscopic thermodynamics from trajectory data using deep learning. We develop a modelling approach: instead of generic neural networks as functional approximators, we use a model-based ansatz for the dynamics following a suitable generalisation of the classical Onsager principle for non-equilibrium systems. This allows the

construction of macroscopic dynamics that are physically motivated and can be readily used for subsequent analysis and control. We discuss applications in the analysis of polymer stretching in elongational flow.

3. 陈景润

简介：陈景润，中国科学技术大学数学科学学院教授，主要从事科学计算与人工智能、工业软件等方面的研究。

报告题目：CAX 设计分析一体化

摘要：本报告将从工业软件的角度，讨论计算机辅助设计（CAD）以及计算机辅助工程（CAE）的一体化问题，并在最优设计的牵引下探讨如何使用深度学习的工具更好地实现科学计算以及逆向设计。

4. 王磊

简介：王磊 2006 年本科毕业于南京大学，2011 年在中国科学院物理研究所获得博士学位，此后在苏黎世联邦理工学院从事计算量子物理的博士后研究，2016 年加入中科院物理所工作。主要研究方向是机器学习与量子多体计算的交叉领域。

报告题目：物质科学研究中的生成模型

摘要：长期以来计算机最显著的特点是算得快、算得准。生成式人工智能赋予了计算机从经验中学习“直觉”的能力，甚至可以用来“创造”新的经验。定量描述这种“直觉”和“创造”的数学工具和统计物理一样，都是自然界中的概率分布。生成模型学习数据样本背后的概率分布，并据此随机采样产生新的数据样本。这个报告将从物理学视角解读 ChatGPT 和 Sora 等生成式人工智能背后的生成模型。基于它们统计物理的深层联系，生成模型在原子尺度物质结构设计和变分自由能计算中具有特别的优势。我将以晶体按需生成[1]和稠密氢状态方程计算[2]为例说明生成模型在物质科学研究中的应用前景。 [1] Z. Cao, X. Luo, J. Lv, and L. Wang, Space Group Informed Transformer for Crystalline Materials Generation, arXiv:2403.15734. [2] Hao Xie, Zi-Hang Li, Han Wang, Linfeng Zhang, and Lei Wang, Deep Variational Free Energy Approach to Dense Hydrogen, Phys. Rev. Lett. 131, 126501 (2023)

五、分会报告 A1 专题：机器学习理论

A1-1 梯度算法的隐式偏差与对抗鲁棒性

李建，清华大学

摘要：训练深度神经网络是一个高度非凸的优化问题。但是在实际中，简单的梯度方法能够找到不仅最小化训练误差，而且展现出很好泛化能力。但是，最近多个理论研究和实验证据表明，梯度方法可能是一把“双刃剑”，一方面它找到的解在干净数据上具有良好的泛化能力，但同时也容易受到小的对抗性扰动的影响。这些现象引发了一系列关于深度学习训练的理论问题：梯度类算法在神经网络训练过程中引入了哪些特定的隐式偏差，以及训练过程的哪些方面影响了深度学习的对抗鲁棒性？我们系统研究了梯度下降训练过程中称之为“特征平均”的一个特定隐式偏差，并且认为这是导致深度神经网络非鲁棒性的重要因素之一。即使数据中存在多个能够分类数据的区分性特征，梯度下降训练的神经网络倾向于学习这些特征的平均，而不是区分并利用每个特征。我们从理论上分析了二层 ReLU 网络在二分类任务中梯度下降训练动态，并严格证明了梯度下降收敛到特征平均的模式。我们同时探讨了特征平均和已知的多个鲁棒性模型的关系，以其该理论结果对于实际训练鲁棒的网络的启发。

A1-2 谱域图机器学习理论与方法

魏哲巍，中国人民大学

摘要：近年来，由于图结构数据的强大表达能力，用机器学习方法分析和挖掘图数据的研究越来越受到重视。图神经网络（GNN）是一类基于深度学习的处理图数据的方法，由于其处理结构化非欧几何数据的出色能力，GNN 在人工智能的各个交叉领域（如智能制药、智慧交通、智慧金融等）展现出了卓越的性能。图卷积神经网络是图神经网络研究中一类重要的方法，它们在拉普拉斯谱域中设计和学习不同的图卷积，具有良好的理论保证和可解释性。本报告拟先介绍图神经网络的任务和一些前沿应用，然后从图傅里叶变换、图卷积的设计和图谱滤波器的多项式近似等方面探讨图卷积神经网络的理论基础，最后将讨论我们在图卷积神经网络领域所做的一些工作和对未来工作的展望。

A1-3 Understanding the implicit bias of stochastic gradient descent: A dynamical stability perspective

吴磊，北京大学数学科学学院

摘要： In deep learning, models are often over-parameterized, which leads to concerns about algorithms picking solutions that generalize poorly. Fortunately, stochastic gradient descent (SGD) always converges to solutions that generalize well even without needing any explicit regularization, suggesting certain “implicit regularization” at work. This talk will provide an explanation of this striking phenomenon from a stability perspective. Specifically, we show that a stable minimum of SGD must be flat, as measured by various norms of local Hessian. Furthermore, these flat minima provably generalize well for two-layer neural networks and diagonal linear networks. As opposed to popular continuous-time analysis, our stability analysis respects the discrete nature of SGD and can explain the effect of finite learning rates, batch size, and why SGD often generalizes better than GD.

A1-4 量子机器学习算法

李彤阳，北京大学计算机学院

摘要： 随着量子计算的高速发展，量子计算机上的机器学习算法受到越来越广泛的关注。本报告讲简要介绍量子机器学习算法的最新进展，从机器学习理论的角度讨论量子计算对于机器学习问题的加速。

六、分会报告 A3 专题：大模型的理解与应用

A3-1 大模型智能体能力对齐

桂韬，复旦大学/现代语言学研究院

摘要： 几十年来，人类一直在寻求创造接近或超越人类智力的人工智能（AI），智能体被誉为实现这一目标的一条有希望的道路。大语言模型（LLM）越来越被视为通用人工智能（AGI）发展的潜在催化剂，激发了创建多功能智能体的希望。本次报告将介绍基于大模型的智能体的综合框架，讨论确保这些智能体符合人类能力和价值观的策略，并提供对此类智能体未来发展的见解。

A3-2 定位即智能

袁洋，清华大学交叉信息研究院

摘要： OpenAI 提出了著名的“压缩即智能”观点，但没有明确其与传统文件压缩算法的差异，也没有给出其与常用算法的联系。相比之下，我认为“定位即智能”是一个更好的视角，不仅拥有优美的理论支撑（范畴论），还可以和大量预训练、多模态算法构建起联系。本报告将介绍在此视角下，关于对比学习与大模型算法的理论分析。

A3-3 文生图大模型的记忆问题

张辉帅，北京大学

摘要： 文生图大模型可以生成高质量的图片，但也被发现它们会记住训练数据。文生图大模型记忆的问题还会产生一些社会影响，比如版权保护问题。本报告首先从理论上探讨了扩散模型的记忆和泛化性质，提出了与训练动态相关的泛化误差理论估计。在早停训练的情况下，基于样本规模的泛化误差呈现多项式小的特征，成功避免了维度灾难。此外，基于数据依赖场景的分析，阐明了目标分布中模式距离对模型泛化的负面影响。本报告还将介绍我们在商业文生图大模型的记忆攻击成果。评估了如 ChatGPT、Copilot 和 Gemini 等商业 T2I 生成系统在版权侵犯上的安全性。实验发现，Copilot 和 Gemini 在面对简单提示词攻击时，仅能阻止 12%和 17%的攻击，而 ChatGPT 则阻止了 84%的攻击。提出了一种更强大的自动破解流程，利用 LLM 优化器生成提示词，在不进行权重更新或梯度计算的情况下，实现了对生成系统安全防护的突破，成功将 ChatGPT 的阻止率降至 11%，并在 76%的情况下生成了受版权保护的内容。

A3-4 Towards Understanding the Working Mechanism of Text-to-Image Diffusion Model

易鸣洋，华为诺亚方舟实验室

摘要： Recently, the strong latent Diffusion Probabilistic Model (DPM) has been applied to high-quality Text-to-Image (T2I) generation (e.g., Stable Diffusion), by injecting the encoded target text prompt into the gradually denoised diffusion image generator. Despite the success of DPM in practice, the mechanism behind it remains to be explored. To fill this blank, we begin by examining the intermediate statuses during the gradual denoising generation process in DPM. The empirical

observations indicate, the shape of image is reconstructed after the first few denoising steps, and then the image is filled with details (e.g., texture). The phenomenon is because the low-frequency signal (shape relevant) of the noisy image is not corrupted until the final stage in the forward process (initial stage of generation) of adding noise in DPM. Inspired by the observations, we proceed to explore the influence of each token in the text prompt during the two stages. After a series of experiments of T2I generations conditioned on a set of text prompts. We conclude that in the earlier generation stage, the image is mostly decided by the special token `[\texttt{EOS}]` in the text prompt, and the information in the text prompt is already conveyed in this stage. After that, the diffusion model completes the details of generated images by information from themselves. Finally, we propose to apply this observation to accelerate the process of T2I generation by properly removing text guidance, which finally accelerates the sampling up to 25\%+.

七、分会报告 A4 专题：脑启发的神经网络模型、算法与应用（I）

A4-1 基于脉冲神经网络的类脑大模型

李国齐，中国科学院自动化研究所

摘要：摘要：类脑计算是受人脑信息处理机制启发，基于神经元和神经环路的结构和功能，以更通用人工智能为目标构建计算系统的技术总称。近年来脉冲神经网络在通用场景逼近传统深度学习的主流网络性能，展现出引领未来智能技术的潜力。本报告介绍脉冲神经网络的模型、算法及其硬件部署以及基于类脑脉冲神经网络大模型的科研进展

A4-2 脉冲神经动力学网络及硬件实现

邓磊，清华大学

摘要：脉冲神经动力学网络是重要的脑启发计算模型，是类脑计算领域的前沿课题，其丰富的

时空动力学被认为具有处理时序传感信号的巨大潜力。然而，目前对于脉冲神经动力学网络模型成分的深入理解较为匮乏，如何利用动力学特性处理时序计算任务获取高应用性能仍未充分解决。本次报告将系统性分析脉冲神经动力学网络的关键模型成分，进而构建集成树突时域异质性的新型脉冲神经动力学网络模型，揭示多尺度时域特征的融合机制，展示其在处理语音信号、视觉信号、脑电信号、机器人位置信号等宽频谱时序传感信号中的准确率和鲁棒性优势以及结合类脑计算芯片实现低功耗信息处理的效率优势，探讨该领域的机遇和挑战。

A4-3 类脑智能算法的思考及未来发展

张铁林，中国科学院脑科学与智能技术卓越创新中心

摘要：当前人工智能算法发展迅速，但是在能耗、鲁棒性、灵活性等方面受到挑战。本次报告将聚焦上述问题，介绍生物在脉冲编码、神经元类型、结构、可塑性等方面的机制机理，并启发相关类脑智能模型。进一步的，针对类脑算法本身的信息处理优势，探讨其在类脑芯片、侵入式脑机接口等方向的未来发展。

A4-4 脉冲视觉：建立机器视觉与生物视觉之间的桥梁

余肇飞，北京大学

摘要：人类获取的信息约 80% 源于视觉，视觉是生物大脑的基本功能，以视觉为主的感知是生物智能形成和发展的源头。脉冲视觉是机器视觉与生物视觉之间的桥梁，一方面利用脉冲/人工神经网络来学习和模拟视觉神经系统，建立类脑视觉模型，回答人类是如何看见世界的；另一方面借鉴生物视觉的计算机理来研究机器视觉，突破机器视觉的计算瓶颈，回答如何帮助机器看清世界。本报告将结合课题组相关工作汇报脉冲视觉方向的最新进展。

八、分会报告 A5 专题：机器学习与微分方程

A5-1 Deep adaptive sampling for numerical PDEs

周涛，中国科学院数学与系统科学研究院

摘要：We present a deep adaptive sampling method for solving PDEs where deep neural networks are utilized to approximate the solutions. More precisely, we propose the failure informed adaptive sampling for PINNs and an adaptive important sampling scheme for deep Ritz. Both approaches can adaptively refine the training

set with the goal of reducing the failure probability. Applications to both forward and inverse PDEs problems will be presented.

A5-2 Friedrichs 学习：通过深度学习求解偏微分方程弱解

黄建国，数学科学学院/上海交通大学

摘要：在本报告中，我们将提出和分析一个 Friedrichs 学习方法，它可通过将偏微分方程等价描述为一个 minmax 问题，然后使用深度学习方法，以数值求解该偏微分方程的弱解。名称“Friedrichs 学习”意在强调该方法与著名数学家 Friedrichs 发展的偏微分方程正对称组理论的密切联系。我们也将介绍该方法在数值求解多类偏微分方程定解问题的计算效果。本报告内容取自报告人与上海交大陈帆博士生，美国佛罗里达大学王春梅教授和美国马里兰大学帕克分校杨海钊教授的合作研究成果。

A5-3 KRnet 流模型密度函数估计方法及高维偏微分方程求解

廖奇峰，上海科技大学信息科学与技术学院

摘要：概率密度函数估计仍是计算科学与工程中的一个难题。通过耦合 Knothe-Rosenblatt (KR) 重排和基于流的生成模型，我们开发了一种可逆传输映射，称为 KRnet，用于高维密度估计。本报告对 Krnet 就行概述，并介绍其用于求解高维偏微分方程的自适应版本。

A5-4 探索扩散模型生成过程中的最佳动力学方程

曹语，自然科学研究院和数学科学学院/上海交通大学

摘要：扩散模型已在计算机视觉领域展现出显著成效，然而，基于常微分方程 (ODE) 的概率流与基于随机微分方程 (SDE) 的扩散模型之间，哪个更为优越以及其适用条件仍不明朗。这一比较之所以充满挑战，是因为它紧密关联着数据分布特性、评分函数训练效果以及其他复杂的数值因素。在本次报告中，我们计划通过探讨两个极限场景——即零扩散 (ODE) 状况与高扩散状态，来寻找解决这一问题的数学路径。我们将证明，在连续时间框架下，评分函数训练误差的时间分布特征，将决定最小化采样误差的动力学方程。这一理论分析的有效性在多种标准分布及真实数据集上得到验证。

九、分会报告 A6 专题：动理学方程与机器学习

A6-1 动理学方程基于神经网络的动态区域分解方法

熊涛, 中国科学技术大学数学科学学院

摘要: 基于动态区域分解的模型杂交算法能大大提高动理学玻尔兹曼方程的模拟效率, 但是目前基于矩阵特征值做指示子需要重构各阶导数且对人为参数较为敏感。我们基于全连接神经网络构造了一组新的指示子。数值上新的指示子有更好的表现。

A6-2 Solving Boltzmann equation using neural sparse representation

王艳莉, 北京计算科学研究中心

摘要: We consider the neural sparse representation to solve Boltzmann equation with BGK and quadratic collision model, where a network-based ansatz that can approximate the distribution function with extremely high efficiency is proposed. Precisely, fully connected neural networks are employed in the time and spatial space so as to avoid the discretization in space and time. Different low-rank representations are utilized in the microscopic velocity for the BGK and quadratic collision model, resulting in a significant reduction in the degree of freedom. We approximate the discrete velocity distribution in the BGK model using the canonical polyadic decomposition. For the quadratic collision model, a data-driven, SVD-based linear basis is built based on the BGK solution. All these will significantly improve the efficiency of the network when solving Boltzmann equation. Moreover, the specially designed adaptive-weight loss function is proposed with the strategies as multi-scale input and Maxwellian splitting applied to further enhance the approximation efficiency and speed up the learning process. Several numerical experiments, including 1D wave and Sod problems and 2D wave problem, demonstrate the effectiveness of these neural sparse representation methods. This is a joint work with Prof. Bin Dong and Dr. Zhengyi Li from Peking University.

A6-3 基于渐近保持格式的深度学习方法求解多尺度动理学方程

吴科科, 上海交通大学数学科学学院

摘要: 本报告以 PINN 求解时间依赖的多尺度线性输运方程为出发点, 发现使用原方程的残差作为损失函数的 PINN 方法由于神经网络的训练方式和潜在的频率偏好等原因只能表征方程的主导阶项或单一尺度表现, 从而无法有效地求解多尺度动理学方程。因此, 我们提出并发展了一系列渐近保持神经网络方法以有效地求解几类典型的多尺度动理学方程。

A6-4 DeepRTE: 辐射输运方程的预训练神经算子

马征, 上海交通大学

摘要: 在这项工作中, 我们提出了一种新颖的神经网络方法来求解稳定的辐射传递方程。辐射传输方程是一个微分积分方程, 描述介质中辐射的传输。它在中子传输、大气辐射传输、传热和光学成像等各个领域都有应用。这所提出的 DeepRTE 方法基于预先训练的基于注意力的神经网络, 能够高精度和高效地求解辐射传递方程。通过数值实验证明了所提出方法的有效性。

十、分会报告 A7 专题: Generative AI for Science

(I)

A7-1 几何深度学习在物质科学中的应用

黄文炳, 中国人民大学高瓴人工智能学院

摘要: 物质科学中的分子、蛋白质、抗体、晶体等原子系统均具有特定的空间几何结构, 构成了一类重要的数据形态——几何图。与社交网络中的拓扑图不同, 几何图中的节点或连边占据了一定的空间位置, 需要满足某些内蕴的物理特性, 比如对称性, 无法被传统图神经网络简单处理。近年来, 几何图神经网络由于有效嵌入了对称性 (不变性和等变性), 具有良好的解释性、泛化性和通用性, 在离散物理系统的表示与生成上得到了广泛应用。本报告将介绍几何图神经网络的定义与性质, 梳理国内外发展现状, 并且介绍团队在分子表示学习、晶体生成、抗体设计与优化等跨领域任务上的应用情况。

A7-2 数值算法驱动的生成式模型架构设计与参数微调

许洪腾, 中国人民大学

摘要：作为人工智能领域的重要分支，生成式模型旨在实现文本、图像等高维复杂数据的生成任务。由于其极高的学术价值和巨大的应用前景，生成式模型一直是人工智能学术界和产业界的研究前沿和关注焦点。在针对生成式模型的研究中，对模型的架构设计与参数微调一直占据着重要地位，并对生成式模型的发展和应用起到了关键作用。本次演讲旨在以生成式模型的函数估计能力（即模型表达能力）和训练难度（包括数值稳定性、计算复杂度等）为切入点，以数值分析理论和数值算法为核心，针对生成式模型中的关键模块设计新型高效的神经网络架构，并提出具有理论保证的、基于正交适配器的微调策略，使得模型在训练和微调过程中，在通用性和表达能力方面具有理论保证，同时具有更少的参数量和更低的计算复杂度，进而在图像生成、数学推理等复杂任务中取得更好的效果。

A7-3 Generative decoding for quantum error-correcting codes

张潘，中国科学院理论物理研究所

摘要：报告将介绍量子纠错码到统计力学，以及最大似然解码与统计力学自由能计算的映射，并介绍如何利用基于自回归神经网络的生成模型计算逻辑算符的条件概率以及最有可能的逻辑算符。

A7-4 基于扩散生成模型的物理系统设计和控制

吴泰霖，西湖大学

摘要：在科学和工程中，控制复杂物理系统的演化以及设计其高维参数是基本问题，广泛应用于流体力学、结构力学、水下机器人等任务。传统的设计和控制方法往往由于复杂的物理动力学而需要大量计算（例如伴随方法）；或者难以适应强非线性、强耦合的场景（例如 PID）。另一方面，最近基于深度学习（例如深度强化学习）的方法也面临对抗样本以及难以优化长期控制序列等问题。在这项工作中，我们提出基于扩散生成模型的物理系统的设计和控制方法。我们将物理系统的状态轨迹、设计参数以及控制序列看成一个联合向量，通过扩散模型从数据中学习其联合概率分布（通过能量函数表示）。在推理采样时，通过同时最小化能量函数加上给定的设计或控制的目标函数来生成新的样本。这样，生成的样本（包含设计参数、控制序列、系统状态轨迹）既符合物理，又能接近最优的设计和控制目标。此外，我们提出进一步的方法，使得能够生成比训练时更复杂的设计参数和摆脱训练集分布限制的控制序列。我们在一维和二维的偏微分方程系统中进行设计和控制的实验。在设计飞机翼型的实验中，尽管我们的模型只

用单个 2D 翼型与流体的交互数据训练，但其能够设计出比训练中更复杂的多个翼型及其编队以提高升阻比。在控制实验中，我们的方法揭示了“快速闭合-缓慢张开”是高效的水母运动模式，与流体动力学领域的已有发现相一致，相比起深度强化学习展现出独特优势。

十一、分会报告 A8 专题：机器学习与化学体系计算（I）

A8-1 机器学习量子蒙特卡洛在化学物理中的应用

陈基，北京大学物理学院

摘要：本报告讲主要介绍近期将基于神经网络的实空间波函数的应用于量子蒙特卡洛方法中，并对各类化学分子的电子基态和激发态性质进行高精度计算的工作进展。

A8-2 AI 集成指导高精度化学合成与低碳化工材料设计

王笑楠，清华大学

摘要：近年来，人工智能（AI）技术的飞速发展给化工、材料、能源等领域带来新的推动力。AI 集成指导的智能化学化工技术有望突破传统研发周期长、成本高、依赖经验的模式，赋能新的科学发现。本报告将介绍人工智能尤其是 AI 大模型在分子设计、化学合成、表征分析、机理发现等方面的进展，以单分子表面合成碳基量子材料、分离膜材料智造及工业催化剂的智能设计合成等典型体系为例分析 AI+的方法与应用实例。报告将介绍领域通用大模型的研究进展，并展望分析相关技术的潜力和优劣。

A8-3 AI 赋能激发态载流子动力学

褚维斌，复旦大学

摘要：激发态载流子动力学是能源和信息领域的核心问题，深入理解和调控它对于推动这些领域的发展至关重要。实时含时密度泛函理论和非绝热分子动力学方法是研究激发态载流子动力学的有力工具，但与传统的基态计算相比，这类方法的计算成本要高出几个数量级。近年来，机器学习在物理和材料科学领域得到了广泛应用，包括直接预测材料的各种性质、构建第一性原理精度的势能面模型、高通量生成分子和晶体结构以及开发更精确的交换关联泛函等。通过将机器学习与非绝热分子动力学方法相结合，我们有望在激发态载流子动力学模拟的尺度、精度和效率方面实现协同突破。在本报告中，我将介绍我们发展了一种新的基于图神经网络的非绝热分子动力学模拟框架，N²AMD。基于该框架我们可以实现大尺度杂化泛函非绝热分子动力学

模拟。N²AMD 不仅大大提高了非绝热分子动力学模拟的效率和精度，而且为超越现有非绝热分子动力学模拟的局限性铺平了道路。

A8-4 材料基础模型：从模拟到生成

陆子恒，微软研究院

摘要：Digital transformation is revolutionizing the field of materials science, with deep learning at the forefront of this paradigm shift. This presentation will unpack the latest advancements in large-scale machine learning tools and their transformative applications in the realm of materials design. We will spotlight the power of deep graph learning, [1] active learning[2], and generative models[3-4] in constructing advanced AI emulators capable of simulating, generating, and designing new materials. Active learning streamlines the learning process by prioritizing the most informative data, whereas graph learning exploits the complex relationships within material structures, and generative models push the envelope in predicting and creating unprecedented material structures. These state-of-the-art tools are not only enhancing our predictive capabilities but also accelerating the design of innovative materials with bespoke properties. We will explore how these AI-driven strategies are being applied to real-world material challenges, demonstrate their potential to catalyze breakthroughs across various sectors, and discuss the ongoing efforts and future directions for integrating deep learning into materials science to foster interdisciplinary collaborations and drive further advancements in this dynamic and critical field.

A8-5 GPU-accelerated Auxiliary-field quantum Monte Carlo with multi-Slater determinant trial states

吕定顺，字节跳动

摘要：AFQMC 作为一个高精度，高并行度的方法，目前在量子多体，量子化学方法得到了广泛的关注。我将汇报我们近期在研究，工程上的努力，进展以及对未来的一些展望。

十二、分会报告 A9 专题：机器学习在空天动力中的应用（I）

A9-1 基于数据融合的流场测量与控制

温新，机械与动力工程学院/上海交通大学

摘要：流动控制可以显著提升飞行器的气动性能，但是现有方法普遍存在控制效率较低和鲁棒性不强等问题，限制了工程应用。为了开发高效控制方法，依赖于精细化的实验测量方法，但是所产生的高频高速复杂流动也对高时空精度测量提出了挑战。本研究开发基于 PIV、PSP 压敏漆和热线风速仪等多种测量技术的数据融合测量方法，克服了数据缺失、采样频率过低以及信噪比低等难题。基于以上测量方法，揭示了振荡器工作机理，提出了振荡射流无舵面翼型环量控制和亚声速矢量喷管等气动控制方法，可将气源消耗显著降低。并采用数据融合方法将多种实验数据相结合，对振荡翼型等复杂流场进行实时感知和预测的方法。以上方法在小型涡喷发动机等真实工况下完成了验证，为无舵面飞行器控制提供技术支撑。

A9-2 人工智能赋能的发动机流场测量系统

韩旺，北京航空航天大学

摘要：推力室内流场的精准测量对高性能火箭发动机的研发与设计具有重要意义，对流场完整信息的获取并进行快速分析也有助于发动机研制的降本增效。然而推力室内流场处于高温高压受限空间环境，严重限制了各种测量手段的应用，使得目前的侵入式测量和光学测量手段都难以完整获得流场多参数的测量数据，面临推力室内流场测不到、喷管外流场测不全的问题。亟需发展基于外流场有限测量数据重构推力室内流场的方法及模型。目前，基于降阶模型方法的流场重构由于其线性假设难以应用于复杂的湍流问题，而基于数据同化的方法需要大量高精度的数值计算，这在现实当中是难以负担的。本报告将讲解课题组最近提出的稀疏数据辅助的生成式模型及其在流场重构方面的应用。

A9-3 深度学习技术与空天动力

安健，西北工业大学航天学院

摘要：介绍火箭机组合动力中深度学习方法的应用

A9-4 宽范围超声速燃烧基础研究

蔡尊，国防科技大学空天科学学院

摘要：针对宽范围超声速燃烧的喷注混合、点火、火焰稳定过程研究进展进行汇报。

十三、分会报告 B1 专题：人工智能的数学理论

B1-1 大模型训练和 Adam：原理探究和新方法 Adam-mini

孙若愚，香港中文大学（深圳）

摘要：Adam 是训练大型基础模型的默认算法。在这次演讲中，我们旨在理解为什么在训练大型基础模型方面，Adam 比 SGD（随机梯度下降）表现得更好，并提出了一个名为 Adam-mini 的内存效率更高的替代方案。首先，我们提供了 SGD 在 Transformer 上的失败的解释：(i) Transformers 是“异构的”：参数块之间的 Hessian 谱差异显著；(ii) 异构性阻碍了 SGD：在具有块异构性的问题上，SGD 的表现不佳。其次，受这一发现的启发，我们引入了 Adam-mini，它根据 Hessian 结构对参数进行分区，并对一个区块中的所有权重分配单一的第二动量项。我们通过实证展示了，在包括 7B 规模的语言模型和 ViT 在内的各种模型上，Adam-mini 在不牺牲性能的情况下，比 Adam 节省了 45-50% 的内存；多个独立研究者验证了 Adam-mini 的优越性，并在 LLaMa3-8B 上验证 Adam-mini 是少有的可超越 Adam 的优化器。

B1-2 Transformer 架构的运行机制研究：模型深度和多头注意力的作用

邹荻凡，香港大学

摘要：本报告深入探讨了 Transformer 架构的运行机制，重点关注了其深度和多头注意力在不同任务中的学习能力和局限性。在报告的第一部分，我们设计了一系列实践序列学习任务，系统地评估了不同深度的 Transformer 在记忆、推理、泛化和上下文泛化方面的性能及局限性。我们的研究结果显示，单层注意力的 Transformer 在记忆任务中表现卓越，但在处理更复杂任务时表现不佳。此外，我们发现至少需要两层的 Transformer 才能有效地实现推理和泛化能力，而上下文泛化能力则可能需要三层的 Transformer 来实现。在报告的第二部分，我们以稀疏线性回归问题为例，详细探讨了训练后的 Transformer 中多头注意力的作用，并揭示了多头注意力在不同 Transformer 层级的运行机制。我们的实验结果表明，在 Transformer 的第一层，每个注意力头都对最终性能至关重要，然而在后续层级，通常只有一个注意力头起主导作用。我

们进一步提出了“预处理-然后-优化”的运行机制，并从理论上证明了多层 Transformer（第一层多个头，后续层只有一个头）能有效实现此机制。此外，我们还证实了在稀疏线性回归问题中，这一机制相较于朴素梯度下降和岭回归算法具有优越性，这一发现与我们的实验结果相符。这些研究结果有助于我们深入理解多头注意力的优势和模型深度的作用，为我们揭示 Transformer 内部更复杂的机制提供了新的视角。

B1-3 大模型的推理与记忆

许志钦，上海交通大学自然科学研究院

摘要：关注现象驱动的深度学习的基础研究，将汇报频率原则、凝聚现象等常见现象。在凝聚现象的基础上，我们展示小初始化下，神经元的凝聚使 Transformer 网络倾向于推理，而大初始化，模型倾向于记忆。

B1-4 生成式 AI 的数学基础与分子设计应用

王宇光，自然科学研究院/上海交通大学

摘要：生成式 AI 在分子设计中发挥着至关重要的作用，可用于蛋白质设计和药物设计。我们开发了基于 500 亿高质量蛋白数据的蛋白质大模型并通过智能体连接下游 20 多个合成生物设计模块，实现了“Protein Design All in One”，从而设计出功能更为优越的新分子和蛋白质。我们的模型采用基于几何深度学习的扩散模型和基于序列的对话型大模型（GPT）两种生成模型。模型的基本模块，如 Transformer 和深度等变图神经网络，结合了贝叶斯统计、粒子方程、调和分析、学习理论，具有高度可解释性，并表现出更强的泛化能力和表示能力。

十四、分会报告 B3 专题：机器学习与交叉（I）

B3-1 基于优化的位错结构与演化的预测方法

戴书洋，武汉大学

摘要：位错是材料塑性形变的主要影响因素，其结构及相应的演化过程会很大程度上影响材料的力学性能，因此对位错结构与演化的快速预测是材料力学性能计算中的一个重要问题。在本次报告中，我们提出一种利用优化方法来预测位错结构与演化的方法，利用非奇异的应力形式给出位错体系能量并构建相应的高效算法。

B3-2 基于非均匀随机游走的自适应采样

胡丹, 自然科学研究院

摘要: 物理信息神经网络 (PINN) 在求解偏微分方程中得到了广泛的应用。对于解的正则性较低的问题, 在 PINNs 中引入自适应采样策略可以提高求解精度。我们提出了一种用于自适应采样的非均匀随机游动过程 (Nurwas)。在 Nurwas 中, 采样点在每个训练步进行一步非均匀随机游动。其中, 随机游走的步长与目标概率密度的平方根成反比。由于目标分布通常由数值解的残差给出, 因此我们可以在没有概率密度函数的显式表达式的情况下通过 Nurwas 进行采样。与此前的自适应采样策略相比, Nurwas 简化了采样过程, 降低了计算量, 并且避免了显式表达概率密度函数带来的困难和限制。

B3-3 Sparse discovery of differential equations based on multi-fidelity Gaussian process

邱越, 重庆大学数学与统计学院

摘要: Sparse identification of differential equations aims to compute the analytic expressions from the observed data explicitly. However, there exist two primary challenges. Firstly, it exhibits sensitivity to the noise in the observed data, particularly for the derivatives computations. Secondly, existing literature predominantly concentrates on single-fidelity (SF) data, which imposes limitations on its applicability due to the computational cost. In this talk, we present two novel approaches to address these problems from the view of uncertainty quantification. We construct a surrogate model employing the Gaussian process regression (GPR) to mitigate the effect of noise in the observed data, quantify its uncertainty, and ultimately recover the equations accurately. Subsequently, we exploit the multi-fidelity Gaussian processes (MFGP) to address scenarios involving multi-fidelity (MF), sparse, and noisy observed data. We demonstrate the robustness and effectiveness of our methodologies through several numerical experiments.

B3-4 pETNNs: Partial Evolutionary Tensor Neural Networks for Solving Time-dependent Partial Differential Equations

赵进, 首都师范大学 交叉科学研究院

摘要: In this talk, we will introduce our recent work for solving time-dependent partial differential equations with both of high accuracy and remarkable extrapolation, called partial evolutionary tensor neural networks (pETNNs). Our proposed architecture leverages the inherent accuracy of tensor neural networks, while incorporating evolutionary parameters that enable remarkable extrapolation capabilities. By adopting innovative parameter update strategies, the pETNNs achieve a significant reduction in computational cost while maintaining precision and robustness. Notably, the pETNNs enhance the accuracy of conventional evolutionary deep neural networks and empowers computational abilities to address highdimensional problems. Numerical experiments demonstrate the superior performance of the pETNNs in solving time-dependent complex equations, including the Navier-Stokes equations, high-dimensional heat equation, highdimensional transport equation and Korteweg-de Vries type equation.

B3-5 Unsupervised transfer learning via adversarial contrastive training

焦雨领, 武汉大学数学与统计学院

摘要: Learning data representation that can be used for downstream supervised learning tasks under unlabeled scenario is critical and challenging. In this paper, we propose a novel unsupervised transfer learning approach via adversarial contrastive training (ACT). Our experimental results achieve outstanding classification accuracy under both fine-tuned linear probe and k -NN protocol on various datasets, which are competitive compared to existing state-of-the-art self supervised learning methods. Moreover, we provide non-asymptotic error bound for the downstream classification task under miss-specified, over-parameterized setting to elucidate how a large amount of unlabeled samples contributes to our learning. Our theoretical results not only suggest that the classification error

of downstream task solely depends on the efficacy of data augmentation employed in ACT when the sample size is large enough, but also serve to bridge the gap in the theoretical understanding of the effectiveness of few-shot learning for downstream tasks.

十五、分会报告 B4 专题：脑启发的神经网络模型、算法与应用（II）

B4-1 神经网络中的具身概念形成、理解和交流

余山，中国科学院自动化研究所

摘要：人类智能的一个独特方面是从感觉运动经验中提取概念，同时能利用概念的认知反过来精确调控感觉运动系统以促进学习，并实现个体之间的交流。目前对于人脑如何实现这一机制，以及如何在人工神经网络中实现这些功能仍缺乏深入的了解。在此我们提出了一个神经网络模型，可以从执行视觉识别任务的经验中抽提概念，并利用所形成的概念空间来学习新技能并在不同系统间通过交流进行知识传递。本报告将介绍这一工作，并讨论其与人脑概念形成机制的关系，以及概念的符号化和交流对于新一代具身人工智能系统的意义。

B4-2 A Brain-inspired Computational Model for Spatio-temporal Sequence Recognition

弭元元，清华大学 心理与认知科学系

摘要：Temporal sequence processing is fundamental in brain cognitive functions. Experimental data has indicated that the representations of ordinal information and contents of temporal sequences are disentangled in the brain, but the neural mechanism underlying this disentanglement remains largely unclear. We investigate how recurrent neural circuits learn to represent the abstract order structure of temporal sequences, and how the disentangled representation of order structure facilitates the processing of temporal sequences. We show that with an appropriate training protocol, a recurrent neural circuit can learn tree-structured attractor

dynamics to encode the corresponding tree-structured orders of temporal sequences. This abstract temporal order template can then be bound with different contents, allowing for flexible and robust temporal sequence processing. Using a transfer learning task, we demonstrate that the reuse of a temporal order template facilitates the acquisition of new temporal sequences, if these sequences share the same or partial ordinal structure. Using a key-word spotting task, we demonstrate that the tree-structured attractor dynamics improves the robustness of temporal sequence discrimination, if the ordinal information is the key to differentiate these sequences.

B4-3 大脑精细模拟：神经科学与人工智能的新前沿

杜凯，北京大学人工智能研究院

摘要： 摘要：随着超级人工智能模型“ChatGPT”的成功，我们对智能的理解正发生变化。ChatGPT 依赖于一个包含近万亿参数的庞大人工神经网络，其智能似乎自动“涌现”。这引发了对智能如何从人类大脑中涌现的思考。人类大脑包含近 1000 亿神经元，分为数千种类型，每种都有独特的树突结构和复杂的离子通道，使得单个神经元的计算能力可与 5-8 层深度学习网络媲美。因此，生物神经元远非简单的“点模型”，而是高度复杂的网络。因此，我们的大脑可能包含一个参数比 ChatGPT 大 4-5 个数量级的超大型神经网络。大脑精细模拟 (Biophysically detailed simulation) 是唯一能够捕捉到树突结构、离子通道和突触复杂性的数学方法。然而，高计算成本严重限制了其在神经科学和人工智能领域的应用。在本次报告中，我将从树突计算研究的历史概述开始，阐明导致大脑精细模拟的理论基础。继此之后，我将探讨有关树突计算的当代理论研究及其在人工智能中的潜在应用，强调这种不断发展的理解如何正在重塑神经科学和人工智能范式。最后，我将介绍我们在基于 GPU 的模拟框架中的最新进展，该框架实现了相比传统基于 CPU 的 NEURON 模拟器高达 1500 倍的速度提升 (Zhang, et. al., Nat Commun, 2023)。这一技术进步为在前所未有的规模上模拟和训练生物真实的神经网络打开了大门，为实现模拟接近人类的生物智能提供了的可能。

B4-4 Dendritic Integration-Inspired Artificial Neural Networks Enhance Data Correlation

刘翀鸣, 上海交通大学数学科学学院

摘要: Incorporating biological neuronal properties into Artificial Neural Networks (ANNs) to enhance computational capabilities poses a formidable challenge in the field of deep learning. Inspired by recent findings indicating that dendrites adhere to quadratic integration rules for synaptic inputs, this study explores the computational benefits of quadratic neurons. We theoretically demonstrate that quadratic neurons inherently capture correlations within structured data, a feature that grants them superior generalization abilities over traditional neurons. This is substantiated by few-shot learning experiments. We further incorporate quadratic neuron into Convolutional Neural Networks (CNNs) with a biologically plausible approach, leading to novel CNN architectures—Dendritic Integration-inspired CNNs (Dit-CNNs). Our Dit-CNNs can compete favorably with state-of-the-art models across multiple classification benchmarks (CIFAR-10, CIFAR-100 and ImageNet-1K), while retaining the simplicity and efficiency of conventional CNNs.

十六、分会报告 B5 专题：反问题中的机器学习方法

B5-1 学习统计逆问题中先验分布的预测函数

贾骏雄, 西安交通大学

摘要: 偏微分方程 (PDE) 的统计逆问题可视为受 PDE 约束的回归问题。从这一视角出发, 我们借鉴概率近似正确 (Probably Approximately Correct, PAC) 贝叶斯学习理论的风格, 提出了一般化的泛化界, 用于学习无限维定义的先验度量。该理论框架在无限维可分函数空间上进行了严格定义, 这使得理论与常规的无限维贝叶斯逆问题方法紧密相连。受到差分隐私概念的启发, 我们提出了一种广义条件, 允许学习到的先验度量依赖于测量数据。在阐述了一般性理论后, 我们针对线性和非线性问题的具体设定进行了说明, 并能轻松地将这些特定情形纳入我们的一般性理论框架中, 以获得具体的泛化界。基于所得的泛化界, 我们构建了在无限维度上定义良好的实用算法。这些工作不仅推进了统计逆问题领域中的学习理论发展, 也为解决实际问题提供了理论指导和算法基础。

B5-2 数据辅助的两阶段方法解决反随机源散射问题

王玉亮, 北京师范大学

摘要: 我们提出了一种数据辅助的两阶段方法, 用于解决 Helmholtz 方程的反随机源散射问题。在第一阶段, 我们使用正则化的 Kaczmarz 方法来生成基于随机 Helmholtz 方程温和解的均值和方差的初始近似值。然后, 根据某种先验准则, 从近似值和相应的真实概况中抽取数据集。第二阶段被构造为一个图像到图像的翻译问题, 我们使用几种数据辅助的方法来处理数据集并获得增强的重构。数值实验表明, 数据辅助的两阶段方法对于具有较少实现的均匀和非均匀介质都提供了令人满意的重构。

B5-3 Physics-aware deep learning framework for the limited aperture inverse obstacle scattering problem

尹运文, 数学学院/东南大学

摘要: It is well known that traditional deep learning relies solely on data, which may limit its performance for the inverse problem when only indirect observation data and a physical model are available. A fundamental question arises in light of these limitations: is it possible to enable deep learning to work on inverse problems without labeled data and to be aware of what it is learning? This talk presents a deep decomposition method (DDM) for such purposes, which does not require ground truth labels. It accomplishes this by providing physical operators associated with the scattering model to the neural network architecture. Additionally, a deep learning based data completion scheme is implemented in DDM to prevent distorting the solution of the inverse problem for limited aperture data.

B5-4 最优运输理论的反问题应用

张一萱, 浙江大学数学科学学院

摘要: 近年来, 由最优运输问题定义的 Wasserstein 距离被广泛应用于数学, 物理学, 计算机科学等领域。该距离不易受到数据中高频噪声的影响, 并且能很好地比较数据之间的差异。本研究引入二次 Wasserstein 距离作为反问题优化形式中的度量函数, 以解决经典的电阻抗断层

成像 (EIT) 问题。我们将介绍 Wasserstein 距离及其梯度的高效计算方法, 从而构建该距离下求解 EIT 问题的优化框架。二维和三维的数值实验结果表明, 该方法能够提升计算的稳定性以及重构的精度, 实验上优于传统的正则化方法。

十七、分会报告 B6 专题: 科学机器学习方法及其在计算力学中的应用

B6-1 NSNO: 基于 Neumann 级数的 Helmholtz 方程神经算子方法

史作强, 清华大学数学中心

摘要: 在报告中, 我们将介绍一种基于 Neumann 级数的神经算子方法, 用于计算 Helmholtz 方程系数到解的映射。这种神经算子具有更好的泛化性和更高的精度。对于反散射问题也可以给出跟传统方法相当精度的重建结果, 同时计算时间大大缩短。

B6-2 人工智能加速超声 CT 影像重建

孙赫, 北京大学

摘要: 超声计算机断层扫描 (超声 CT) 是一种具有巨大临床应用潜力的无辐射人体组织成像新模式。不同于传统 B 超设备仅收集组织反射回波, 超声 CT 同时采集透射和反射声波, 因此能实现更高的成像分辨率。超声 CT 通常通过求解波动方程约束的反问题——全波形反演问题, 来重建组织的声速分布图。然而, 现有的全波形反演方法依赖于复杂的波动方程数值求解器, 难以满足医疗应用所需的准实时性能。本文介绍了一种新型的神经解算子方法——神经 Born 序列算子 (NBSO), 旨在提升大尺度波动方程模拟的精度和效率, 并基于此实现准实时超声 CT 影像重建。我们的方法在软组织和骨组织实验中均表现出极高的效率和准确性, 为超声 CT 的临床转化和广泛应用提供了新的解决方案。

B6-3 物理机理神经网络中的物理模型校准方法

孟旭辉, 数学与应用学科交叉研究院/华中科技大学

摘要: 物理机理神经网络在偏微分方程正问题和反问题求解中得到了较多应用, 尤其是反问题求解中存在无需划分网格和多次求解正问题导致的计算量偏大问题。然而, 在反问题求解时, 我们需要预先给定反问题对应的方程模型, 当方程模型指定出现错误时, 将会导致预测精度较

低。在此次报告中，我们通过采用在方程中引入额外的神经网络来纠正可能预设错误的模型，并基于贝叶斯方法量化模型预设错误导致的不确定性。此外，我们也可将此方法与符号回归结合，得到正确的控制方程或物理模型。

B6-4 计算力学的两种范式对比分析、融合与展望

张伟伟，西北工业大学 航空学院

摘要：人工智能的蓬勃发展为微分方程的求解和计算力学的发展带来了变革性的技术途径。以有限差分法、有限体积法和有限元法等为代表的传统计算力学范式，与以物理信息神经网络、极限学习机等为代表的智能范式各有优劣，他们的互相促进和深度融合，将为计算力学的发展带来了新的可能。报告深入对比分析了传统范式与智能范式各自的特点及优劣，以报告人团队近期的研究为例，归纳了两种范式互相促进的最新进展和典型案例，并探讨了两种范式融合的初步方案。报告最后将展望人工智能在工程计算领域更广泛的应用，涵盖大模型和智能流体力学计算软件等。

十八、分会报告 B7 专题：Generative AI for Science (II)

B7-1 通用、准确、生成式预训练分子大模型 Uni-Mol

汪鸿帅，北京深势科技有限公司

摘要：早期的分子表征方法是通过特征工程为分子创建“指纹”，利用化学和物理参数表征分子，以区分不同分子并预测其化学性质。三维分子表征能够精确地描述分子中原子的空间排列，保留了分子所有必要的信息，没有丢失任何关于分子结构的细节。对分子进行三维表征的工作量巨大，因为它涉及到大量的数据处理和复杂的计算。这不仅需要高性能的计算硬件来处理原子间相互作用的量子力学计算，还需要先进的机器学习算法来从这些数据中提取有用的信息和模式。近年来 AI for Science 新范式快速发展，各种底层要素快速涌现，使得 3D 分子表征成为可能。Uni-Mol 即是全球首个发表的 3D 分子表征大模型。Uni-Mol 是由深势科技于 2022 年 5 月发布的一款基于分子三维结构的通用大模型，论文被机器学习顶会 ICLR 2023 接收[1]。Uni-Mol 是分子世界的基座大模型，基于 Uni-Mol 可实现分子生成、性质预测等业界关注的

多种通用能力，覆盖众多行业。Uni-Mol 性能优越、模型泛化能力强，获得了大量学界和业界研究人员的广泛关注。我们在蛋白-药物构象生成[2]、单步逆合成设计[3]、MOF 碳捕捉材料性质预测[4]、OLED 发光性质预测[5]、药物毒理性预测[6]等众多场景，以 Uni-Mol 作为基座大模型进行了成功的应用，展现了强大的通用性，在不同场景中均能通过迁移和微调达到该场景的领域最优。Uni-Mol 大模型展现出的通用性构成了其在 AI for Science 领域中的巨大潜力。深势科技正在将 Uni-Mol 与产品功能深度融合，以用户手可及的交互形式，赋能药物开发、化学工程、材料设计等相关领域。我们预期 Uni-Mol 将被更多团队使用，作为通用大模型，对分子的世界进行探索。

B7-2 基于 AI 的晶体结构生成：算法及应用

陈品，中山大学计算机学院

摘要：近来 AlphaFold3 发布在生物学和药物设计领域引起了广泛的关注，并被认为是在蛋白质结构预测领域的重要突破。类似的，在材料领域，仅通过化学组分直接预测晶体结构一直是计算材料领域追求的目标之一，也出现相关计算软件如 CALYPSO、USPEX 等。在本报告中，汇报人借鉴了计算机视觉领域经过广泛验证的扩散模型方法，将其应用于材料结构预测问题，成功开发出了基于材料性质预测材料结构的 AI 模型。为了验证模型的有效性，汇报人基于有限的超导材料数据，成功设计出一批潜在具有较高超导转变温度 (T_c) 的超导材料。基于 HPC-AI 研究范式，构建了超过 1 千万的虚拟晶体结构数据库，用于新材料设计与探索。

B7-3 人工智能生成与泛化的最优传输理论与方法

孙剑，西安交通大学

摘要：最优传输主要关注以最小代价实现两个概率测度之间的迁移或对齐。作为一种应用数学工具，其理论、算法与高效计算一直以来是数学与机器学习所关注的重要研究方向。本报告将从人工智能生成模型与泛化能力角度，介绍最优传输的基本模型，计算方法与神经网络高效实现。进一步介绍我们所提出的保持关键点关系的最优传输模型、算法及其所引导的人工智能生成模型，并介绍用于提升人工智能跨域泛化能力的最优传输分布对齐模型与算法。最后总结与思考最优传输对大模型的基础作用与可能的未来研究方向。

B7-4 Con-CDVAE：晶体结构的条件生成方法

叶财渊，中国科学院物理研究所（标间，和刘家轩合住）

摘要：近年来，在使用生成机器学习模型生成新的晶体材料方面取得了一定的进展，但仍然不能很好地基于目标性质有效地生成晶体。本报告将介绍一种基于晶体扩散变分自编码器 (CDVAE) 开发的，用于条件晶体生成的 Con-CDVAE 模型。在 Con-CDVAE 中，我们通过引入新的模块，设计了合适的训练方法，使得模型具备条件生成的能力。我们对 Con-CDVAE 进行了多方面的测试，并利用期在近期的晶体集邮竞赛上取得了不错的成绩。最后我会简单介绍一下中国科学院物质科学数据中心。

十九、分会报告 B8 专题：机器学习与化学体系计算

(II)

B8-1 机器学习在非绝热动力学中的应用

兰峥岗, South China Normal University

摘要：我们将讨论机器学习在非绝热动力学中的应用，包括利用监督学习建立势能面加速非绝热动力学模拟；利用无监督学习分析动力学轨迹理解激发态分子运动；利用时间序列建模预测动力学演化。

B8-2 科研智能体助力理实交融的机器化学家探索

陈林江, 中国科学技术大学

摘要：传统的化学研究范式主要是依赖于“试错”手段，面对庞大的化学空间，配方和工艺的搜索常常止步于局部最优，带来了化学品和新材料的创制瓶颈。中国科学技术大学研发的机器化学家平台由化学数据库、算力大脑服务器、操作系统、精准操控的机器人与智能工作站组成，初步具备科学思维，覆盖了设计、制备、表征、测试、优化等科研全流程，在催化剂、发光分子、光学薄膜材料等的创制研究中实现了 3-5 个数量级的效率提升，且适用范围随着平台升级继续扩大，在国际上引领了数据智能驱动的化学研究新范式。本次汇报将展示大模型助力以及大模型“智能体”驱动的机器化学家实验探索实例。

B8-3 解析、数值、回归和智能融合驱动的动力学方程标度和分类理论

胡素磊, 合肥微尺度物质科学国家研究中心/中国科学技术大学

摘要：在本报告中，报告人通过解析推导、数值模拟、回归学习和智能优化融合方法，建立材

料生长动力学方程的标度和分类理论。通过此理论，我们可以实现借助于实验观察数据的微观原子尺度能量信息的提取，进而为构建基于实验和第一性原理数据的新科学模型打下基础。

B8-4 算法和算力驱动的第一性原理计算材料模拟

胡伟，中国科学技术大学

摘要：高性能计算是现代超级计算机上加速第一性原理密度函数理论计算材料模拟的有力工具。在本次报告中，我们介绍了平面波软件 PWDFT、原子轨道线性标度软件 HONPAS 和间断有限元低标度软件 DGDFT 的发展，所采用的先进高性能并行数值算法。PWDFT 是标准的赝势平面波高性能计算软件，采用了先进迭代对角化算法和的低秩分解近似算法，适合于千原子内高性能并行计算，支持 GPU-CUDA 和 DCU-HIP 加速计算，性能加速显著优于 VASP。HONPAS 采用严格局域的数值原子轨道基组，主要面向上千原子体系的杂化泛函线性标度计算，采用多级 MPI 动态并行，可扩展到数万核心。DGDFT 基于间断伽辽金方法，采用在自洽场迭代过程中动态产生的自适应局域基函数离散，其精度媲美于标准平面波基组，同时，DGDFT 方法采用两级并行化策略和异构加速，具有高度可扩展性，可以在新一代神威超级计算机使用近四千万核心并行计算，依然保持 60% 的并行效率，可以用于研究包含数百万原子复杂金属体系的电子结构性质。

二十、分会报告 B9 专题：机器学习在空天动力中的应用（II）

B9-1 结合第一性原理的 AI 技术及其在航空航天领域的创新应用

徐辉，上海交通大学

摘要：随着人工智能技术的飞速发展，其在航空航天领域的应用已成为推动行业创新的重要力量。本报告旨在探讨如何将第一性原理与 AI 技术相结合，以实现航空航天领域中复杂问题的深入理解和高效解决。报告围绕团队在基于谱物理神经网络及神经算子发展多尺度问题模拟方法，融合流形与物理知识的流场降阶模型开展跨声速抖振机理研究，基于第一性原理和数据驱动发展复杂系统强化学习控制框架及应用实践等相关工作，深入探讨第一性原理在提高模型泛化能力和解释性方面的关键作用，并展示其在航空航天领域中的广泛应用前景，从而为未来的技术创新提供坚实的理论基础和实践指导。

B9-2 固体推进剂分子模拟及其燃烧反应动力学研究

陈东平, 北京理工大学机电学院

摘要: 燃烧反应动力学是航空航天动力系统中重要的基本问题, 然而受限于固体推进剂本身的复杂性, 导致其燃烧反应动力学研究无法解决真实燃料体系中的核心问题。在本次报告中, 报告人将从固体推进剂分子模拟出发, 着重介绍近 6 年来我们课题组在固体推进剂等含能材料体系燃烧反应动力学方面的工作。报告人将以 HTPB 推进剂、NEPE 推进剂、双基推进剂、PBX 炸药体系的 DeepMD 力场开发为例, 展开讨论固体推进剂分子模拟的近期进展和未来发展方向。进一步基于含能材料分子模拟结果, 简要介绍自主原创的含能材料反应动力学建模技术 (EM-HyChem) 以及其初步应用成果。

B9-3 机器学习在燃烧振荡中的应用

韩啸, 航空发动机研究院/北京航空航天大学

摘要: Application of machine learning in combustion instability

二十一、分会报告 C1 专题: 深度学习理论与进展 (I)

C1-1 实例更优的加速梯度下降算法

方聪, 智能学院

摘要: 梯度下降算法在近年来被广泛用于求解机器学习模型。关于一阶算法的一项开创性的工作来源于 Yurii Nesterov 在 1985 年前后设计的基于动量的加速梯度下降算法。加速梯度下降算法对于最小化一般性满足梯度 Lipschitz 连续的凸函数能取得最优性。这里最优性的度量来自于满足条件的最难目标函数。然而在实际中, 机器学习问题对应的优化目标可能并没有那么困难。一个常见观测是目标函数的 Hessian 矩阵与数据相关, 其奇异值往往递减迅速。在这次报告, 我们研究设计对每个实例问题更优的算法。对于光滑的凸优化问题, 我们将提出一个新算法, 通过对空间分块优化, 实现实例更优的收敛速度。

C1-2 Learning with norm-based neural networks: model capacity and computational-statistical gaps

刘方辉, University of Warwick

摘要: In this talk, I will discuss some fundamental questions in modern machine

learning: - What is the suitable model capacity of over-parameterized models? -

What is the suitable function space for feature learning? - Which function can be learned by two-layer neural networks, statistical and/or computational efficiently? - What is the computational-statistical gap behind this? My talk will partly answer the above questions, both theoretically and empirically.

C1-3 深度学习中的随机矩阵最新进展

廖振宇, 华中科技大学

摘要: 深度学习已成为现代机器学习的基石。报告聚焦深度神经网络模型, 针对高维高斯混合数据, 给出了显式和隐式深度神经网络的定量刻画, 从而建立了显式和隐式网络、浅层和深度网络之间的显示联系。报告进一步介绍了高维统计和随机矩阵理论在深度学习理论中的最新进展。

C1-4 复杂学习系统中的多重下降现象

曹原, 香港大学

摘要: 最近的研究表明, 超参数化机器学习问题中存在双下降现象。尽管最近的一系列工作对这一现象进行了研究, 但双下降现象在理论上尚未被完全理解。在本文中, 我们研究了一类多成分预测模型中的多下降现象。首先, 我们考虑一个使用两种类型的随机特征的“双随机特征模型”(double random feature model, DRFM) 并研究了 DRFM 在岭回归中的预测风险。我们在高维框架下计算了 DRFM 的预测风险的精确极限, 并以此从理论上证明了 DRFM 的风险曲线可能呈现三重下降。我们还将研究扩展到“多随机特征模型”(multiple random feature model, MRFM), 并且证明了集成 K 种随机特征的 MRFM 可能呈现 K 重下降的现象。我们的分析指出, 在学习多成分预测模型时, 风险曲线通常存在特定数量的下降阶段。

二十二、分会报告 C2 专题: 机器学习与统计

C2-1 高维核回归的若干新现象

林乾, 清华大学

摘要: 我们简单介绍一下核回归在高维数据中的若干新结果, 并用它解释一些在神经网络中观察到的良性过拟合现象。

C2-2 A Statistical Theory of Regularization-Based Continual Learning

林伟, 北京大学数学科学学院

摘要: We provide a statistical analysis of regularization-based continual learning on a sequence of linear regression tasks, with emphasis on how different regularization terms affect the model performance. We first derive the convergence rate for the oracle estimator obtained as if all data were available simultaneously. Next, we consider a family of generalized ℓ_2 -regularization algorithms indexed by matrix-valued hyperparameters, which includes the minimum norm estimator and continual ridge regression as special cases. As more tasks are introduced, we derive an iterative update formula for the estimation error of generalized ℓ_2 -regularized estimators, from which we determine the hyperparameters resulting in the optimal algorithm. Interestingly, the choice of hyperparameters can effectively balance the trade-off between forward and backward knowledge transfer and adjust for data heterogeneity. Moreover, the estimation error of the optimal algorithm is derived explicitly, which is of the same order as that of the oracle estimator. In contrast, our lower bounds for the minimum norm estimator and continual ridge regression show their suboptimality. A byproduct of our theoretical analysis is the equivalence between early stopping and generalized ℓ_2 -regularization in continual learning, which may be of independent interest. Finally, we conduct experiments to complement our theory.

C2-3 尺度率的多阶段性质与极限

杨朋昆, 清华大学统计与数据科学系

摘要: 尺度率描述了模型的测试性能随着参数量、数据量、计算量的增加而提高的现象。有效的尺度率意味着可以通过扩大规模来解决机器学习问题。然而, 在许多实际应用中, 简单扩大规模的边际收益可能是递减的。本报告将基于最近邻方法, 讨论非参数方法尺度率的分阶段特点: 在第一阶段, 样本复杂度与数据维度多项式相关, 泛化误差快速下降; 而在第二阶段, 样本复杂度与数据维度指数相关, 泛化误差缓慢下降。我们研究表明, 实例的复杂度与分布的复

杂度对于确定泛化误差起到至关重要的作用。研究表明，实例的复杂度与分布的复杂度对于确定泛化误差起到至关重要的作用。

C2-4 Ternary Quantization for Distributed Mean Estimation

毛晓军，数学科学学院

摘要：The increasing size of data has created a pressing need for protection of communication and data privacy, spurring significant interest in quantization. This paper proposes a novel scheme for variance reduced correlated quantization that is designed for data with bounded support and distributed mean estimation. Our method achieves a theoretical reduction in mean square error for fixed and randomized designs compared to the correlated quantization method under different levels and dimensions scenarios. We conducted several synthetic data experiments to illustrate the effectiveness of our approach and to provide a good approximation of the reduced mean square error based on our theory. We also applied our proposed method to real-world data with different learning tasks, and it produced promising results.

二十三、分会报告 C3 专题：科学机器学习

C3-1 地球计算

白磊，上海人工智能实验室

摘要：理解并预测我们赖以生存的地球系统是人类自古至今的核心诉求之一。随着以卫星、雷达为代表的各种数据感知手段的日益丰富和大规模计算平台的完善，人工智能技术在处理大规模地球科学数据方面的优势逐渐显现，为气象预报、环境监测、应对气候变化等提供了全新的思路。本次报告将首先简要回顾分享人团队过去在人工智能技术方面的积累，然后介绍分享人近期基于人工智能方法在地球科学尤其是全球气象气候预报方面的研究进展，最后探讨人工智能技术在更多地球科学领域的交叉研究前景。

C3-2 之江实验室生物智能计算平台的特色应用

刘扶芮，之江实验室

摘要：以智能计算驱动重点领域的科研范式变革是之江实验室新一轮五年规划“1397”划定的重点任务之一。本报告将介绍生命科学计算研究中心的支撑设施-生物智能计算平台的建设背景与当前进展，并结合 AI4S 领域的具体场景，阐述该平台的特色算法和产业化应用。

C3-3 神经网络是否可以被严谨地解释清楚？以及可解释性技术在大模型上的应用落地

张拳石，电院 上海交通大学

摘要：虽然近年来神经网络的可解释性研究得到了广泛的关注，但是大部分可解释性研究依然停留在工程技术层面，大量根本性问题尚未得到解决，尚缺少相对严谨的理论体系从根本机理层面统一解释神经网络的知识表达和其表征性能。比如，证明神经网络内在决策逻辑是否可以被严谨地解释为符号化的概念，如何量化神经网络的知识表征，什么是决定神经网络泛化性和鲁棒性的第一性原理，等等。本次报告将介绍如何在博弈交互理论体系下严谨地量化神经网络所建模的概念表征，如何证明解释的严谨性，如何通过概念表征层面的解释提升大模型的训练效率节省成本，如何对大模型安全性进行量化评估。

C3-4 大模型中的复杂推理

刘鹏飞，上海交通大学-电院

摘要：大模型复杂推理能力的提升是人工智能领域的一个重要研究方向。尽管大语言模型在许多任务上表现出色，但在涉及复杂推理的领域，如高等数学问题解决，它们的表现仍有待提高。本报告将探讨提升大模型复杂推理能力的关键策略，包括预训练数据的优化、监督式微调技术的改进、奖励模型的设计以及推理过程的评估方法。我们将讨论当前研究的最新进展，分析存在的挑战，并探讨未来可能的研究方向。

二十四、分会报告 C4 专题：统计机器学习与 AI4Math

C4-1 AI for Permutation: Trainable Sparse Optimal Transport

邱怡轩，上海财经大学

摘要：Learning permutations is a classical and fundamental task in statistics and machine learning, and is closely related to sorting, ranking, and selection problems. However, permutation matrices are discrete by nature, making it hard to embed trainable permutations in end-to-end and gradient-based learning frameworks.

This issue also greatly hinders contemporary artificial intelligence to deal with discrete and structural problems. More recently, the close connections between optimal transport (OT) and permutations have received massive attentions in the machine learning community, and various differentiable soft-permutation operators have been proposed based on the entropic-regularized OT formulation. However, the approximate permutation matrices learned from these methods are completely dense, thus losing the selection functionality of a genuine permutation operator. In this work, we consider the sparse OT problem based on the quadratic regularization, and study its differentiation properties. We show that although the learned approximate permutation is not smooth, it possesses Bouligand derivatives as a generalization to the gradient. We also derive a closed-form expression for the Bouligand derivative of the learned permutation, and implement it as a deep learning layer that supports automatic differentiation. Various numerical experiments are conducted to illustrate the effectiveness of such trainable permutations in machine learning tasks such as neural network pruning and neural-network-based kNN classifiers.

C4-2 Constrained Policy Optimization with Explicit Behavior Density (CPED) for Offline Reinforcement Learning

王文佳, 香港科技大学 (广州) / 香港科大霍英东研究院

摘要: Due to the inability to interact with the environment, offline reinforcement learning (RL) methods face the challenge of estimating the Out-of-Distribution (OOD) points. Existing methods for addressing this issue either control policy to exclude the OOD action or make the Q function pessimistic. However, these methods can be overly conservative or fail to identify OOD areas accurately. To overcome this problem, we propose a Constrained Policy optimization with Explicit Behavior density (CPED) method that utilizes a flow-GAN model to explicitly estimate the density of behavior policy. By estimating the explicit density, CPED can accurately identify the safe region and enable optimization within the region,

resulting in less conservative learning policies. We further provide theoretical results for both the flow-GAN estimator and performance guarantee for CPED by showing that CPED can find the optimal Q -function value. Empirically, CPED outperforms existing alternatives on various standard offline reinforcement learning tasks, yielding higher expected returns.

C4-3 图神经网络的泛化理论分析

刘勇，高瓴人工智能学院/中国人民大学

摘要：近年来，图神经网络受到大家广泛关注，在算法层面得到了长足的进展，然而关于图神经网络的理论分析相对较少，尤其对于其泛化分析基本还处于刚起步的阶段。本报告从转导学习的角度研究图神经网络的泛化性，并详细对比分析了几种常见图神经网络模型的泛化能力，为图神经网络新算法的设计提供了一定的理论指导。

C4-4 大语言模型的数学推理能力

刘征瀛，华为诺亚

摘要：大语言模型在众多任务和浅层推理上已经达到很不错的性能，但在长链条推理任务如数学推理和各种符号推理上仍有巨大的改进空间。本次报告将介绍近期我们在使用大语言模型进行符号推理任务比如定理证明、数学应用题求解的一些尝试，以及介绍我们提出的“硬化”的概念和相关的工作，有助于理解大语言模型的可解释性，提升推理效率及泛化性能。

二十五、分会报告 C5 专题：数据科学中的随机模型与算法（I）

C5-1 基于随机梯度重构的地震波成像方法

柴利慧，中山大学

摘要：我们提出了基于随机批次方法的梯度重构策略，用于地震波成像的反演迭代过程中。该方法有机的结合了诸如冻结高斯逼近等渐进方法与机器学习中的随机梯度下降方法的思想，能够充分发挥数值方法的并行性能，为求解高维高精度的反演成像问题提供了工具。同时，我们

从理论上证明了该方法与确定性的梯度重构方法具有相同的收敛速率。该项工作是与胡奕啸博士、黄忠亿教授、杨旭教授共同完成的。

C5-2 Graph Fourier Transforms on Directed Graphs

成诚, 数学学院/中山大学

摘要: Graph signal processing provides an innovative framework to process data on graphs. The widely used graph Fourier transform on the undirected graph is based on the eigen-decomposition of the Laplacian. In many engineering applications, the data is time-varying and pairwise interactions among agents of a network are not always mutual and equitable, such as the interaction data on a social network. Then the graph Fourier transform on directed graph is in demand and it should be designed to reflect the spectral characteristic for different directions, decompose graph signals into different frequency components, and to efficiently represent the graph signal by different modes of variation. In this talk, I will present our recent work on the graph Fourier transforms on directed graphs which are based on the singular value decompositions of the Laplacians.

C5-3 Uncertainty Quantification in Scientific Machine Learning

郭玲, 数学系/上海师范大学

摘要: Neural networks (NNs) are currently changing the computational paradigm on how to combine data with mathematical laws in physics and engineering in a profound way, tackling challenging inverse and ill-posed problems not solvable with traditional methods. However, quantifying errors and uncertainties in NN-based inference is more complicated than in traditional methods. In this talk, we will present a comprehensive framework that includes uncertainty modeling, new and existing solution methods, as well as Information bottleneck based uncertainty quantification for neural function regression and neural operator learning.

C5-4 针对奇异摄动问题的算子学习

黄忠亿, 清华大学数学科学系

摘要: Generalization of DeepONets for Learning Operators Arising from a Class of Singularly Perturbed Problems

二十六、分会报告 C6 专题: AI for Imaging

C6-1 医学 PET 成像与仪器研发

胡战利, 中国科学院深圳先进技术研究院

摘要: 医学成像技术是借助于某种能量与生物体进行相互作用, 为生物组织研究和临床诊断提供影像信息的一门科学。它涉及的范围包括超声成像、CT 成像、磁共振成像和 PET 成像等。PET 成像作为核医学尖端技术的代表, 已经在肿瘤、心血管和神经等疾病早期诊断中发挥了重要作用。其中, PET/CT 和 PET/MR 是目前临床中最先进的多模态成像设备之一。本报告主要介绍课题组在医学 PET 智能成像与科研仪器研发方面所开展的一些工作。

C6-2 智能锥束 CT 成像

牛田野, 深圳湾实验室

摘要: 锥束 CT 成像可提供生物医学组织成分和结构定量信息, 尚存组织成分定量辨识难、动态定量成像功能弱、组织器官图像量化难等问题。申请人在前期工作中针对上述问题: ①提出了稀疏性原理低对比度分辨方法, 设计了基于空间低频分布特性的散射修正方法, 解决了锥束 CT 组织成分定量辨识难题; ②研发了动态呼吸下单次扫描四维成像方案, 提高了动态定量分辨能力; ③提出了多特征组学鲁棒计算新方法, 实现了锥束 CT 组织器官图像精准量化, 提高了临床应用准确率; ④自主研发了多款定量成像设备, 其中太空锥束 CT 入选中国国际航空航天博览会医疗器械展。近五年, 以一作/通讯发表 TMI、TCI、CCR 等期刊论文 30 篇, 其中 JCR 一区论文 12 篇, 他引 463 次, 授权发明专利 13 项, 主持重点研发计划等国家级项目 4 项, 获中国体视学会技术发明二等奖(序 1)。

C6-3 图像恢复的无监督学习算法

庞彤瑶, 清华大学

摘要: 图像恢复是指从退化或有限的测量中恢复高质量图像, 这在科学和医学等多个领域有着广泛的应用。近年来, 深度学习已经成为解决包括图像恢复在内的许多问题的一个重要工具。大多数深度学习方法是监督学习, 这需要大量成对的训练数据, 包括真实图像。在本次报告中,

我将介绍几种无监督方法，这些方法仅不需要成对的训练数据，同时仍表现出与监督学习相当的性能，在收集干净的图像和建立高质量的训练数据集非常困难的实际复杂图像恢复任务中具有交大的潜力。

C6-4 Generative Networks for Inverse Problems without Ground-Truth Data

Oscar Leong, University of California, Los Angeles

摘要: We consider solving ill-posed imaging inverse problems under a generic forward model. Because of the ill-posedness present in such problems, prior models that encourage certain image-based structure are required to reduce the space of possible images when finding a solution. Many approaches based on machine learning try to learn the underlying image generation model given samples from the data distribution of interest, and use this to solve a constrained inverse problem; however, in many applications, ground-truth images may be unavailable. In contrast, we propose to either select or learn an image generation model from noisy measurements alone, without incorporating prior constraints on image structure. We first show how, given a collection of candidate models, the Evidence Lower Bound (ELBO) of a variational distribution can be used to select an appropriate prior. Then, we showcase how in the absence of available priors, one can directly learn the underlying image model from a set of noisy measurements using the ELBO. We assume crucially that the ground-truth images share common, low-dimensional structure. The learned model leverages this structure in its architecture, which consists of a shared generator with a compressed latent space where each measurement posterior is learned variationally. This allows the model to learn global properties of the data distribution from noisy observations without overfitting. We illustrate our framework on a variety of inverse problems, ranging from denoising to compressed sensing problems inspired by black-hole imaging.

C6-5 基于 AI 的冷冻电镜图像数据处理技术

张法, 医学技术学院 / 北京理工大学

摘要：冷冻电镜三维重构已成为当前解析生物大分子三维结构的第一选择。然而，冷冻电镜图像数据存在高噪声、信息缺失等特征，随着数据规模和尺度的持续增大，冷冻电镜图像数据的高效处理已成为当前结构生物学和计算机科学具有挑战性的难点问题之一。本报告将从计算的角度介绍冷冻电镜三维重构存在的挑战性问题，并着重介绍我们在冷冻电镜图像数据降噪、重构和柔性结构动力学分析等方面的最新研究成果。

二十七、分会报告 C7 专题：深度学习求解复杂流体的高效算法及分析（I）

C7-1 求解偏微分方程的随机神经网络方法

王飞，数学与统计学院/西安交通大学

摘要：传统数值方法在处理高维问题、复杂区域划分以及时间迭代所引起的误差累积方面面临着许多挑战。与此同时，基于优化训练的神经网络方法由于缺乏高效的优化算法，导致精度不足、训练速度缓慢，而且难以控制误差。为了综合利用这两种方法的优点并弥补它们的不足，本报告探讨了将随机神经网络方法与传统数值格式相结合来求解各种偏微分方程的方法。该方法不仅充分利用了神经网络的强大逼近能力，避免了经典数值方法的局限，还能够解决神经网络在精度和训练效率方面的问题。数值算例表明，随机神经网络方法能够以较少的自由度获得高精度的解，并且能够通过时空一体的数值格式高效地求解时间依赖问题，从而避免了时间离散格式引发的误差累积。进一步地，该方法还展示了求解高维问题的潜力。

C7-2 稀疏观测数据下深度神经算子学习重建高温热化学非平衡流场

洪启臻，中国科学院力学研究所

摘要：The hypersonic flow is in a thermochemical nonequilibrium state due to the high-temperature caused by the strong shock compression. In a thermochemical nonequilibrium flow, the distribution of molecular internal energy levels strongly deviates from the equilibrium distribution (i.e., the Boltzmann distribution). It is intractable to directly obtain the microscopic nonequilibrium distribution from existed experimental measurements usually described by macroscopic field variables

such as temperature or velocity. Motivated by the idea of deep multi-scale multi-physics neural network (DeepMMNet), we develop in this paper a data assimilation framework called DeepStSNet to accurately reconstruct the quantum state-resolved thermochemical nonequilibrium flowfield by using sparse experimental measurements of vibrational temperature and pre-trained deep neural operator networks (DeepONets). In particular, we first construct several DeepONets to express the coupled dynamics between field variables in the thermochemical nonequilibrium flow and to approximate the state-to-state (StS) approach, which traces the variation of each vibrational level of molecule accurately. These proposed DeepONets are then trained by using the numerical simulation data, and would later be served as building blocks for the DeepStSNet. We demonstrate the effectiveness and accuracy of DeepONets with different test cases showing that the density and energy of vibrational groups as well as the temperature and velocity fields are predicted with high accuracy. We then extend the architectures of DeepMMNet by considering a simplified thermochemical nonequilibrium model, i.e., the 2T model, showing that the entire thermochemical nonequilibrium flowfield is well predicted by using scattered measurements of full or even partial field variables. We next consider a more accurate and complex thermochemical nonequilibrium model, i.e., the StS-CGM model, and develop a DeepStSNet for this model. In this case, we employ the coarse-grained method, which divides the vibrational levels into groups (vibrational bins), to alleviate the computational cost for the StS approach in order to achieve a fast but reliable prediction with DeepStSNet. We test the present DeepStSNet framework with sparse numerical simulation data showing that the predictions are in excellent agreement with the reference data for test cases. We further employ the DeepStSNet to assimilate a few experimental measurements of vibrational temperature obtained from the shock tube experiment, and the detailed non-Boltzmann vibrational distribution of molecule oxygen is reconstructed by using the sparse experimental data for the first time. Moreover, by considering

the inevitable uncertainty in the experimental data, an average strategy in the predicting procedure is proposed to obtain the most probable predicted fields. The present DeepStSNet is general and robust, and can be applied to build a bridge from sparse measurements of macroscopic field variables to a microscopic quantum state-resolved flowfield. This kind of reconstruction is beneficial for exploiting the experimental measurements and uncovering the hidden physicochemical processes in hypersonic flows.

C7-3 Deep adaptive basis Galerkin method for evolution equations

顾亦奇, 电子科技大学数学科学学院

摘要: Neural network-based numerical methods for differential equations have been widely developed in recent years. In this work, we study deep neural networks (DNNs) for solving high-dimensional evolution equations with oscillatory solutions. Unlike other existing methods (e.g. PINNs) that deal with time and space variables simultaneously, we propose a deep adaptive basis Galerkin (DABG) method which adopts the spectral-Galerkin method for time variables and the network-based method for high-dimensional space variables. We also establish estimates of the solution error. Numerical examples, including high-dimensional linear parabolic and hyperbolic equations, and a nonlinear Allen-Cahn equation are presented to demonstrate the advantage of the DABG method over other existing methods.

C7-4 基于机器学习的流动建模和控制

谢芳芳, 浙江大学航空航天学院

摘要: 随着科技的不断发展, 机器学习技术逐渐成为各个领域的热门话题。传统的流动建模和控制方法往往依赖于经验和试错, 而机器学习技术则为飞行器设计提供了更加快捷、高效的方法。本报告首先基于机器学习理论, 充分利用流动系统中已有的仿真数据和实验数据, 提出了一系列流动降阶建模方法与多源/多精度气动数据融合方法, 实现对跨声速流场及气动特性的快速精确预示和非定常流动问题的参数优化, 为发展飞行器流动系统数字孪生模型提供技术支撑。其次, 针对实际工业应用中常见的流动减阻问题, 基于专家演示强化学习和 transformer

架构的元模型构建了一个训练时间少、样本需求小、具有泛化能力的无模型自适应流动控制智能体，避免了复杂流动控制系统的建模过程，为建立新一代飞行器智能流动控制系统打下了基础。

二十八、分会报告 C8 专题：科学智能中的高性能计算

C8-1 HPC+AI 驱动的微观尺度模拟初探

贾伟乐，中国科学院计算技术研究所

摘要：从原子尺度出发的微观科学计算面临着计算方法和效率的挑战。科学智能(AI for Science)的发展，尤其是高性能计算与人工智能方法的结合为其注入了新的活力。本报告将汇报我们组在相关领域的一些工作，包括利用高性能计算用于高精度的数据产生、科学智能的大规模并行训练和 HPC+AI 在第一性原理精度分子动力学大规模模拟的工作。最终总结这个领域目前面临的挑战。

C8-2 大模型高效可扩展并行策略

李士刚，北京邮电大学

摘要：大模型的并行策略越来越复杂，这给系统开发及性能优化人员带来沉重负担。为应对上述挑战，提出一种深度学习自动分布式并行框架 AutoDDL。相比已有框架，AutoDDL 可以描述更高维度并行算法，从而拓宽了并行策略空间，可实现端到端通信最优并行策略的自动搜索，使大模型获得更好的并行可扩展性。针对大模型流水线并行，提出一种融合二阶优化方法的流水线并行方案 PipeFisher。该方案通过在流水线空泡中自动填充二阶计算负载，提升模型收敛速率及 GPU 硬件利用率，可显著降低大模型端到端训练时间。

C8-3 精度可复现模型弹性训练技术

杨海龙，北京航空航天大学

摘要：随时深度学习模型参数规模不断增长，使用云资源按需付费成为更加经济的模型训练方式。为了应对云资源的动态性，弹性训练技术可以使深度学习训练任务在资源弹性伸缩的情况下继续其训练过程，并通过高效的调度方案提高训练吞吐量。然而，现有的弹性训练技术在使用不同规模资源训练时，会出现模型精度训练难以复现的问题，并且无法高效利用异构资源提升训练吞吐量。本报告重点探讨精度可复现模型弹性训练技术，特别是训练过程与资源分配解

耦，精度不确定性溯源，以及异构资源任务调度策略等，在确保模型在弹性训练中精度可复现的前提下，高效利用异构计算资源，提高模型训练性能。

C8-4 光电混合的高性能网络互连

赵世振，John Hopcroft 计算机科学中心/上海交通大学

摘要：随着 AI 大模型的兴起，大规模分布式训练变得愈发重要。然而，随着摩尔定律的失效，传统基于电交换的数据中心网络架构越来越难以满足大模型训练对带宽的需求，导致网络逐渐成为分布式训练的瓶颈。为了绕过电交换芯片的限制，我们提出用光交换机替代大容量电交换机，构建光电混合数据中心网络。该报告将从拓扑、路由、网络协议等方面探讨光电混合组网的挑战与解决思路，旨在构建出可落地的光电混合网络，消除 AI 大模型的通信瓶颈。

C8-5 融合气象大模型的高性能资料同化系统

琚锡廷，清华大学

摘要：大气资料同化是数值天气预报中的关键技术。集合资料同化的精度依赖于大量的模式成员运算，因此被认为是当前一个具有挑战性的计算任务。基于传统算法实现大范围、高分辨率和更多集合成员的快速资料同化仍面临计算挑战，限制了现有算法和应用的探索空间。而人工智能天气模型发展迅速，在预报精准度逼近再分析资料内蕴模式的同时，计算性能数量级提升，表现出了接近甚至超越传统数值模式的能力，对此我们构建了一个融合气象大模型的高性能集合资料同化系统。

二十九、分会报告 C9 专题：机器学习与最优化方法（I）

C9-1 组合优化问题的机器学习求解方法

严骏驰，上海交通大学

摘要：本报告将介绍项目组在组合优化机器学习求解方面的研究，覆盖 AI 辅助的问题求解、AI 原生的问题求解，以及在 EDA 等设计任务的应用。最后也对后续研究做出简单展望。

C9-2 基于大语言模型的云计算调度方法设计

王祥丰，计算机科学与技术学院/华东师范大学

摘要：云计算运筹优化是运筹学与计算机科学技术深度交叉融合的研究方向。云计算运筹优化

效率不仅仅关系到云计算服务的使用体验，而且是云计算提供商降本增效的核心。传统运筹学方法已经被广泛应用到云计算运筹优化中，而随着以大模型为代表的新一代人工智能技术的出现，对云计算运筹优化技术提出了更高的挑战。大模型技术也为云计算运筹优化提供了新的可能，促使我们利用大模型进一步设计新型云计算运筹优化方法。这个报告将给大家介绍大模型与云计算运筹优化如何深度融合。

C9-3 On Representing (Mixed-Integer) Linear Programs by Graph Neural Networks

陈子昂, Massachusetts Institute of Technology

摘要: While Mixed-integer linear programming (MILP) is NP-hard in general, practical MILP has received roughly 100-fold speedup in the past twenty years. Still, many classes of MILPs quickly become unsolvable as their sizes increase, motivating researchers to seek new acceleration techniques for MILPs. With deep learning, they have obtained strong empirical results, and many results were obtained by applying graph neural networks (GNNs) to making decisions in various stages of MILP solution processes. We study the theoretical foundation and discover a fundamental limitation: there exist feasible and infeasible MILPs that all GNNs will, however, treat equally, indicating GNN's lacking power to express general MILPs. Then we show that linear programs (LPs) without integer constraints do not suffer from this limitation and that, by restricting the MILPs to unfoldable ones or by adding random features, there exist GNNs that can reliably predict MILP feasibility, optimal objective values, and optimal solutions up to prescribed precision. We also show that second-order GNNs can represent branching strategies for MILPs and extend some results to quadratic programs. Small-scale numerical experiments are conducted to validate our theoretical findings.

C9-4 An Effective AI-Driven Algorithm for Decentralized Optimization

袁坤, 北京大学

摘要: Most decentralized optimization algorithms are handcrafted. While endowed with strong theoretical guarantees, these algorithms generally target a broad

class of problems, thereby not being adaptive or customized to specific problem features. This talk discusses data-driven decentralized algorithms trained to exploit problem features to boost convergence. Existing learning-to-optimize methods typically suffer from poor generalization or prohibitively vast search spaces. In addition, they face more challenges in decentralized settings where nodes must reach consensus through neighborhood communications without global information. To resolve these challenges, this paper first derives the necessary conditions that successful decentralized algorithmic rules need to satisfy to achieve both optimality and consensus. Based on these conditions, we propose a novel mathematics-inspired learning-to-optimize framework for decentralized optimization. Empirical results demonstrate that our learned algorithms outperform handcrafted algorithms and exhibit strong generalizations. Algorithms trained with 100 iterations perform robustly when running 100,000 iterations during inferences. Moreover, our algorithms trained with synthetic datasets perform well on problems involving real data, higher dimensions, and different loss functions.

三十、分会报告 D1 专题：深度学习理论与进展（II）

D1-1 分析随机特征模型和两层神经网络的对偶框架

龙吉昊，上海算法创新研究院

摘要： We consider the problem of learning functions in the F_p and Barron spaces, which are relevant for understanding random feature models (RFMs), two-layer neural networks, and kernel methods. Through a duality analysis, we reveal an equivalence between the approximation and estimation for learning functions in the two spaces. This allows us to focus on the easier task between approximation and estimation when examining the learnability of these function spaces. Furthermore, we demonstrate the flexibility and versatility of our duality framework through comprehensive analyses of two specific applications: 1. Random feature learning beyond the kernel regime. We prove that RFMs can learn functions in the F_p space

without the curse of dimensionality as long as $p > 1$. This result implies that RFMs can work well beyond the kernel regime since the F_p space is strictly larger than the associated reproducing kernel Hilbert space (RKHS) when $p < 2$. 2. The L -infty learnability of RKHSs. By leveraging the duality principle, we establish both lower and upper bounds for learning functions in a RKHS under the L -infty norm, linking the L -infty learnability to the eigenvalue decay of the associated kernel. We then apply these bounds to dot-product kernels and identify conditions under which the learning either suffers from or overcomes the curse of dimensionality. In particular, these results imply that learning with (random) ReLU features is generally intractable under the L -infty norm. To establish the aforementioned dual equivalence, we introduce an information-based complexity measure. We show that this complexity can effectively control minimax estimation errors in various settings, which might be of independent interest.

D1-2 矩阵信息论在人工智能中的应用

袁洋，清华大学交叉信息研究院

摘要：矩阵信息理论是比传统信息论更一般的理论，它将经典的信息论概念，如熵、KL 散度和交叉熵等拓展到了矩阵版本。这种更一般的视角帮助我们提出了一个结合了均匀性损失和对齐损失的统一框架，适用于对比学习和非对比学习方法。在此框架内，我们提出了 Matrix-SSL，这是一种在 ImageNet 和 MS-COCO 数据集上超越了先前方法的新方法。此外，我们证明了两种双分支自监督学习方法（Barlow Twins 和 Spectral contrastive learning）的损失函数在优化过程中，也同时优化了矩阵互信息和矩阵联合熵。这一见解引出了 M-MAE，这是一种利用基于矩阵的熵估计作为正则化的新方法，实验证明其比 U-MAE 更有效。

D1-3 Understanding Implicit Bias in Input Spaces

张景昭，清华大学

摘要：One explanation for the strong generalization ability of neural networks is implicit bias. Yet, the definition and understanding of implicit bias in non-linear contexts remains mysterious. In this work, we propose one form of implicit

bias, characterized by the count of connected regions in the input space with the same predicted label. Compared against parameter-dependent biases (e.g., norm or normalized margin), region count can be better adapted to nonlinear, overparameterized models, because the count value is determined by the function mapping and is invariant to reparametrization. Empirically, we found that small region counts align with geometrically simple decision boundaries and empirically correlate with good generalization performance. We also observe that good hyperparameter choices such as larger learning rates and smaller batch sizes can induce small region counts. We further establish the theoretical connections between region count and the generalization bound, and explain how gradient descent can induce small region count in nonlinear neural networks.

D1-4 两层神经网络全局最小值的几何性质与局部恢复

罗涛, 上海交通大学数学科学学院

摘要: 在宽松假设下, 我们研究两层神经网络在全局最小值附近的损失景观的几何结构。通过利用新技术, 我们展示: (i) 随着样本量的增加, 具有零泛化误差的全局最小值如何在几何上与其他全局最小值分离; 以及 (ii) 梯度流动力学的局部收敛性质和速度。我们的结果表明, 在过参数化条件下, 两层神经网络可以在局部恢复。

三十一、分会报告 D2 专题: Recent Advances on Learning to Optimize

D2-1 可靠成像的深度学习逆问题求解

戴文睿, 上海交通大学电子信息与电气工程学院

摘要: 报告针对压缩感知和相位检索等线性和非线性成像逆问题, 重点介绍近期在基于深度学习的可解释优化求解和可靠重建方面的研究进展。首先, 针对压缩感知等线性成像逆问题, 构造自由结构的深度展开网络, 通过可微分编程拓展传统的稀疏表示, 展开 ISTA 等传统迭代算法, 保证稳定收敛同时松弛深度神经网络的结构限制, 支持 ResNets、Transformer 等常见神

神经网络。针对稀疏先验，引入神经网络的隐式正则，建立深度均衡模型的优化求解，保证收敛至驻点。进一步对于扩散生成模型的成像逆问题求解，建立基于条件后验均值的统一引导方式，进行协方差优化。针对相位检索的非线性问题，提出基于去噪先验的 Wirtinger 流框架，大幅降低稳定重建的过采样率。在去噪器更为宽松的半李普希茨条件下，对于广泛有噪/无噪场景下的实信号和复信号重建，保证几何速率收敛到最优点。

D2-2 可信预测与时序可解释方法

顾心悦，阿里巴巴达摩院

摘要：在电力预测等高风险决策场景中，预测方法不仅要提供高精度的结果，更要确保结果可理解、可信赖，从而支撑业务安全、高效地开展。阿里巴巴达摩院决策智能实验室根据相关场景研发了可信、透明的预测模型与模型解释方法。一方面，提出了白盒的预测模型，一种基于 GAM 的可编辑方法，在人理解机器之后，也让机器有能力融合人提出的专业经验，以提升小样本与极端场景的预测表现。另一方面，设计了面向时间序列的通用黑盒可解释方法，将预测结果拆解到不同输入片段及其交互效应上，使各片段对于模型的重要性一目了然，解决了一线业务人员对于预测结果的理解和信任问题。相应的方法已经产品化并在多个实际场景中成功落地。

D2-3 凝聚现象在多种算法训练的神经网络中的普遍性

李雨晴，上海交通大学/数学科学学院

摘要：在本次报告中，我们将探讨三个主要研究领域。首先，通过 Huang 和 Yau 提出的神经切线层次框架（NTH），研究了有限宽度深度残差网络（ResNet）的神经切线核（NTK）动态行为。结果显示，对于具有光滑和 Lipschitz 激活函数的 ResNet，层宽 m 相对于训练样本数 n 的要求可从四次方降至三次方；其次，我们关注神经网络在不同初始化尺度下的行为差异。基于 Luo 等人的研究，我们展示了两层神经网络和卷积神经网络的初始凝聚现象的相图。凝聚指的是训练过程中权重向量集中于孤立方向的现象，这是非线性学习的关键特征，有助于提升神经网络的泛化能力；最后，我们研究了广泛应用的神经网络正则化技术——Dropout 算法。通过严格的理论推导，得出了随机修正方程，为 Dropout 的离散迭代过程提供有效近似。此外，实验证据表明，Dropout 在神经网络训练结束时确实促进了凝聚现象。

D2-4 ODE-based Learning to Optimize

谢中林，北京大学数学科学学院

摘要: Recent years have seen a growing interest in understanding acceleration methods in optimization through the lens of ordinary differential equations (ODEs). Despite the theoretical advancements, translating the rapid convergence observed in continuous-time models to discrete iterative methods poses significant challenges. In this talk, we present a comprehensive framework integrating the dynamical inertial Newton with asymptotic vanishing damping equation (DIN-AVD) for developing optimization methods through a deep synergy of theoretical insights and learning-based approaches. We first establish the convergence condition for ensuring the convergence of the solution trajectory of DIN-AVD. Then, we show that provided the stability condition, another relaxed requirement on the coefficients of DIN-AVD, the sequence generated through the forward Euler discretization of DIN-AVD converges, which gives a large family of practical optimization methods. In order to select the best optimization method in this family for certain problems, we introduce the stopping time, the time required for an optimization method derived from DIN-AVD to achieve a predefined level of suboptimality. Then, we formulate a novel learning to optimize (L2O) problem aimed at minimizing the stopping time subject to the convergence and stability conditions. To navigate this learning problem, we present an algorithm combining stochastic optimization and the penalty method (StoPM). We also leverage the conservative gradient to address the challenges posed by non-smoothness inherent in neural networks. The convergence of StoPM using the conservative gradient is proved. Empirical validation of our framework is conducted through extensive numerical experiments across a diverse set of optimization problems. These experiments showcase the superior performance of the learned optimization methods, which highlight the robustness and adaptability of our framework, underscoring its potential to advance the ODE viewpoint of optimization and L2O.

三十二、分会报告 D3 专题: Learning dynamical models from data: algorithms, analyses and applications

D3-1 Predicting and modulating complex dynamics using data-driven and machine learning techniques

林伟, 复旦大学数学科学学院/智能复杂体系实验室

摘要: In the era of the data science, model-free techniques are developed overwhelmingly. When the experimentally-collected data are generated by dynamical systems, the missions of reconstruction, prediction, and modulation only based on these data are highly anticipated to be achieved for these systems. Here, we introduce several directions of progresses made by our research group in developing the model-free techniques using machine learning techniques and dynamical systems theory. We use representative systems of physical or/and biological significance to demonstrate the developed techniques. We hope that all the methods can shed a light on deciphering and controlling the hidden dynamics that dominate the evolutions of any systems in real-world.

D3-2 From Data to Dynamics: Deep Modeling of Unknown Differential Equations

吴开亮, 南方科技大学

摘要: This talk will introduce effective methods for learning unknown time-dependent differential equations from measurement data. We will explore the importance of using many short bursts of trajectory data instead of a few long trajectories. We will present several data-driven modeling strategies using deep neural networks. It will be shown that residual networks are particularly suitable for equation discovery, as they can produce an exact time integrator for numerical predictions. Additionally, the deep learning of unknown partial differential

equations in modal or nodal spaces will be discussed, along with recent advances in structure-preserving learning approaches for unknown equations.

D3-3 RoeNet: 基于数据预测双曲系统的间断解

熊诗颖, 浙江大学航空航天学院

摘要: 预测从训练数据集中无法观察到的未来不连续现象一直是科学机器学习中的一个难题。我们提出了一种新的范式, 通过在短时间窗口内给定的训练数据和有限的不连续信息, 预测各种双曲型偏微分方程不连续现象的出现和演化。我们的方法受到经典 Roe 求解器的启发, 并通过设计动力学演化网络结构和基于广义逆的矩阵处理模块, 使数据驱动的预测器能够满足 Roe 求解器的基本数学约束。通过预测双曲问题间断解的演化, 我们展示了数据驱动的 RoeNet 在准确性和鲁棒性方面优于经典的 Roe 求解器和具有弱先验知识的深度神经网络。

D3-4 动态高斯混合近似: 从数据中学习随机微分方程

祝爱卿, National University of Singapore

摘要: 从观测数据中学习未知的随机微分方程是一项重要且具挑战性的任务。目前的方法通常使用神经网络来逼近漂移和扩散函数, 并通过近似密度来构建基于似然的损失以训练这些网络。然而, 这些方法通常依赖于单步随机数值格式, 因此需要具有足够高时间分辨率的数据。在本次报告中, 我们将介绍对参数化 SDE 的转移密度的新近似方法: 一种基于动力系统随机扰动理论的高斯密度近似, 以及它的扩展形式, 动态高斯混合近似 (DynGMA)。由于更加精确的密度近似, 我们的方法在学习未知漂移和扩散函数以及从轨迹数据中计算不变分布方面表现出优于基线方法。而且它能够处理低时间分辨率和可变甚至不可控时间步长的轨迹数据, 例如由 Gillespie 随机模拟生成的数据。我们还将展示多个不同场景下的实验结果。

三十三、分会报告 D4 专题: 机器学习与交叉 (II)

D4-1 GAN 训练的三个阶段

高卫国, 复旦大学

摘要: 生成对抗网络 (GAN) 是高效生成模型, 但会出现模式混合和模式崩溃等现象。我们将 GAN 训练分为三个连续阶段——拟合、精化和崩溃, 提供一个全局刻画, 强调了模式混合和精化阶段以及模式崩溃和崩溃阶段之间的强相关性。为分析每个阶段的原因和特征, 我们提出一个

新的理论框架，整合了 GAN 的连续和离散方面，克服了现有工作只关注一个方面的缺点。我们还发展了一个指标来检测从精化到崩溃的阶段转变，并将其集成到“早期停止”算法中以优化 GAN 训练。在合成数据集和真实数据集上的实验结果证实了我们的理论分析并验证了我们算法的有效性。这是和厉茗的合作工作。

D4-2 准二维库伦体系中核函数的指数函数求和估计及其 $O(N)$ 复杂度的随机分批分子动力学模拟

干则成，香港科技大学（广州）

摘要：准二维库仑多体粒子系统在许多前沿科技领域具有重要的应用。该类系统本身的对称性破缺导致了大量有趣的集群行为的涌现，但也给基于粒子的多体模拟带来了巨大的挑战。我们提出了一种新的算法框架，来降低该类体系模拟中由于库仑相互作用的长程性质导致的 $O(N^2)$ 复杂度，其中 N 为粒子数量。首先，我们引入了一个高效的指数函数求和（SOE）近似，将该近似用于对库伦势作 Ewald 分解之后的长程核函数的非周期方向，我们实现了关于粒子间距的一致收敛性，同时将复杂性降低到 $O(N^{7/5})$ 。实际分子动力学模拟中，我们发现 SOE 近似之后的 k -空间能量表达式在周期方向上有指数衰减的性质。因此，我们在周期方向引入 k -空间的随机分批重要性采样方法。我们证明这种随机分批近似是无偏的，具有方差减小的效果，且能够进一步将计算成本降低到 $O(N)$ 。我们通过数值示例展示了我们算法的性能和精度。与传统的 Ewald2D 方法相比，它实现了 2~3 个数量级的加速，使得在单核上能够进行高达 10^6 个粒子的分子动力学模拟。因此，该方法非常适合于准二维约束下，库仑多体系统的大规模分子模拟。

D4-3 传统无网格方法求解黎曼流形上的向量场方程

蒋诗晓，上海科技大学 数学科学研究所

摘要：我将介绍广义移动最小二乘法求解嵌入 3 维欧式空间的 2 维流形上的向量场偏微分方程，流形上的点是由随机点云数据集给出，该方法利用流形上的导数恒等于欧式空间导数在流形切空间上的投影的性质，推导给出了流形上向量场拉普拉斯的恒等式，再利用移动最小二乘法进行插值离散。该方法的问题是离散后的矩阵维度和非零元个数依赖于欧式空间维度，造成较高的计算复杂度，我们于是基于某种投影方法将矩阵维度和计算复杂度降低到了匹配流形的内在维度。我们展示的数值结果包含向量场拉普拉斯矩阵的特征值问题、线性 Poisson 方程、线性含时演化方程和非线性 Burgers 方程。

D4-4 非局部模型的理论 and 计算

张继伟，武汉大学数学与统计学院

摘要：非局部效应行为在自然界是普遍存在的，然而人们更多的把关注的目光放在了局部的偏微分方程模型，主要因为很多时候非局部作用的影响不大，可以利用局部模型进行简化。但是当人们对于极端异构环境中的复杂过程越来越感兴趣，特别是在有些情况下如材料出现断裂和破碎时，经典的偏微分方程模型在描述这些过程时是不适定的。本报告将简单介绍我们在基础数学理论、快速算法以及高效实现方面做的努力和尝试。尤其是三维非局部模型的有限元实现涉及高维积分、巨大的计算量和存储量，因此在算法设计、网格生成、数据结构优化、面向并行平台的算法架构等诸多方面都提出了挑战，这里将简要介绍高维有限元算法实现方面的一些思路和进展，以及可能需要的机器学习等算法。

三十四、分会报告 D5 专题：数据科学中的随机模型与算法（II）

D5-1 Wasserstein 梯度流的保结构原始对偶方法

魏朝祯，电子科技大学/数学科学学院

摘要：我将提出一种新的数值方法来求解一类具有密度依赖迁移率的连续性方程，这类方程在生物和材料科学中得到了广泛的应用，如肿瘤生长、蜂群动力学、固态润湿/去湿和薄膜表面活性剂动力学等。这类方程可以看作是某种迁移率诱导的广义 Wasserstein 输运距离的梯度流。通过利用该变分结构，并结合 Wasserstein 输运距离的动力学刻画，我们构造了一个全离散 JKO 格式，最终化为一个具有严格凸目标函数和线性约束的最小化问题，该问题可由原始对偶算子分裂格式及其加速版本求解。该方法具有保界性、质量守恒性和能量耗散性，克服了强非线性和退化性带来的稳定性问题。我将展示一组数值模拟算例来说明我们算法的有效性。

D5-2 微分方程高效神经网络方法及自适应数据采样技巧

于海军，中国科学院数学与系统科学研究院

摘要：近年来，发现和求解（偏）微分方程的深度学习的方法逐渐成为计算数学学科的一个重要研究内容。常见的微分方程深度学习的方法有基于函数逼近的 DeepRitz, DGM/PINN, WAN 等，

和基于算子逼近的 PDE-net, DeepONet, FNO, OnsagerNet 等. 在函数逼近类方法中, 尽管 DeepRitz 和 PINN 已经成为此类方法的两个事实基准, 但在应用到复杂问题中还存在各式各样的困难. 比如: DeepRitz 方法不方便处理非梯度系统, 而且对数值积分误差比较敏感; 而 PINN 方法虽然使用简单, 但是无法处理弱解. 本报告中, 我们将结合 DeepRitz 和 PINN 的优点提出一种新方法, 并借鉴函数逼近论和自适应网格等思想对网络结构和自适应数据采样进行分析.

D5-3 抽样不变测度的弱生成网络抽样方法

周翔, City University of Hong Kong

摘要: 本报告介绍基于福克普朗克方程的弱形式构造的高效实用的直接从随机微分方程生成满足其不变测度的 iid 样本. 不同于传统直接模拟, 生成网络直接训练映射产生数据 而无需求解其概率密度函数. 基于弱形式和数据驱动测试函数, 新的计算方法比传统基于 PINN 的有效提高模式坍塌的处理能力.

D5-4 基于 Gromov-Monge 正则化的单调生成器

邹东勉, 昆山杜克大学

摘要: 生成对抗网络具有模式坍塌和训练欠稳等缺点. 为了克服这些缺点, 我们在保留结构特征的同时将数据分布映射到低维空间中, 然后构建低维分布到参考分布的最优传输. 如此构建的生成器表现出单调性. 数值实验证实该方法有效生成高质量图像, 并且在模式坍塌和训练欠稳等问题上更具优势.

三十五、分会报告 D6 专题: 机理驱动的深度学习方法 和理论

D6-1 Exploring low-dimensional data structures by deep neural networks with applications on operator learning

刘皓, 数学系

摘要: Deep neural networks have demonstrated a great success on many applications, especially on problems with high-dimensional data sets. In spite of that, most

existing statistical theories are cursed by data dimension and cannot explain such a success. To bridge the gap between theories and practice, we exploit the low-dimensional structures of data set and establish theoretical guarantees with a fast rate that is only cursed by the intrinsic dimension of the data set. Autoencoder is a powerful tool in exploring data low-dimensional structures. In our work, we analyze the approximation error and generalization error of autoencoders and its application in operator learning. Our results provide fast rates depending on the intrinsic dimension of data sets and show that deep neural networks are adaptive to low-dimensional structures of data sets.

D6-2 基于学习的多尺度流动多连续体模型

王亚婷, 西安交通大学

摘要: 数值均匀化是近似求解多尺度问题的一个经典方法, 它通过求解粗尺度上的有效参数来捕获原始系统在粗网格上的宏观行为, 从而加快仿真速度。然而, 这种方法通常假设尺度分离, 且解的异质性可以通过每个粗块中的平均值来近似。对于复杂的多尺度问题, 单个有效参数可能精度不足。我们针对复杂多尺度问题, 介绍了一个基于学习的多连续体模型来丰富均质化方程, 通过给定观测数据来提高单连续体模型的准确性。该模型可解析每个粗网格块内的多个相互作用介质, 并描述它们之间的质量传递。数值实验证明该基于数据的多连续体模型可显著改善模拟结果。

D6-3 关于“复杂微环境中细胞迁移随机动力学”的深度学习方法

徐新鹏, 广东以色列理工学院

摘要: 细胞迁移是生物体中一种基本的过程, 对于多种生物学现象至关重要, 包括胚胎发育、创伤愈合、组织再生、免疫反应以及肿瘤进展。量化并揭示细胞迁移在响应外部生化或生物物理刺激时的随机性生物物理机制仍然具有挑战性。在本研究中, 我们提出了一种通用的物理信息机器学习方法——深度 Onsager-Machlup 方法 (DOM, 基于随机轨迹或细胞的概率分布函数) 来研究不同刺激下的随机细胞迁移。作为示例, 我们考虑了具有物理和/或生物学意义的各种势能下的被动和主动布朗粒子的随机动力学。我们展示了随着对系统物理知识输入的系统性增加, 我们的深度学习方法的准确性变得越来越准确和可靠。此外, 即使对于被动布朗粒子, 我

们的 DOM 方法也比传统的基于统计物理的方法显示出许多优势。使用更少的数据（较少的轨迹或更少的粒子数量）就可以恢复势能景观。将机器学习整合到量化细胞迁移中有助于全面理解动态复杂微环境中的细胞运动性。我们的 DOM 方法为生物系统中不同长度和时间尺度的预测建模提供了一种资源高效且可解释的机器学习框架。

D6-4 卷积神经网络的特征稀疏性及其驱动的学习策略研究

冯寒，数学系/香港城市大学

摘要：在本次演讲中，我们将结合稀疏编码理论探讨卷积神经网络（CNN）的特征提取能力。首先，我们将介绍一个针对深度稀疏编码模型的框架，并确立这些模型的唯一性和稳定性所需的条件。此框架通过稀疏性提供了一种坚实的数据表征学习方法。接着，我们将转向对 CNN 近似稀疏特征能力的严格分析，为其在特征提取和表示学习中的有效性提供理论基础。在这些理论见解的基础上，我们提出了一种专为 CNN 设计的新颖特征稀疏学习策略。我们进行了广泛的数值实验来验证我们所提策略的性能和效果。结果表明，我们的策略在特征表征和图像处理任务中取得了显著的改进。这不仅加深了我们对 CNN 理论的理解，也为设计更高效、更强大的神经网络模型开辟了新的途径。

三十六、分会报告 D7 专题：深度学习求解复杂流体的高效算法及分析（II）

D7-1 基于人工智能的流场可视化测量与计算

蔡声泽，浙江大学

摘要：粒子跟踪测速等流动可视化技术在航空航天、生物医学等领域有着广泛应用，对理解流动机理具有重要意义。其中，如何实现高浓度粒子分布、大动态速度范围等场景下的双帧粒子匹配，并实现全局流场的反问题重建计算，是可视化测速技术存在的主要挑战。针对该问题，本报告介绍基于图神经网络的三维粒子运动估计与跟踪测速方法，并在神经网络中依据最优传输思想挖掘相邻两帧粒子特征之间的匹配关系；随后，采用物理启发式神经网络，实现从稀疏测量到全局流场的反问题计算，能从少量和带噪声的速度测量反演三维时空速度场和压力场，在仿生流场计算、生物体内流体测量等场景中成功应用。

D7-2 使用算子神经学习弥合多尺度气泡动力学中的尺度差异

林晨森, 复旦大学

摘要: 气泡生长动力学是一个复杂的过程, 涉及从气泡形成的微观力学到气泡与周围热流体动力学之间的宏观相互作用的广泛的物理现象。传统的气泡动力学模型, 包括原子尺度的方法和基于连续体的方法, 通常将气泡动力学划分为不同尺度的特定模型。为了弥合微观随机流体模型和基于连续体的流体模型在气泡动力学中的断层, 我们开发了一种复合神经算子模型, 通过新颖的神经网络架构, 整合多体耗散粒子动力学 (mDPD) 模型与基于连续体的 Rayleigh-Plesset (RP) 模型, 从而统一了微观和宏观尺度下非线性气泡动力学的分析。该神经网络架构包括一个深度算子网络, 用于学习气泡在压力变化下的平均生长行为, 以及一个长短期记忆网络, 用于学习微观气泡动力学中的统计波动特征。训练和测试数据是通过进行 mDPD 和 RP 模拟, 针对初始气泡半径在 0.1 到 1.5 微米范围内的非线性气泡动力学产生的。结果表明, 训练后的复合神经算子模型可以准确预测跨尺度的气泡动力学, 在变化的外部压力下对气泡半径随时间变化的评估具有 99% 的准确性, 同时在微观气泡生长动力学中包含了正确的依赖于尺寸的随机波动。该复合神经算子是首个能够捕捉微观流体现象中随机波动的多尺度气泡生长动力学的深度学习代理模型, 为未来多尺度流体动力学建模研究提供了新的思路。

D7-3 基于 MIONet 的 PDE 混合迭代法

金鹏展, 北京大学大数据分析与应用技术国家工程实验室

摘要: 对于 PDE 求解, 我们结合有限元的传统数值迭代求解器和神经算子 MIONet, 提出了一种基于 MIONet 的混合迭代方法。我们系统地分析了其理论性质, 包括收敛条件、谱行为和收敛率。我们展示了常用磨光器对应的混合迭代理论结果, 包括 Richardson (阻尼 Jacobi) 和 Gauss-Seidel。我们给出了混合迭代法收敛率的上界, 该上界蕴含了混合迭代收敛率关于模型修正周期的最优点。此外, 几个数值算例, 包括一维 (二维) 泊松方程的混合 Richardson (Gauss-Seidel) 迭代, 验证了我们的理论结果, 并反映出优秀的加速效果。作为一种无网格加速方法, 它具有巨大的实际应用潜力。

D7-4 物理增强流场智能计算方法

金晓威, 哈尔滨工业大学

摘要: 流体运动理论上可用 Navier-Stokes 方程描述, 但由于对流项带来的方程高度非线性,

仅在少数情况可求得方程解析解。对复杂工程流动问题，数值模拟难以高效精准计算流场，实验或现场测量获得流场丰富细节成本高。针对上述问题，围绕机器学习高维空间非线性函数逼近能力，采用机器学习隐式或显式地逼近控制方程与模型的高维空间解，提出了物理增强流场智能计算方法，包括机器学习显式求解微分方程方法、数值计算-机器学习混合方法、物理增强微分方程数据求解方法。

三十七、分会报告 D9 专题：机器学习与最优化方法

(II)

D9-1 Hyper-gradient in bilevel optimization: efficient computation by Krylov Subspace and enhanced investigation in reinforcement learning

高斌，中国科学院数学与系统科学研究院

摘要: Bilevel optimization, with broad applications in machine learning, has an intricate hierarchical structure. Hyper-gradient-based methods have emerged as a common approach to large-scale bilevel problems. However, the computation of hyper-gradient which involves a Hessian inverse vector product, confines the efficiency and is regarded as a bottleneck. In this talk, for bilevel problems with lower-level strong convexity, we propose a provable subspace-based framework, which constructs a sequence of low-dimensional approximate Krylov subspaces able to dynamically and incrementally approximate the Hessian inverse vector product with less effort, and thus leads to a favorable estimate of the hyper-gradient. In the context of bilevel reinforcement learning (RL), a class of structured bilevel optimization problem, by employing the fixed point equation associated with the regularized RL, we characterize the hyper-gradient via fully first-order information, thus circumventing the assumption of lower-level convexity. Moreover, we propose both model-based and model-free bilevel reinforcement learning algorithms, both of which are provable to enjoy the optimal convergence rate.

D9-2 面向海量数据和异构数据的高效深度学习

林涛, 西湖大学

摘要: 现代深度学习的成功在很大程度上依赖于大量高质量的数据。随着端侧设备和生成模型的不断发展, 海量数据将持续产生, 为模型能力的进化提供机遇和挑战。在本讲座中, 我将从数据的采样、生成、和协同这三个角度剖析深度表征学习, 并提出以数据为中心的优化与泛化理论和高效训练算法。

D9-3 分布式优化算法

严明, 香港中文大学(深圳)

摘要: 随着大数据和人工智能的发展, 分布式算法已经成为解决大规模问题不可或缺的工具。分布式算法可以看作是多智能体系统通过信息交互来处理分布式数据的一种方法。在追求计算效率的同时, 分布式优化还需要考虑通信效率。我将介绍一些提高分布式算法计算和通信效率的方法, 主要包括: 算法加速, 通信压缩, 异步并行等。

D9-4 一类复合优化问题的最优复杂度上界和下界

朱桢源, 北京大学数学科学学院

摘要: 我们研究了找到复合问题 $\min_x f(x) + h(Ax - b)$ 的近似解的最优下界和上界复杂度, 其中 f 是光滑的, h 是凸的。假设 h 的近似算子容易计算, 对于强凸、凸和非凸的 f , 我们分别设计了一系列和增广拉格朗日函数法紧密相关的高效一阶算法, 它们分别具有复杂度 $\tilde{O}(\kappa_A \sqrt{\kappa_f} \log(1/\epsilon))$ 、 $\tilde{O}(\kappa_A \sqrt{L_f} D / \sqrt{\epsilon})$ 和 $\tilde{O}(\kappa_A L_f \Delta / \epsilon^2)$ 。这里, κ_A 是组合中矩阵 A 的条件数, L_f 是 f 的光滑常数, κ_f 是强凸情况下 f 的条件数, D 是初始点距离, Δ 是初始函数值差。三种情况的紧致下界复杂度也被推导出, 并且与上界在忽略对数常数下相匹配, 从而证明了所提出的上下界的最优性。

三十八、分会报告 E1 专题: 学生分会 (I)

E1-1 基于机器学习的结构优化算法

梅子健, 人工智能与数据科学学院

摘要：人类社会的高速发展离不开各类复杂结构和系统，如飞行器、汽车等，如何优化是工业和学术界一直以来研究的问题。优化方法大体可以分为两种，基于梯度的方法和基于无梯度的方法，其中基于梯度的方法虽然效率高，但是往往容易陷入局部最优，且有时难以目标函数关于参数的导数，而基于无梯度的方法，虽然有概率收敛到全局最优，但代价是需要更多的迭代次数和仿真成本。随着机器学习的发展，越来越多基于深度学习或深度学习赋能的方法被提出来提高结构优化的效率与效果。本次汇报将简要回顾现有的一些基于深度学习的结构优化方法，并介绍一下陈老师课题组在该方面的研究成果。

E1-2 神经网络损失景观的深度嵌入原则

白志威，数学科学学院

摘要：在本报告中，我们将深入探讨深度神经网络与浅层神经网络之间的关系，特别关注它们损失景观中的临界点。我们发现了一种深度嵌入原则，即深度神经网络的损失景观“包含”了所有比它浅网络损失景观的临界点。具体来说，我们利用临界提升的方法，可以将任意网络的临界点映射到更深层网络的临界流形中，同时保持输出不变。为研究这一原则的实际影响，我们进行了多方面的数值实验。实验结果表明，深度神经网络在训练过程中实际上会遇到这些提升后的临界点，导致不同深度的网络具有相似的训练动力学和泛化表现。我们的理论和实验证据表明，提升后的临界点表现出更高的退化性。深度嵌入原则还为批量归一化 (Batch Normalization) 和更大数据集在优化方面的优势提供了见解，并实现了网络层结构上的压缩等实际应用。深度嵌入原则的发现揭示了深度学习损失景观深度上的层级结构，为进一步研究深度在神经网络中的作用奠定了基础。

E1-3 泛化误差的相变行为

严菁，数学科学学院

摘要：在机器学习领域，泛化误差与样本数量之间的关系是一个重要课题。本次报告将探讨这种关系，重点关注其中存在的急剧转变。

E1-4 Loss Spike in Training Neural Networks

李晓龙，上海交通大学数学科学学院

摘要：In this work, we investigate the mechanism underlying loss spikes observed during neural network training. When the training enters a region with a lower-

loss-as-sharper (LLAS) structure, the training becomes unstable, and the loss exponentially increases once the loss landscape is too sharp, resulting in the rapid ascent of the loss spike. The training stabilizes when it finds a flat region. We explain this phenomenon from the frequency perspective, observing that the deviation in the first eigen direction is predominantly influenced by low-frequency components. Since low-frequency information is captured very quickly (frequency principle), a rapid descent is subsequently observed. Inspired by our analysis of loss spikes, we revisit the link between the maximum eigenvalue, flatness and generalization. The maximum eigenvalue of the loss Hessian is a good measure of sharpness but not a good measure for generalization. Furthermore, we experimentally observe that loss spikes can facilitate condensation, causing input weights to evolve towards the same direction. And our experiments show that there is a correlation (similar trend) between λ_{max} and condensation. This observation may provide valuable insights for further theoretical research on the relationship between loss spikes, λ_{max} , and generalization.

三十九、分会报告 E2 专题：学生分会（II）

E2-1 通过 Fokker-Planck 方程探究半梯度 Q 学习的隐式偏好

尹树雨，电子信息与电气工程学院

摘要：半梯度 Q 学习 (Semi-gradient Q-learning) 在许多领域中都有应用，但由于半梯度方法没有对应的显式损失函数，其在参数空间中的动力学和隐式正则化难以研究。本课题引入 Fokker-Planck 方程，在二维参数空间进行了等效损失景观的可视化。该可视化展示了梯度方法对应的损失景观中的全局极小点在等效损失景观中可能会转化为鞍点，并由此展现了半梯度 Q 学习的隐式正则化。最终，我们展示了在高维参数空间和神经网络的设定下等效损失景观中依然存在由全局极小点转化而来的鞍点。本课题为理解半梯度方法在参数空间中的隐式正则化提供了新的视角。

E2-2 Input gradient annealing neural network for solving low-temperature Fokker-Planck equations

杭良慨, 数学科学学院

摘要: We present a novel yet simple deep learning approach, called input gradient annealing neural network (IGANN), for solving stationary Fokker-Planck equations. Traditional methods, such as finite difference and finite elements, suffer from the curse of dimensionality. Neural network based algorithms are meshless methods, which can avoid the curse of dimensionality. However, at low temperature, when directly solving a stationary Fokker-Planck equation with more than two metastable states in the generalized potential landscape, the small eigenvalue introduces numerical difficulties due to a large condition number. To overcome these problems, we introduce the IGANN method, which uses a penalty of negative input gradient annealing during the training. We demonstrate that the IGANN method can effectively solve high-dimensional and low-temperature Fokker-Planck equations through our numerical experiments.

E2-3 Quantifying Training Difficulty and Accelerating Convergence in Neural Network-Based PDE Solvers

陈楚淇, 数学系/香港科技大学

摘要: Neural network-based methods have emerged as a powerful tool for solving partial differential equations (PDEs) in scientific and engineering applications, particularly when dealing with complex domains or incorporating empirical data. These methods leverage neural networks as basis functions to approximate the PDE solutions. However, training such networks can be challenging, often resulting in limited accuracy. In this paper, we delve into the training dynamics of neural network-based PDE solvers, focusing on the impact of eigenvalue distribution and initialization techniques. We assess training difficulty by analyzing the eigenvalue distribution of the gradient descent kernel and introduce the concept of truncated effective rank to quantify this difficulty. A larger truncated

effective rank leads to faster convergence of the training error. Through experiments and theoretical analyses, we showcase the effectiveness of two initialization techniques, Partition of Unity (PoU) and variance scaling, in enhancing the truncated effective rank. Finally, we experimentally validate the efficacy of these techniques on popular PDE solving frameworks, including PINN, Deep Ritz, and DeepONet models, demonstrating their ability to improve training efficiency and accuracy.

E2-4 大型 Transformer 是更好的 EEG 学习者

王冰鑫, 香港科技大学数学系

摘要: 脑电图 (EEG) 长期以来一直是揭开人类大脑复杂性的有力工具。基于深度学习的脑电图特征提取方法可应用于帕金森病和癫痫等疾病的检测。在这些模型中, 预训练的大型 Transformer 模型在自然语言处理和计算机视觉领域取得了卓越的表现。然而, 带注释的脑电图数据有限, 这对将这些模型的成功扩展到基于脑电图的应用提出了独特的挑战。为了解决这一差距, 我们提出了 AdaCE, 这是一种无需训练的适配器, 旨在将时间序列脑电图数据转换为时空二维伪图像或文本形式。所提出的方法允许在脑电图数据分析中无缝集成预训练的视觉模型和语言模型, 而无需引入额外的神经网络层。在包括多模态帕金森、癫痫发作识别和睡眠 EDF 在内的各种脑电图数据集上的实验结果证明了 AdaCE 优于基线方法。总的来说, 我们提出了一个实用的迁移学习框架, 以证明在 EEG 预测任务中利用预训练视觉和语言转换模型的能力的可行性。