目 录

一、会务指南

中国机器学习与科学应用大会由北京大学数学科学学院、北京大学国际机器学习研究中心和北京大学北京国际数学研究中心主办，上海交通大学自然科学研究院和上海交通大学数学科学学院协办，会议围绕数学的机器学习、科学的机器学习、工业的机器学习进行深入交流研讨。

会议时间： 2022 年 8 月 18（周四）-19 日（周五）

会议地点： 北京大学中关新园（北京市海淀区中关村北大街 126 号）

酒店网址：pkugv.pku.edu.cn

酒店电话：（8610）62752288

会议网站： c2sml.cn

大会邀请报告直播/回放链接： https://www.koushare.com/topicIndex/i/CSML2022

直播二维码

会务组联系方式：

田　甜：ttian@math.pku.edu.cn

刘艳云：yyliu@pku.edu.cn

周俊艺：zhoujunyi@pku.edu.cn

二、会议组织

大会主席：

鄂维南（院士，北京大学）

会议学术委员会(按姓氏拼音排序)

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）

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(按姓氏拼音排序)：

包　刚（教授，浙江大学）

董　彬（副教授，北京大学）

鄂维南（院士，北京大学）

金　石（教授，上海交通大学）

许志钦 (副教授，上海交通大学)

张平文（院士，北京大学）

邀请大会报告人(按姓氏拼音排序)：

史作强（清华大学）

王立威（北京大学）

吴　昊（同济大学）

许志钦（上海交通大学）

张林峰（深势科技）

会议组织委员会(按姓氏拼音排序)：

董彬 (北京大学)，许志钦 (上海交通大学)，朱军 (清华大学)

三、会议日程

2022 年 8 月 18-19 日，北京大学中关新园 1 号楼

<table>
<tr><th>时间</th><th>活动内容</th><th>主持人</th><th>地点</th></tr>
<tr><td colspan="4">8 月 18 日</td></tr>
<tr><td>7:30-8:30</td><td colspan="2">签到</td><td rowspan="7">B1 群英宴会厅</td></tr>
<tr><td>8:30-8:35</td><td colspan="2">开场：董彬</td></tr>
<tr><td>8:35-8:40</td><td>致辞：鄂维南院士</td><td rowspan="3">董彬</td></tr>
<tr><td>8:40-9:25</td><td>邀请报告
吴昊（同济大学）
多体系统运动学降阶的机器学习方法</td></tr>
<tr><td>9:30-10:15</td><td>邀请报告
史作强（清华大学）
半监督学习的微分方程模型</td></tr>
<tr><td>10:15-10:30</td><td colspan="2">茶歇</td></tr>
<tr><td>10:30-12:00</td><td>圆桌论坛</td><td>董彬</td></tr>
<tr><td>12:00-12:15</td><td colspan="2">合影</td><td>1 号楼门口</td></tr>
<tr><td>12:15-13:30</td><td colspan="2">午餐</td><td>时光西餐厅（二层）</td></tr>
<tr><td rowspan="2">13:30-15:30
分会 A</td><td>A1 专题：图像科学中的机器学习方法——前沿进展及挑战（一）
（腾讯会议：998-210-141）
梁　栋：Deep Fast MR Imaging: Deep Learning meets Nuclear Spins（线上）
孟德宇：重思图像复原的基本方法论（线上）
高　跃：超图计算（线上）
邱凌云：Non-line-of-sight Imaging</td><td>史作强</td><td>观湖 B 会议室（二层）</td></tr>
<tr><td>A2 专题：复杂度和泛化理论
（腾讯会议：774-190-458）
谢泽柯：Deep Learning Dynamics: On Minima Selection and Saddle-Point Escaping
张景昭：Neural Network Weights Do Not Converge to Stationary Points: An Invariant Measure Perspective
焦雨领：Error analysis of GAN（线上）
周岐轩：PINN 的失效与其隐式正则化</td><td>罗涛</td><td>观湖 C 会议室（二层）</td></tr>
</table>

	A3 专题：科学计算与机器学习 【Tencent Meeting: 512-466-594 (webinar)】 包承龙：Learning Robust Imaging Models without Paired Data 张小群：A stochastic three-block splitting algorithm and its application to quantized deep neural networks（线上） 石　磊：Classification with Deep Neural Networks	明平兵 李颖洲 （线上）	映塔 A 会议室 （二层）
	A4 专题：最优输运理论、计算和应用 （腾讯会议：118-391-078） 孟　澄：Importance Sparsification for Sinkhorn Algorithm 李　磊：Learning geodesics under spherical WFR metric and its application to generation of weighted samples（线上） 包承龙：PROXIMAL POINT METHODS FOR COMPUTING (UN-)BALANCED WASSERSTEIN DISTANCE 王子豪：最优输运在语义信息匹配中的应用（线上）	吴昊	观湖 A 会议室 （二层）
	A5 专题：机器学习与计算数学（一） （腾讯会议：905-676-200） 王　东：Theory and algorithms on archetype analysis（线上） 郭　玲：Uncertainty Quantification in Scientific Machine Learning: Methods and Comparisons（线上） 蒋诗晓：Radial Basis Functions for Approximating Differential Operators on Closed Manifolds 于海军：谱方法与深度神经网络逼近	于海军	观湖 D 会议室 （二层）
	A6 专题：语言智能分论坛 （腾讯会议：902-151-182） 秦　兵：大模型背景下可信自然语言处理（线上） 张家俊：多语言机器翻译中知识迁移的不对称性问题研究 冯　洋：开放域对话的语义信息流控制与一致性检测 高　莘：面向用户模拟与满意度的任务导向推荐系统评估	严睿	映塔 B 会议室 （二层）
15:30-16:00	茶歇		
16:00-18:00 分会 B	**B1 专题：图像科学中的机器学习方法—前沿进展及挑战（二）** （腾讯会议：824-800-070） 刘海广：3D Model reconstruction from 1D Small angle X-ray Scattering data（线上） 孙　剑：Deep Learning in Non-Euclidean Space（线上） 张　意：基于机器学习的智能成像研究进展（线上） 全宇晖：TBA（线上）	包承龙	观湖 A 会议室 （二层）

	B2 专题：反问题中的机器学习方法 （腾讯会议：104-878-091） 吕锡亮：Imaging Conductivity from Current Density Magnitude using Neural Networks（线上） 邱　越：Low-rank Methods for Bayesian Inverse Problems 闫　亮：Deep learning approach for Bayesian inverse Problems(线上） 臧耀华：Weak Adversarial Networks: A Deep Learning Framework for Solving High Dimensional Inverse Problems	陆帅 （线上）	观湖 C 会议室 （二层）
	B3 专题：机器学习中的优化问题（一） （腾讯会议：931-700-565） 陈彩华：Algorithmic Design for Wasserstein DRO Based Trustworthy Machine Learning（线上） 刘　歆：Decentralized Optimization Over the Stiefel Manifold by an Approximate Augmented Lagrangian Function（线上） 范金燕：Convergence properties of the stochastic Levenberg-Marquardt method（线上） 丁　超：Perturbation analysis of nonsmooth optimization on manifolds	文再文	映塔 A 会议室 （二层）
	B4 专题：机器学习与计算数学（二） （腾讯会议：545-973-882） 陈景润：Random feature method for solving PDEs（线上） 凌舒扬：Near-Optimal Bounds for Generalized Orthogonal Procrustes Problem via Generalized Power Method（线上） 杨　将：A Local Deep Learning Method for Solving High Order Partial Differential Equations（线上） 朱圣鑫：Distance geometry, quadratic form and parameter-turning for graph matching（线上）	王东 （线上）	观湖 D 会议室 （二层）
	B5 专题：机器学习与张量网络 （腾讯会议：730-765-124） 程　嵩：Revisit Tensor Networks Machine Learning Modeling 谢志远：粗粒化结构在神经网络中的应用 冉仕举：Functional tensor network solving many-body Schrödinger equation and multi-variable differential equations 李　伟：Thermal Tensor Networks: Recent Progress and Application to Quantum Magnetism	张潘 王磊	观湖 B 会议室 （二层）

	B6 专题：机器学习与神经科学 （腾讯会议：690-484-830） 弭元元：A brain-inspired method for motion pattern recognition 冷卢子未：类脑计算——从芯片到算法（线上） 肖彦洋：Smoothness in nature and smoothness in DNN（线上） 闵　斌：The representational geometry of sequence working memory in macaque prefrontal cortex	张耀宇 （线上）	映塔 B 会议室 （二层）
	8 月 19 日		
8:00-8:45	**邀请报告** 张林峰（北京深势科技有限公司；北京科学智能研究院） AI for Science 开源社区发展的实践与思考	董彬	B1 群英宴会厅
8:45-9:30	**邀请报告** 许志钦（上海交通大学）（线上） "Simple" implicit regularizations in deep learning		
9:30-10:15	**邀请报告** 王立威（北京大学） L2 is not the right loss for PINN when solving nonlinear PDE		
10:15-10:30	茶歇		
10:30-12:30 分会 C	**C1 专题：机器学习的工业应用与挑战** （腾讯会议：644-458-567） 洪蓝青：OoD 泛化与自动驾驶 Corner Cases（线上） 吕俊龙：\texttt{Para-CFlows}：ξC^kξ-universal Diffeomorphism Approximators as Superior Neural Surrogates（线上） 唐睿明：AutoML 在推荐系统特征交互建模中的应用 徐　航：Perception system in Autonomous Driving and its application in AutoML（线上） 张世枫：Generative Modelling with AI Lossless Compression	李震国 （线上） 徐航 （线上）	观湖 C 会议室 （二层）
	C2 专题：机器学习与多尺度建模 （腾讯会议：710-401-432） 毛志平：DeepMMnet for hypersonics: Predicting the coupled flow and finite-rate chemistry behind a normal shock using neural-network approximation of operators 孙　琪：Layer-Parallel Training of Residual Networks with Auxiliary-Variable Networks（线上） 杨　畅：Asymptotic preserving scheme for anisotropic elliptic equations with deep neural network 周一舟：Learning Thermodynamically Stable and Galilean Invariant Partial Differential Equations for Non-Equilibrium Flows	马征	观湖 B 会议室 （二层）

<table>
<tr><td rowspan="4"></td><td>C3 专题：机器学习中的优化问题（二）
（腾讯会议：931-700-565）
韩德仁：非线性不定正则临近点算法
赵欣苑：THOR, Trace-based Hardware-driven layer-ORiented Natural Gradient Descent Computation
郦旭东：Non-convex Factorization and Manifold Formulations in Low-rank Matrix Optimization（线上）
文再文：On the convergence analysis of variational Monte Carlo methods</td><td>文再文</td><td>映塔 A 会议室（二层）</td></tr>
<tr><td>C4 专题：机器学习与计算数学（三）
（腾讯会议：748-428-519，密码：0819）
方礼冬：DeePN2: A deep learning-based non-Newtonian hydrodynamic model（线上）
胡　丹：A non-gradient method for solving elliptic partial differential equations with deep neural networks（线上）
贺巧琳：Deep learning-based method for solving incompressible Navier-Stokes equation and Chan-Hilliard equation（线上）
李　颖：DeLISA: Deep learning based iteration scheme approximation for solving PDEs（线上）</td><td>李颖（线上）</td><td>观湖 D 会议室（二层）</td></tr>
<tr><td>C5 专题：强化学习分会
（腾讯会议号：266-597-802）
卢宗青：Fully Decentralized Multi-Agent RL
吴　翼：多样性强化学习：不光要赢，还要赢得精彩（线上）
杨耀东：合作博弈的通用求解框架
张伟楠：探索强化学习大模型</td><td>俞扬</td><td>观湖 A 会议室（二层）</td></tr>
<tr><td>C6 专题：机器学习与燃烧数值模拟（一）
（腾讯会议：275-603-342）
常军涛：冲压发动机内部流场场信息智能重构方法初步探讨（线上）
贾　明：机器学习在发动机喷雾燃烧模拟和优化中的应用（线上）
杨　斌：基础燃烧数据与燃烧反应动力学模型优化
张　弛：航空发动机燃烧振荡预测——机器学习图像分析方法
蔡伟伟：深度学习赋能的多维燃烧诊断技术</td><td>陈正
张天汉</td><td>映塔 B 会议室（二层）</td></tr>
<tr><td>12:30-13:30</td><td colspan="2">午餐</td><td>时光西餐厅（二层）</td></tr>
</table>

13:30-15:30 分会 D	**D1 专题：因果推断与统计学习中人工智能应用的新进展** （腾讯会议：431 346 285，密码：949714） 廖振宇：Random Matrix Methods for Machine Learning: “Lossless”Compression of Large Neural Networks 苗　旺：Paradoxes and solutions for semiparametric fusion learning with external summary statistics 杨朋昆：Functional approximation perspective on neural networks and statistical models 张　政：Nonparametric Estimation of the Continuous Treatment Effect with Measurement Error	刘林 （线上）	映塔 A 会议室 （二层）
	D2 专题：AI+复杂信息系统：机器学习在 ICT 领域的机遇与挑战 （腾讯会议：874-286-612，密码：220819） 程祥乐：Flow Neural Network and Beyond for AI-Native Network 范起瑞：机器学习方法和技术在光纤通信中的应用（线上） 罗　龙：机器学习在光通信中的应用 缪　赟：AI+光感知（线上） 孙赵亿：AI+数据无损压缩（线上） 田　洋：漫谈神经网络的统计物理：从 NTK、最优输运到非平衡态热力学（线上） 周　李：人工智能在算力优化领域的机遇和挑战（线上）	孙杰 （线上）	观湖 D 会议室 （二层）
	D3 专题：材料科学中的机器学习方法 （腾讯会议：489-548-013） 欧阳润海：Improving Symbolic Regression for Predicting Materials Properties with Iterative Variable Selection（线上） 戴付志：DP 方法在材料研究中的应用 戴佳钰：极端条件材料的结构演化（线上） 薛德祯：主动学习在材料开发中的应用（线上）	王涵	观湖 A 会议室 （二层）
	D4 专题：机器学习与燃烧数值模拟（二） （腾讯会议：275-603-342） 王建春：基于数据驱动的可压缩湍流的反卷积模型（线上） 王海鸥：机器学习在湍流燃烧中的应用 万凯迪：基于卷积神经网络反应模型的湍流燃烧数值模拟 朱　通：基于深度学习实现碳氢燃料燃烧的反应分子动力学模拟 陈东平：机器学习在含能材料反应动力学中的应用	韩旺 张天汉	映塔 B 会议室 （二层）

	D5 专题：机器学习辅助分子模拟的新进展 （腾讯会议：788-325-696） 陈　翔：The Electrolyte Project 李　贺：DeepH: 深度学习 DFT 哈密顿量 李文菲：DeePKS+ABACUS：AI 辅助的电子结构方法 王亦楠：Mg-Y 合金深度学习势构建及锥面刃位错分解滑移反应研究	张林峰	观湖 B 会议室（二层）

四、邀请报告

1. 许志钦

简介：上海交通大学自然科学研究院/数学科学学院长聘教轨副教授。2012 年本科毕业于上海交通大学致远学院。2016 年博士毕业于上海交通大学，获应用数学博士学位。2016 年至 2019 年，在纽约大学阿布分校和柯朗研究所做博士后。主要研究方向是机器学习和计算神经科学。论文发表于 JMLR, AAAI, NeurIPS, CiCP 等学术期刊和会议。现为 Journal of Machine Learning 的 managing editor.

报告题目："Simple" implicit regularizations in deep learning

摘要：为什么神经网络的模型看起来如此复杂却通常能泛化好？为了理解这个问题，我们研究深度学习的训练，并发现其中一些偏向简单的隐式正则化效应。首先是神经网络从低频到高频学习的频率原则。这次报告，我将展示一些基于频率原则的理解和应用。其次是参数的凝聚效应，其使大网络有效神经元数目远少于实际神经元数目。我将从能量景观的嵌入原则和梯度流理解凝聚的机制。我们进一步推导 Dropout 的隐式正则化，发现其能促进凝聚的形成，限制模型拟合中的复杂度。这些隐式正则化都体现神经网络在训练过程有偏向用简单函数拟合数据的特点，以达到较好的泛化。最后，我将讨论一些神经网络中重要的理论问题。

2. 吴昊

简介：吴昊于 2002 年和 2007 年自清华大学获计算机科学与技术专业学士和博士学位，而后赴德国柏林自由大学数学系从事博士后工作，并于 2017 年至 2018 年间兼任柏林 Zuse 研究所的独立研究组组长。另外，在德期间以主持人身份承担德国科学基金会 DFG 与柏林爱因斯坦基金会的科研项目各一项。2018 年获国家级青年人才计划支持，进入同济大学数学科学学院工作。主要从事计算数学、机器学习与计算分子生物学的交叉研究。具体方向包括随机过程的算子理论以及生物大分子的机器学习分析方法等。在知名期刊（包括 Science、Nat. Commun. 和 PNAS 等顶级期刊）和机器学习知名会议（NIPS、UAI 和 MSML 等）发表论文 30 余篇。

报告题目：多体系统运动学降阶的机器学习方法

摘要：以分子动力学为代表的多体系统模拟技术在科学与工程的诸多领域中都扮演着极为重要的角色。而随着高性能计算、云计算的飞速发展以及大规模模拟算法的日益进步，多体系统模拟所能达到系统空间尺度和维度不断提高。在这一发展趋势下，如何利用模拟数据对高维多体系统运动学实现有效降维，并在此基础上作进一步的分析与建模，就成为了一个挑战性的问题。在本报告中，我们将从该问题的数学描述出发，讨论目前被广泛使用的转移算子方法和近年来兴起的基于信息论的方法，并重点介绍深度学习技术相关研究中所发挥的重要作用。

3. 张林峰

简介：张林峰博士本科毕业于北京大学元培学院，博士毕业于普林斯顿大学。他在 AI for Science 相关领域成果丰富，曾在 SCI 索引期刊及机器学习顶会发表论文 40 余篇，并发展了 AI for Science 领域最大的开源社区 DeepModeling。2020 年，林峰作为核心开发者的工作获得高性能计算领域最高奖 ACM 戈登贝尔奖（GORDON BELL PRIZE）。他创立的深势科技致力于运用人工智能和分子模拟算法，结合先进计算手段，为人类文明最基础的生物医药、能源、材料和信息科学与工程需求打造新一代微尺度工业设计和仿真平台。同时，他也在作为北京科学智能研究院学术副院长，推动 AI for Science 开源基础设施建设和前沿科学问题研究。

报告题目：AI for Science 开源社区发展的实践与思考

摘要：AI for Science 正在成为学界和业界极为关注的话题。它将有望推动科研范式和产业形态的新一轮升级，同时也面临着很多挑战。在本报告中，我将分享在推动 DeepModeling——一个致力于打造

AI for Science 时代基础设施平台的开源社区——过程中的实践与思考。我将着重探讨算法、软件、和应用场景如何融合，“开源模式”在这一语境下的必要性和困难，及其对算力基础设施的系统需求。此外，我也将结合在深势科技和北京科学智能研究院的实践，分享一系列从创新到落地的实际案例。

4. 史作强

简介：清华大学丘成桐数学科学中心副教授，北京雁栖湖应用数学研究院兼职研究员，主要研究方向为偏微分方程数值方法，图像处理和机器学习中的微分方程模型，非线性非平稳信号时频分析等，在 ACHA，SIAM 系列期刊，Advances in Mathematics，ARMA 等国际知名学术期刊发表文章 50 余篇。

报告题目：半监督学习的微分方程模型

摘要：半监督学习是机器学习重要的研究领域，对于少量标注数据的问题具有重要的意义。在本报告中，我们将探讨半监督学习中的各类微分方程模型，包括 Laplace 方程，无穷 Laplace 方程，双调和方程，对流扩散方程，通过理论分析和实际算例展示不同微分方程模型的应用范围。

5. 王立威

简介：北京大学教授。长期从事机器学习理论研究。在机器学习国际权威期刊会议发表高水平论文 150 余篇。担任机器学习与计算机视觉顶级期刊 IEEE TPAMI 编委。多次担任国际机器学习旗舰会议 NeurIPS，ICML，ICLR 领域主席与高级领域主席。入选 AI’s 10 to Watch，是该奖项自设立以来首位获此荣誉的中国学者。

报告题目：L2 is not the right loss for PINN when solving nonlinear PDE

摘要：The Physics-Informed Neural Network (PINN) approach is a new and promising way to solve partial differential equations using deep learning. The L2 Physics-Informed Loss is the de-facto standard in training Physics-Informed Neural Networks. In this paper, we challenge this common practice by investigating the relationship between the loss function and the approximation quality of the learned solution. In particular, we leverage the concept of stability in the literature of partial differential equation to study the asymptotic behavior of the learned solution as the loss approaches zero. With this concept, we study an important class of high-dimensional non-linear PDEs in optimal control, the Hamilton-Jacobi-Bellman(HJB) Equation, and prove that for general Lp Physics-Informed Loss, a wide class of HJB equation is stable only if p is sufficiently large. Therefore, the commonly used L2 loss is not suitable for training PINN on those equations, while $L\infty$ loss is a better choice. Based on the theoretical insight, we develop a novel PINN training algorithm to minimize the $L\infty$ loss for HJB equations which is in a similar spirit to adversarial training. The effectiveness of the proposed algorithm is empirically demonstrated through experiments.

五、分会报告 A1 专题：图像科学中的机器学习方法——前沿进展及挑战（一）

A1-1 Deep Fast MR Imaging: Deep Learning meets Nuclear Spins

梁栋，中国科学院深圳先进技术研究院

摘要：磁共振成像是目前医学成像技术中功能最强大，技术门槛最复杂的技术之一。然而，相对其他成像模态较为缓慢的采集速度（较长的扫描时间）制约了磁共振成像在临床上的广泛应用。借助采样样本的数量与扫描时间的正比关系，通过减少采集的数据量（稀疏采样）来缩短磁共振采集时间是快速磁共振成像领域的研究热点。过去十年基于压缩感知的快速成像技术极大的提高了磁共振的采集速度，但是同时也面对着很多挑战如重建时间长、参数选择困难，丢失细节等。近年来，借助深度学习的强大学习能力来解决上述问题已经成为研究热点。本次报告将介绍这一领域的前沿进展及面临的挑战。

A1-2 重思图像复原的基本方法论

孟德宇，西安交通大学

摘要：针对图像复原，如遥感图像复原，问题，传统方法论主要分为模型驱动与数据驱动两类。其中模型驱动主要通过认识数据，预先设计合理的损失与正则项，从而达到良好复原效果。而数据驱动主要通过借鉴计算机视觉领域通用有效深度网络的构建技巧，通过端到端机器学习的方式来获得针对退化遥感图像的显式复原函数，从而便于泛化使用。然而，针对遥感图像的特殊内涵，两种方法论均存在内在的缺陷。本报告中，将尝试对已有底层图像技术进行内在功能的分析，从而反思其局限性，进而讨论如何对图像复原问题能够更加合理设计方法论的可能策略。

A1-3 超图计算

高跃，清华大学

摘要：本报告主要介绍超图计算理论及方法。超图是一种广义的图结构，因其具有较强的数据样本间非线性高阶关联的刻画和挖掘能力而被广泛应用于数据分类、检索等任务中。本报告将介绍超图计算的基础理论及方法，包括超图结构建模、超图神经网络和超图高效计算。最后介绍了超图计算在计算机 4 视觉、数据挖掘和智慧医疗等领域的应用。

A1-4 Non-line-of-sight Imaging

邱凌云，清华大学

Abstract： Non-line-of-sight imaging aims at recovering obscured objects from multiple scattered light. It has recently received widespread attention due to its potential applications such as autonomous driving, rescue operations, and remote sensing. However, in cases with high measurement noise, obtaining high-quality reconstructions remains a challenging task. In this work, we establish a unified regularization framework, which can be tailored for different scenarios, including indoor and outdoor scenes with substantial background noise under both confocal and non-confocal settings. The proposed regularization framework incorporates sparseness and non-local self-similarity of the hidden objects as well as smoothness of the measured signals. We show that the estimated signals, albedo, and surface normal of the hidden objects can be reconstructed robustly even with high measurement noise under the proposed framework. Reconstruction results on synthetic and experimental data show that our approach recovers the hidden objects faithfully and outperforms state-of-the-art reconstruction algorithms in terms of both quantitative criteria and visual quality.

六、分会报告 A2 专题：复杂度和泛化理论

A2-1 Deep Learning Dynamics: On Minima Selection and Saddle-Point Escaping

谢泽柯，东京大学

Abstract： Stochastic Gradient Descent (SGD) and its variants are mainstream methods for training deep networks in practice. SGD is known to find a flat minimum that generalizes well. However, the theoretical mechanism behind stochastic optimization for deep networks is underexplored. Our work focuses on theoretically analyzing deep learning dynamics and designing novel optimization dynamics by using a proposed diffusion theoretical framework. We revealed how minima selection quantitatively depends on the minima sharpness and the hyperparameters. To the best of our knowledge, we are the first to prove that SGD favors flat minima exponentially more than sharp minima, while Gradient Descent with injected Gaussian noise favors flat minima only polynomially more than sharp minima. We further prove that Adam can escape saddle points efficiently, but cannot select flat minima as SGD does. This mathematically explains why SGD generalizes better, while Adam generalizes worse but converges faster. Inspired by the theoretical analysis, we propose a novel adaptive optimization framework, called Adaptive Inertia (Adai), which uses parameter-wise momentum hyperparameters. The proposed Adai provably escapes saddle points efficiently and favors flat minima as well as SGD.

A2-2 Neural Network Weights Do Not Converge to Stationary Points: An Invariant Measure Perspective

张景昭，清华大学

Abstract: This work examines the deep disconnect between existing theoretical analyses of gradient-based algorithms and the practice of training deep neural networks. Specifically, we provide numerical evidence that in large-scale neural network training (e.g., ImageNet + ResNet101, and WT103 + TransformerXL models), the neural network's weights do not converge to stationary points where the gradient of the loss is zero. Remarkably, however, we observe that even though the weights do not converge to stationary points, the progress in minimizing the loss function halts and training loss stabilizes. Inspired by this observation, we propose a new perspective based on ergodic theory of dynamical systems to explain it. Rather than studying the evolution of weights, we study the evolution of the distribution of weights. We prove convergence of the distribution of weights to an approximate invariant measure, thereby explaining how the training loss can stabilize without weights necessarily converging to stationary points. We further discuss how this perspective can better align optimization theory with empirical observations in machine learning practice.

A2-3 Error analysis of GAN

焦雨领，武汉大学

Abstract: I will talk about error of GAN.

A2-4 PINN 的失效与其隐式正则化

周岐轩，上海交通大学

摘要： 伴随机器学习新浪潮，以 DeepRitz 与 PINN 为代表的神经网络数值求解偏微分方程迅速流行。然而（在不超过 3 维的物理问题上）其表现堪忧：实践上，鲜有比肩传统数值方法的实例；理论上，缺乏稳定、收敛性证明。在这个报告中，我们将讨论 PINN 等方法在有间断问题上的表现。首先，对带有间断系数的椭圆型 PDE，我们通过实验展示 PINN 并不能成功逼近方程的真实解。我们通过引入辅助方程的手段，证明 PINN 对带间断系数的椭圆型方程的解的拟合是失效的。进一步，我们指出这

种失效中依然带有一系列规律，也即意味着 PINN 方法在此类问题上具有的隐式正则化。最后，我们将所有结果扩展到拟线性椭圆 PDE 上。

七、分会报告 A3 专题：科学计算与机器学习

A3-1 Learning Robust Imaging Models without Paired Data

包承龙，清华大学

Abstract: The observations in practical imaging systems always contain complex noise such that classical approaches are difficult to obtain satisfactory results. In recent years, deep neural networks directly learned a map between the noisy and clean images based on the training on paired data. Despite its promising results in various tasks, collecting the training data is difficult and time-consuming in practice. In this talk, in the unpaired data regime, we will discuss our recent progress for building AI-aided robust models and their applications in image processing. Leveraging the Bayesian inference framework, our model combines classical mathematical modeling and deep neural networks to improve interpretability. Experimental results on various real datasets validate the advantages of the proposed methods.

A3-2 A stochastic three-block splitting algorithm and its application to quantized deep neural networks

张小群，上海交通大学

Abstract: Deep neural networks (DNNs) have made great progress in various fields. In particular, the quantized neural network is a promising technique making DNNs compatible on resource-limited devices for memory and computation saving. In this work, we consider a nonconvex minimization model with three blocks to train quantized DNNs and propose a new stochastic three-block alternating minimization (STAM) algorithm to solve it. We develop a convergence theory for the STAM algorithm and obtain an ξ\epsilonξ-stationary point with optimal convergence rate ξ\mathcal{O} (\epsilon^{-4})ξ. The experiments on training quantized DNNs are carried out on different network structures on CIFAR-10 and CIFAR-100 datasets. The test accuracy indicates the effectiveness of STAM algorithm for training binary quantization DNNs.：

A3-3 Classification with Deep Neural Networks

石磊，复旦大学

Abstract: Classification with deep neural networks (DNNs) has made impressive advancements in various learning tasks. Due to the unboundedness of the target function, generalization analysis for DNN classifiers with logistic loss remains scarce. This talk will report our recent progress in establishing a unified framework of generalization analysis for both bounded and unbounded target functions. Our analysis is based on a novel oracle-type inequality, which enables us to deal with the boundedness restriction of the target function. In particular, for logistic classifiers trained by deep fully connected neural networks, we obtain the polynomial convergence rates only by requiring the H\"{o} lder smoothness of the conditional probability. Under certain circumstances, such as when decision boundaries are smooth and the two classes are separable, the derived convergence rates can be independent of the input dimension. This talk is based on joint work with Zihan Zhang and Prof. Ding-Xuan Zhou.

A3-4 Deep-learning ab initio calculation method: DeepH

徐勇，清华大学

Abstract: Deep-learning ab initio calculation method: DeepH

八、分会报告 A4 专题：最优输运理论、计算和应用

A4-1 Importance Sparsification for Sinkhorn Algorithm

孟澄，中国人民大学

Abstract: Sinkhorn algorithm has been used pervasively to approximate the solution to optimal transport (OT) and unbalanced optimal transport (UOT) problems. However, its application in practice is limited due to its high computational complexity. To alleviate the computational burden, we propose a novel importance sparsification method, called Spar-Sink, to approximate regularized OT and UOT distances efficiently. Specifically, we leverage natural upper bounds for unknown optimal transport plans to construct effective sampling probabilities, and construct a sparse kernel matrix to accelerate Sinkhorn iterations, reducing the cost from O(n^2) to O(n) for a sample of size n. Theoretically, we show the proposed estimators for the regularized OT and UOT distances are consistent under mild regularity conditions. Experiments on various synthetic datasets demonstrate Spar-Sink leads to smaller estimation errors compared with mainstream competitors.

A4-2 Learning geodesics under spherical WFR metric and its application to generation of weighted samples

李磊，上海交通大学

Abstract: We present a framework for learning geodesics under spherical WFR metric. A loss based on the KL divergence using the inverse mapping is adopted to overcome the difficulty brought by weight change. The framework can be used for generating weighted samples which could be useful in Bayes setting.

A4-3 PROXIMAL POINT METHODS FOR COMPUTING (UN-)BALANCED WASSERSTEIN DISTANCE

包承龙，清华大学

Abstract: In this talk, we discuss a general model which unifies both balanced and unbalanced optimal transport problems. Then we propose two sorts of fast and accurate proximal point algorithms to solve this generalized optimal transport problem: one is the Primal proximal point algorithm (Primal PPA), another is the Dual proximal point algorithm (Dual PPA). We shall prove the asymptotically superlinear convergence rate of these algorithms under certain assumptions which are satisfied in the cases of balanced OT and unbalanced OT with KL divergence by utilizing the concept of metric subregularity. We also propose a semismooth Newton method to solve the subproblem, which fully exploits the sparse structure of our problem and obtains a quadratic convergence rate with relatively low computational cost. We compare our algorithms with gurobi and Sinkhorn algorithm to demonstrate the fast convergence speed with the high accuracy of our algorithms.

A4-4 最优输运在语义信息匹配中的应用

王子豪，香港科技大学

摘要：语义信息广泛存在与语言和知识的各个层次，匹配语义信息有重要意义。最优输运作为匹配概率分布的重要方法，可以应用在词，句子，知识等层次上。本报告介绍集中不同层次的语义信息匹配问题，以及最优输运方法在其中的应用。

九、分会报告 A5 专题：机器学习与计算数学（一）

A5-1 Theory and algorithms on archetype analysis

王东，香港中文大学（深圳）

Abstract: Archetypal analysis is an unsupervised learning method that uses a convex polytope to summarize multivariate data. For fixed k, the method finds a convex polytope with k vertices, called archetype points, such that the polytope is contained in the convex hull of the data and the mean squared distance between the data and the polytope is minimal. In this talk, we will discuss the consistency of archetypal analysis and describe probabilistic methods for approximate archetypal analysis.

A5-2 Uncertainty Quantification in Scientific Machine Learning: Methods and Comparisons

郭玲，上海师范大学

Abstract: 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. Although there are some recent works on uncertainty quantification (UQ) in NNs, there is no systematic investigation of suitable methods towards quantifying the total uncertainty effectively and efficiently even for function approximation, and there is even less work on solving partial differential equations and learning operator mappings between infinite-dimensional function spaces using NNs. In this talk, we will present a comprehensive framework that includes uncertainty modeling, new and existing solution methods, as well as evaluation metrics and post-hoc improvement approaches. To demonstrate the applicability and reliability of our framework, we will also present an extensive comparative study in which various methods are tested on prototype problems, including problems with mixed input-output data, and stochastic problems in high dimensions.

A5-3 Radial Basis Functions for Approximating Differential Operators on Closed Manifolds

蒋诗晓，上海科技大学

Abstract: In this talk, we will introduce a class of kernel methods, Radial Basis Functions (RBF), to approximate Laplace-Beltrami, covariant derivative, and Bochner, Hodge, Lichnerowicz Laplacians of vector fields defined on smooth manifolds without boundaries. Each operator is estimated using an ambient space formulation followed by a projection onto the local tangent space of the manifold using a tangential projection tensor matrix P. When the manifold is unknown and identified only by a point cloud data, we present a novel second-order SVD method for approximating the projection matrix P. All Laplacian operators have two natural discrete estimators: symmetric and non-symmetric formulations, and each formulation has its own practical and theoretical advantages. We establish the spectral convergence for both formulations of Laplacian operators. Numerically, we validate the theory via various examples of simple manifolds embedded in Euclidean spaces as well as of unknown manifolds. Moreover, we will briefly discuss some possible applications of the RBF methods, including solving PDEs, model order reduction, parameter estimation.

A5-4 谱方法与深度神经网络逼近

于海军，中国科学院数学与系统科学研究院

摘要： 整流线性单元(ReLU)是深度网络中最常用的的激活函数. ReLU 网络由于易于训练, 且具有较好的逼近性质，在图像处理等领域得到广泛应用。但是导数的不连续性限制了其在一些对光滑性要求较高场合的应用，比如求解偏微分方程的深度 Ritz 方法和 PINN 方法等。针对此问题，我们分析了整流

幂次单元(RePU)网络的逼近性质。通过使用一类稳定的构造算法，我们揭示了深度 RePU 网络同谱方法多项式逼近之间的紧密关联。针对 RePU 非全局 Lipschitz 连续的问题，我们提出了正则化的整理幂次单元(RePUr), 并在几类典型问题，包括函数回归，微分方程学习和求解, 模型降维等问题，验证了 RePU 和 RePUr 网络的优越性。

十、分会报告 A6 专题：语言智能分论坛

A6-1 大模型背景下可信自然语言处理

秦兵，哈尔滨工业大学

摘要：当前，大规模预训练自然语言理解模型已经在多个自然语言理解任务上面取得了令人印象深刻的性能，成为自然语言理解的新范式。然而，当前大模型给出的自然语言处理结果在可信程度上还比较薄弱，主要体现在三个方面：模型的稳定性差、可解释性弱、泛化能力不足。本次报告介绍我们如何在大模型背景下进行可信自然语言理解的问题。首先，针对大模型稳定性差的问题，介绍因果机制引导的稳定自然语言理解方法；其次，针对大模型可解释性弱的问题，介绍基于神经符号的可解释自然语言理解方法。

A6-2 多语言机器翻译中知识迁移的不对称性问题研究

张家俊，中国科学院自动化研究所

摘要：多语言机器翻译旨在设计一个模型可以同时翻译多种语言，不仅提升了训练和部署的效率，同时也便利了语言之间的知识迁移。我们研究发现多语言机器翻译中的知识迁移存在不对称性，也就是翻译任务 A 与翻译任务 B 共同训练时，A 可以提升 B 的性能，但是 B 可能会降低 A 的性能。这种知识迁移的不对称性问题在多语言机器翻译中尚未得到充分研究，针对该问题，我们将介绍一种基于模糊聚类的多语言机器翻译方法。

A6-3 开放域对话的语义信息流控制与一致性检测

冯洋，中国科学院计算技术研究所

摘要：随着高质量对话数据的不断构建以及大规模预训练语言模型的广泛应用，开放域对话取得了巨大的进步。开放域对话需要兼顾语义逻辑以及文字表述，由于对话的灵活性，其在语义逻辑上面临着更大的挑战，包括保证前后文的连贯性以及一致性。针对这些挑战，我们提出了引入高层语义信息流的开放域对话回复生成方法，同时基于该方法构建语义流预训练模型用于对话评估，并进一步提出了基于问询的开放域对话一致性评估框架，以此来保证对话的连贯性和一致性。

A6-4 面向用户模拟与满意度的任务导向推荐系统评估

任昭春，山东大学

摘要：作为对话系统开发过程中的重要部分，任务导向对话系统的评估问题受到广泛的关注。现有的方法主要使用离线测试和人工评测，存在单轮评估的限制以及花费过高的问题。针对这一问题，我们提出面向用户模拟与满意度的对话系统自动评估方法。首先，我们提出用户满意度模拟任务，通过对用户对话过程中满意度的模拟提升自动评估结果的和人工评测的一致性。在此基础上，我们提出了基于隐喻的用户模拟器，通过对用户心智模型的建模提升用户模拟器的真实性。此外，我们设计了基于测试器的评估框架，以实现可靠的用户模拟器自动评估。我们在多个任务导向对话数据集上进行了实验，验证了研究的有效性。

十一、分会报告 B1 专题：图像科学中的机器学习方法—前沿进展及挑战（二）

B1-1 3D Model reconstruction from 1D Small angle X-ray Scattering data

刘海广，微软研究院科学智能中心

Abstract: X-ray Solution scattering methods are widely used in the study of structure and dynamics of macromolecules. Translating the 1D SAXS curve to 3D density map is an ill-defined problem, yet it is possible to obtain 3D models given proper restraints as additional information. An auto-encoder neural network model for 3D protein density maps was trained to compress 3D shape information into vectors in a low-dimension latent space, and the vectors are optimized using genetic algorithms to build 3D models that are consistent with the scattering data. The algorithm was implemented using Python with the TensorFlow framework and tested with experimental data, demonstrating capacity and robustness of accurate model reconstruction even without using prior model size information.

B1-2 Deep Learning in Non-Euclidean Space

孙剑，西安交通大学

Abstract: The traditional deep networks are commonly defined in Euclidean space, either in the 3D / 2D image space or sequential data space. However, in realistic scenario, the data maybe irregular or distributed on manifold / graph. In such cases, the traditional deep network does not fully take advantages of the underlying data structure in non-Euclidean space. Along this research direction, in this talk, I will introduce the research backgrounds, advances in research on geometric deep learning approach in the non-Euclidean space, with applications to 3D object recognition, image segmentation and domain adaptation.

B1-3 基于机器学习的智能成像研究进展

张意，四川大学

摘要：近年来，机器学习在计算机视觉、图像处理领域已经取得了丰硕的成果，但是在医学成像领域，主要的应用研究仍然集中在图像分析，在成像领域的研究才刚刚起步。同时，这些技术在推广过程中，面临泛化性、可解释性等一系列问题。本报告主要介绍针对这些问题，课题组在该领域的一系列工作及研究进展，主要包括低剂量 CT 重建、快速 MRI 重建等相关工作，最后将展望基于机器学习的智能成像方法未来发展的趋势。

B1-4 TBA

全宇晖，华南理工大学

十二、分会报告 B2 专题：反问题中的机器学习方法

B2-1 Imaging Conductivity from Current Density Magnitude using Neural Networks

吕锡亮，武汉大学

Abstract: In this talk we propose a deep neural network based reconstruction technique for imaging the conductivity from the magnitude of the internal current density. It is achieved by formulating the problem as a relaxed weighted least-gradient problem, and then approximating its minimizer by standard fully connected deep neural networks. The approximation error and statistical error, explicitly in terms of properties of the neural networks (e.g., depth, total number of parameters, and the bound of the network parameters) are given. Several numerical experiments are proposed to show robustness of the method.

B2-2 Low-rank Methods for Bayesian Inverse Problems

邱越，上海科技大学

Abstract: In this talk, I will introduce our recent work on low-rank methods for Bayesian inverse problems. For linear problems with Gaussian noise and Gaussian prior, the posterior is also Gaussian and characterized by the posterior mean and covariance. We propose a low-rank Arnoldi method to approximate the large dense posterior covariance matrix by making use of tensor computations. For nonlinear systems, the posterior is not Gaussian anymore, however, can often be approximated by a Gaussian distribution using the ensemble Kalman filter (EnKF) or the extended Kalman filter (ExKF). We propose a randomized low-rank method to reduce the computational complexity of the EnKF. We use numerical experiments to show the efficiency of our low-rank methods.

B2-3 Deep learning approach for Bayesian inverse Problems

闫亮，东南大学

Abstract: Obtaining samples from the posterior distribution of Bayesian inverse problems(BIPs) is a long-standing challenging, especially when the forward operator is modeled by partial differential equation (PDE). In this talk, we will show you how to leverage the deep learning's capabilities to tackle this challenge. Several fast and efficient deep neural network (DNN)-based approaches for accelerating simulations in sample generation will be described. A novel framework based on invertible neural networks using normalizing flow is also demonstrated.

B2-4 Weak Adversarial Networks: A Deep Learning Framework for Solving High Dimensional Inverse Problems

臧耀华，华为杭州研究所

Abstract: We present a weak adversarial network approach to numerically solve a class of inverse problems, including electrical impedance tomography. The weak formulation of the PDE for the given inverse problem is leveraged, where the solution and the test function are parameterized as deep neural networks. Then, the weak formulation and the boundary conditions induce a minimax problem of a saddle function of the network parameters. As the parameters are alternatively updated, the network gradually approximates the solution of the inverse problem. Theoretical justifications are provided on the convergence of the proposed algorithm. The proposed method is completely mesh-free without any spatial discretization, and is particularly suitable for problems with high dimensionality and low regularity on solutions. Numerical experiments on a variety of test inverse problems demonstrate the promising accuracy and efficiency of this approach. This presentation is based on the joint work with Gang Bao (Zhejiang U.), Xiaojing Ye (Georgia State U.) and Haomin Zhou (Georgia Tech.)

十三、分会报告 B3 专题：机器学习中的优化问题（一）

B3-1 Algorithmic Design for Wasserstein DRO Based Trustworthy Machine Learning

陈彩华，南京大学

Abstract: Wasserstein Distributionally Robust Stochastic Optimization (DRSO) is concerned with finding decisions that perform well on data that are drawn from the worst-case probability distribution within a Wasserstein ball centered at a certain nominal distribution. In recent years, it has been shown that various DRSO formulations of learning models admit tractable convex reformulations. However, most existing works propose to solve these convex reformulations by general-purpose solvers, which are not well-suited for tackling large-scale problems. In this talk, we focus on Wasserstein distributionally robust support vector machine (DRSVM) problems (with fairness) and logistic regression (DRLR) problems (with fairness), and propose two novel first order algorithms to solve them. The updates in each iteration of these algorithms can be computed in a highly efficient manner. Our numerical results indicate that the proposed methods are orders of magnitude faster than the state-of-the-art, and the performance gap grows considerably as the problem size increases.

B3-2 Decentralized Optimization Over the Stiefel Manifold by an Approximate Augmented Lagrangian Function

刘歆，中国科学院数学与系统科学研究院

Abstract: We study the decentralized optimization problem over the Stiefel manifold, which is defined on a connected network of d agents. The objective is an average of d local functions, and each function is privately held by an agent and encodes its data. The agents can only communicate with their neighbors in a collaborative effort to solve this problem. In existing methods, multiple rounds of communications are required to guarantee the convergence, giving rise to high communication costs. In contrast, this paper proposes a decentralized algorithm, called DESTINY, which only invokes a single round of communications per iteration. DESTINY combines gradient tracking techniques with a novel approximate augmented Lagrangian function. The global convergence to stationary points is rigorously established. Comprehensive numerical experiments demonstrate that DESTINY has a strong potential to deliver a cutting-edge performance in solving a variety of testing problems.

B3-3 Convergence properties of the stochastic Levenberg-Marquardt method

范金燕，上海交通大学

Abstract: In this talk, we propose a stochastic Levenberg-Marquardt algorithm based on trust region. We show that the trust region subproblem defined is probabilistically first-order accurate when the batch size is appropriate. The algorithm is proved to converge globally and the complexity of the algorithm is also given.

B3-4 Perturbation analysis of nonsmooth optimization on manifolds

丁超，中国科学院数学与系统科学研究院

Abstract: In this talk, we will present some recent results on perturbation analysis of nonsmooth optimization on manifolds with their applications on the convergent analysis of manifold augmented Lagrangian method for nonsmooth optimization on matrix manifolds.

十四、分会报告 B4 专题：机器学习与计算数学（二）

B4-1 Random feature method for solving PDEs

陈景润，中国科学技术大学

Abstract: We will introduce the random feature method for solving partial differential equations, which combines the advantages of traditional and machine learning-based methods. Its performance will be demonstrated through a series of problems with or without explicit solutions and with or without complex geometries.

B4-2 Near-Optimal Bounds for Generalized Orthogonal Procrustes Problem via Generalized Power Method

凌舒扬，上海纽约大学

Abstract: Given multiple point clouds, how to find the rigid transform (rotation, reflection, and shifting) such that these point clouds are well aligned? This problem, known as the generalized orthogonal Procrustes problem (GOPP), has found numerous applications in statistics, computer vision, and imaging science. While one commonly-used method is finding the least squares estimator, it is generally an NP-hard problem to obtain the least squares estimator exactly due to the notorious nonconvexity. In this work, we apply the semidefinite programming (SDP) relaxation and the generalized power method to solve this generalized orthogonal Procrustes problem. In particular, we assume the data are generated from a signal-plus-noise model: each observed point cloud is a noisy copy of the same unknown point cloud transformed by an unknown orthogonal matrix and also corrupted by additive Gaussian noise. We show that the generalized power method (equivalently alternating minimization algorithm) with spectral initialization converges to the unique global optimum to the SDP relaxation, provided that the signal-to-noise ratio is high. Moreover, this limiting point is exactly the least squares estimator and also the maximum likelihood estimator. In addition, we derive a block-wise estimation error for each orthogonal matrix and the underlying point cloud. Our theoretical bound is near-optimal in terms of the information-theoretic limit (only loose by a factor of the dimension and a log factor). Our results significantly improve the state-of-the-art results on the tightness of the SDP relaxation for the generalized orthogonal Procrustes problem, an open problem posed by Bandeira, Khoo, and Singer in 2014.

B4-3 A Local Deep Learning Method for Solving High Order Partial Differential Equations

杨将，南方科技大学

Abstract: Recent years, deep learning based methods are being employed to resolve the computational challenges of high-dimensional partial differential equations (PDEs). But the computation of the high order derivatives of neural networks is costly, and high order derivatives lack robustness for training purposes. We propose a novel approach to solve PDEs with high order derivatives by simultaneously approximating the function value and derivatives. We first introduce intermediate variables to rewrite the PDEs into a system of low order differential equations as what is done in the local discontinuous Galerkin method. The intermediate variables and the solutions to the PDEs are simultaneously approximated by a multi-output deep neural network. The whole process only relies on low order derivatives. We also derive a priori error estimate of the local deep Deep Learning method when solving some elliptic PDEs. We prove that the neural network solutions will converge if we increase the training samples and network size without any constraint on the ratio of training samples to the network size. Besides, our results suggest that the mixed residual method can recover high order derivatives better than the deep Ritz method. Numerous numerical examples are carried out to demonstrate that our local deep learning is efficient, robust, flexible, and is particularly well-suited for high-dimensional PDEs with high order derivatives.

B4-4 Distance geometry, quadratic form and parameter-turning for graph matching

朱圣鑫，北京师范大学珠海校区

Abstract: We will discuss some recent progress in graph matching problem.

十五、分会报告 B5 专题：机器学习与张量网络

B5-1 Revisit Tensor Networks Machine Learning Modeling

程嵩，北京雁栖湖应用数学研究院

Abstract: The rich connection between machine learning and tensor networks was revealed in recent years, while it was always intriguing that these two independently developed fields could be successfully combined. This talk will provide a partial review and general description of the development of interdisciplinary research between those two fields. The content will primarily focus on tracking down where their connections come from and what we can do with those connections.

B5-2 粗粒化结构在神经网络中的应用

谢志远，中国人民大学物理系

摘要：基于图像的局域关联特性，我们将重正化群的粗粒化结构引入到神经网络的构造中，在图像分类问题中得到了一些初步的成功结果。如果时间允许，本报告也将介绍该结构在数据蒸馏中的一些初步应用。

B5-3 Functional tensor network solving many-body Schrödinger equation and multi-variable differential equations

冉仕举，首都师范大学

Abstract: We propose the functional tensor network (FTN) approach to solve the many-body Schrodinger equation and generally the partial differential equations of many variables. Provided the orthonormal functional bases, we represent the coefficients of the many-body wave-function as tensor network. The observables, such as energy, can be calculated simply by tensor contractions. Simulating the ground state becomes solving a minimization problem defined by the tensor network. An efficient gradient-decent algorithm based on the automatically differentiable tensors is proposed. We here take matrix product state (MPS) as an example, whose complexity scales only linearly with the system size. We apply our approach to solve the ground state of coupled harmonic oscillators, and achieve high accuracy by comparing with the exact solutions. Reliable results are also given with the presence of three-body interactions, where the system cannot be decoupled to isolated oscillators. Our approach is simple and with well-controlled error, essentially different from the highly-nonlinear neural-network solvers. Our work extends the applications of tensor network from quantum lattice models to the systems in the continuous space. FTN can be used as a general solver of the differential equations with many variables.

B5-4 Thermal Tensor Networks: Recent Progress and Application to Quantum Magnetism

李伟，中科院理论物理研究所

Abstract: In this talk, I will briefly review the recent progress in thermal tensor networks for finite-temperature quantum many-body simulations, and its applications in quantum magnetic material studies.

十六、分会报告 B6 专题：机器学习与神经科学

B6-1 A brain-inspired method for motion pattern recognition

弭元元，重庆大学

Abstract: Spatio-temporal information processing is fundamental in both brain functions and AI applications. Current strategies for spatio-temporal pattern recognition usually involve explicit feature extraction followed by feature aggregation, which requires a large amount of labeled data. In the present study, motivated by the subcortical visual pathway and early stages of the auditory pathway for motion and sound processing, we propose a novel brain-inspired computational model for generic spatio-temporal pattern recognition. The model consists of two modules, a reservoir module and a decision-making module. The former projects complex spatio-temporal patterns into spatially separated neural representations via its recurrent dynamics, the latter reads out neural representations via integrating information over time, and the two modules are linked together using known examples. Using synthetic data, we demonstrate that the model can extract the frequency and order information of temporal inputs. We apply the model to reproduce the looming pattern discrimination behavior as observed in experiments successfully. Furthermore, we apply the model to the gait recognition and key words spotting tasks, and demonstrate that our model accomplishes the recognition in an event-based manner and outperforms deep learning counterparts when training data is limited.

B6-2 类脑计算——从芯片到算法

冷卢子未，华为技术有限公司

摘要：传统 AI 面临计算量大，高能耗，泛化性差，鲁棒性低等问题。类脑计算在算法和芯片层面模拟大脑原理，具有事件触发、异步计算、稀疏响应、可塑性强等特点，有望实现低时延、低能耗的在线学习。芯片层面，类脑芯片模拟大脑神经元的功能特性、信号传递和学习方式，突破传统芯片“存算分离”的冯诺依曼架构。算法层面，类脑计算可分为自上而下和自下而上两种路径。前者利用大脑处理信息的广义原理如预测编码、自由能原理、自监督学习等去更好的解决 AI 问题，后者从脉冲神经元模型、突触动力学、大脑微环路等仿脑结构出发搭建低能耗、高效网络。这些方向逐渐交叉融合，产生了很多有趣的工作。

B6-3 Smoothness in nature and smoothness in DNN

肖彦洋，中国科学院深圳先进技术研究院

Abstract: Smoothness is widely observed among natural signals such as electrode recordings and natural images. On the other hand, deep learning networks are also observed to favor lower frequency components in the signals during the training process. In this talk, we will discuss both the phenomenon from the viewpoint of orthogonal polynomial and singular value decomposition (a.k.a. principal component analysis). We find that the lower frequency affinity can be explained by the eigenvalues in the linearized gradient decent dynamics, and the magnitude of the eigenvalues depends on the smoothness of the activation function tightly. The same eigenvalue analysis could be extended to answer other related questions, we demonstrate it by describing the Loss function landscape around the optimal point.

B6-4 The representational geometry of sequence working memory in macaque prefrontal cortex

闵斌，上海脑科学与类脑研究中心

Abstract: How the brain stores a sequence in memory remains largely unknown. We investigated the neural code underlying sequence working memory using two-photon calcium imaging to record thousands of neurons in the prefrontal cortex of macaque monkeys memorizing and then reproducing a sequence of

locations after a delay. We discovered a regular geometrical organization: The high-dimensional neural state space during the delay could be decomposed into a sum of low-dimensional subspaces, each storing the spatial location at a given ordinal rank, which could be generalized to novel sequences and explain monkey behavior. The rank subspaces were distributed across large overlapping neural groups, and the integration of ordinal and spatial information occurred at the collective level rather than within single neurons. Thus, a simple representational geometry underlies sequence working memory.

十七、分会报告 C1 专题：机器学习的工业应用与挑战

C1-1 OoD 泛化与自动驾驶 Corner Cases

洪蓝青，华为诺亚方舟实验室

摘要：近年来，自动驾驶技术因其在减少事故和提高效率方面的巨大潜力而备受关注。道路图像中的物体检测作为视觉感知系统中的一个重要模块，在自动驾驶中起着至关重要的作用。尽管当前的自动驾驶模型在行人、车等常见目标的检测中取得较为准确的结果，但是对于罕见难例（如高速路上的动物，马路上的婴儿车，翻倒的卡车等）仍然难以检出，存在潜在的安全风险。对于自动驾驶 Corner Case 问题，目前尚未有一个公认的定义和描述。Corner Case 可以是对于数据而言的，也可以是对于模型而言的。如何理解和定义自动驾驶 Corner Case，定位模型弱点，并结合激光点云、多摄像头输入、时序信息等，提出通用的开放世界物体检测算法，使检测模型的泛化能力大幅提升，是一个挑战难题。本报告将介绍自动驾驶 Corner Case 公开数据集 CODA 的构建，以及相关的算法探索。

C1-2 \texttt{Para-CFlows｝：ξC^kξ-universal Diffeomorphism Approximators as Superior Neural Surrogates

吕俊龙，华为诺亚方舟实验室

Abstract: Invertible neural networks based on Coupling Flows (\texttt{CFlows｝) have various applications such as image synthesis and data compression. The approximation universality for \texttt{CFlows｝ is of paramount importance to ensure the model expressiveness. In this talk, we prove that \texttt{CFlows｝ can approximate any diffeomorphism in ξC^kξ-norm if its layers can approximate certain single-coordinate transforms. Specifically, we derive that a composition of affine coupling layers and invertible linear transforms achieves this universality. Furthermore, in parametric cases where the diffeomorphism depends on some extra parameters, we prove the corresponding approximation theorems for parametric coupling flows named \texttt{Para-CFlows｝. In practice, we apply \texttt{Para-CFlows｝ as a neural surrogate model in contextual Bayesian optimization tasks, to demonstrate its superiority over other neural surrogate models in terms of optimization performance and gradient approximations.

C1-3 AutoML 在推荐系统特征交互建模中的应用

唐睿明，华为技术有限公司

摘要：如何建模特征之间的交互关系，在推荐系统中至关重要。传统的 Logistic Regression 模型，简单有效，依赖专家经验进行特征交互设计。Factorization Machine 模型有效建模了二阶特征交互。而深度学习模型以其优秀的特征表达能力，在推荐系统占据一席之地。在深度学习模型中，特征交互的设计也是尤为重要。我们对现有的几种设计方式进行总结，并且提出了创新性的方案。AutoML，在计算视觉中的图像分类问题中，自动设计出各种新颖的神经网络，不断刷新各类竞赛的精度上限。我们也会介绍 AutoML 技术在特征交互的设计上如何发挥优势。

C1-4 Perception system in Autonomous Driving and its application in AutoML

徐航，华为诺亚实验室

Abstract: In recent years, autonomous driving has become a very popular topic in the car industry and Internet companies. Since 2003, the rapid development of deep learning technology has brought revolutionary changes to the autonomous driving system. This talk will introduce some components and specific tasks of the autonomous driving perception system, such as obstacle detection, driving road segmentation, lane line detection, point cloud detection, etc., to make the audience have an understanding of this direction. Furthermore, we will briefly introduce some of our recent research of AutoML for object detection in this field to enable an efficient and accurate perception system.

C1-5 Perception system in Autonomous Driving and its application in AutoML

张世枫，华为诺亚方舟实验室

Abstract: It was estimated that the world produced 5.9e13 GB of data in 2020, resulting in the enormous costs of both data storage and transmission. Fortunately, recent advances in deep generative models have spearheaded a new class of so-called "neural compression" algorithms, which significantly outperform traditional codecs in terms of compression ratio. Unfortunately, the application of neural compression garners little commercial interest due to its limited bandwidth; therefore, developing highly efficient frameworks is of critical practical importance. In this talk, we focus on efficient lossless compression algorithms with various types of generative models, in which the novel methods with state-of-the-art compression ratio and real-time compression bandwidth are achieved. We also discuss the generative ability of generative model for AI codec. The potential of generative ability of image compression is theoretically analysised, and dynamic generative model is proposed which greatly improves the AI codec algorithms.

十八、分会报告 C2 专题：机器学习与多尺度建模

C2-1 DeepMMnet for hypersonics: Predicting the coupled flow and finite-rate chemistry behind a normal shock using neural-network approximation of operators

毛志平，厦门大学

Abstract: In high-speed flow past a normal shock, the fluid temperature rises rapidly triggering downstream chemical dissociation reactions. The chemical changes lead to appreciable changes in fluid properties, and these coupled multiphysics and the resulting multiscale dynamics are challenging to resolve numerically. Using conventional computational fluid dynamics (CFD) requires excessive computing cost. Here, we propose a totally new efficient approach, assuming that some sparse measurements of the state variables are available that can be seamlessly integrated in the simulation algorithm. We employ a special neural network for approximating nonlinear operators, the DeepONet, which is used to predict separately each individual field, given inputs from the rest of the fields of the coupled multiphysics system. Specifically we predict five species in the non-equilibrium chemistry downstream of a normal shock at high Mach numbers as well as the velocity and temperature fields. We show that upon training, DeepONets can be over five orders of magnitude faster than the CFD solver employed to generate the training data and yield good accuracy for unseen Mach numbers within the range of training. Outside this range, DeepONet can still predict accurately and fast if a few sparse measurements are available. We then propose a composite supervised neural network, DeepMMnet, that uses multiple pre-trained DeepONets as building blocks and scattered measurements to infer the set of all seven fields in the entire domain of interest. Two DeepMMnet architectures are tested, and we demonstrate the accuracy and capacity for efficient data assimilation.

C2-2 Layer-Parallel Training of Residual Networks with Auxiliary-Variable Networks

孙琪，同济大学

Abstract: Gradient-based methods for the distributed training of residual networks typically require a forward pass of the input data, followed by back-propagating the error gradient to update model parameters, which becomes time-consuming as the network goes deeper. To break the algorithmic locking and exploit synchronous module parallelism in both the forward and backward modes, auxiliary-variable methods have attracted much interest lately but suffer from significant communication overhead and lack of data augmentation. In this work, a novel joint learning framework for training realistic ResNets across multiple compute devices is established by trading off the storage and recomputation of external auxiliary variables. More specifically, the input data of each independent processor is generated from its low-capacity auxiliary network, which permits the use of data augmentation and realizes forward unlocking. The backward passes are then executed in parallel, each with a local loss function that originates from the penalty or augmented Lagrangian methods. Finally, the proposed AuxNet is employed to reproduce the updated auxiliary variables through an end-to-end training process. We demonstrate the effectiveness of our methods on ResNets and WideResNets across CIFAR-10, CIFAR-100, and ImageNet datasets, achieving speedup over the traditional layer-serial training method while maintaining comparable testing accuracy.

C2-3 Asymptotic preserving scheme for anisotropic elliptic equations with deep neural network

杨畅，哈尔滨工业大学

Abstract: In this work, a new asymptotic preserving (AP) scheme is proposed for the anisotropic elliptic equations. Different from previous AP schemes, the actual one is based on first-order system least-squares for second-order partial differential equations, and it is uniformly well-posed with respect to anisotropic strength. The numerical computation is realized by a deep neural network (DNN), where least-squares functionals are employed as loss functions to determine parameters of DNN. Numerical results show that the current AP scheme is easy for implementation and is robust to approximate solutions or to identify the anisotropic parameter in various 2D and 3D tests.

C2-4 Learning Thermodynamically Stable and Galilean Invariant Partial Differential Equations for Non-Equilibrium Flows

周一舟，北京大学

Abstract: In this work, we develop a method for learning interpretable, thermodynamically stable and Galilean invariant partial differential equations (PDEs) based on the conservation-dissipation formalism of irreversible thermodynamics. As governing equations for non-equilibrium flows in one dimension, the learned PDEs are parameterized by fully connected neural networks and satisfy the conservation-dissipation principle automatically. In particular, they are hyperbolic balance laws and Galilean invariant. The training data are generated from a kinetic model with smooth initial data. Numerical results indicate that the learned PDEs can achieve good accuracy in a wide range of Knudsen numbers. Remarkably, the learned dynamics can give satisfactory results with randomly sampled discontinuous initial data and Sod's shock tube problem although it is trained only with smooth initial data.

十九、分会报告 C3 专题：机器学习中的优化问题（二）

C3-1 非线性不定正则临近点算法

韩德仁，北京航空航天大学

摘要：本报告回顾求解优化和算子方程问题的临近点算法，介绍其进展。

C3-2 THOR, Trace-based Hardware-driven layer-ORiented Natural Gradient Descent Computation

赵欣苑，北京工业大学

Abstract: It is well-known that second-order optimizer can accelerate the training of deep neural networks, however, the huge computation cost of second-order optimization makes it impractical to apply in real practice. In order to reduce the cost, many methods have been proposed to approximate a second-order matrix. Inspired by KFAC, we propose a novel Trace-based Hardware-driven layer-ORiented Natural Gradient Descent Computation method, called THOR, to make the second-order optimization applicable in the real application models. Specifically, we gradually increase the update interval and use the matrix trace to determine which blocks of Fisher Information Matrix (FIM) need to be updated. Moreover, by resorting the power of hardware, we have designed a Hardware-driven approximation method for computing FIM to achieve better performance. To demonstrate the effectiveness of THOR, we have conducted extensive experiments. The results show that training ResNet-50 on ImageNet with THOR only takes 66.7 minutes to achieve a top-1 accuracy of 75.9 % under an 8 Ascend 910 environment with MindSpore, a new deep learning computing framework. Moreover, with more computational resources, THOR can only takes 2.7 minutes to 75.9 % with 256 Ascend 910.

C3-3 Non-convex Factorization and Manifold Formulations in Low-rank Matrix Optimization

郦旭东，复旦大学

Abstract: In this talk, we consider the geometric landscape connection of the widely studied manifold and factorization formulations in low-rank positive semidefinite (PSD) and general matrix optimization. We establish an equivalence on the set of first-order stationary points (FOSPs) and second-order stationary points (SOSPs) between the manifold and the factorization formulations. We further give a sandwich inequality on the spectrum of Riemannian and Euclidean Hessians at FOSPs, which can be used to transfer more geometric properties from one formulation to another. We also discuss applications of our findings to some machine learning problems.

C3-4 On the convergence analysis of variational Monte Carlo methods

文再文，北京大学

Abstract: The variational Monte Carlo (VMC) method is very promising for solving many-body quantum problems at a manageable cost. In this talk, we provide a rigorous analysis of the VMC method, by showing the convergence of the stochastic gradient method with respect to the number of iterations and the cost of sampling.

二十、分会报告 C4 专题：机器学习与计算数学（三）

C4-1 DeePN2: A deep learning-based non-Newtonian hydrodynamic model

方礼冬，Michigan State University

Abstract: A long-standing problem in the modeling of non-Newtonian hydrodynamics of polymeric flows is the availability of reliable and interpretable hydrodynamic models that faithfully encode the underlying micro-scale polymer dynamics. The main complication arises from the long polymer relaxation time, the complex molecular structure, and heterogeneous interaction. DeePN2, a deep learning-based non-Newtonian hydrodynamic model, has been proposed and has shown some success in systematically passing the micro-scale structural mechanics information to the macro-scale hydrodynamics for suspensions with simple polymer conformation and bond potential. The model retains a multi-scaled nature by mapping the polymer configurations into a set of symmetry-preserving macro-scale features. The extended constitutive laws for these macro-scale features can be directly learned from the kinetics of their micro-scale counterparts. In this paper, we develop DeePN2 using more complex microstructural models. We show that DeePN2 can faithfully capture the broadly overlooked viscoelastic differences arising from the specific molecular structural mechanics without human intervention.

C4-2 A non-gradient method for solving elliptic partial differential equations with deep neural networks

胡丹，上海交通大学

Abstract: Deep learning has achieved wide success in solving Partial Differential Equations (PDEs), with particular strength in handling high dimensional problems and parametric problems. Nevertheless, there is still a lack of a clear picture on the designing of network architecture and the training of network parameters. In this work, we developed a non-gradient framework for solving elliptic PDEs based on Neural Tangent Kernel (NTK): 1. ReLU activation function is used to control the compactness of the NTK so that solutions with relatively high frequency components can be well expressed; 2. Numerical discretization is used for differential operators to reduce computational cost; 3. A dissipative evolution dynamics corresponding to the elliptic PDE is used for parameter training instead of the gradient-type descent of a loss function. The dissipative dynamics can guarantee the convergence of the training process while avoiding employment of loss functions with high order derivatives. It is also helpful for both controlling of kernel property and reduction of computational cost. Numerical tests have shown excellent performance of the non-gradient method.

C4-3 Deep learning-based method for solving incompressible Navier-Stokes equation and Chan-Hilliard equation

贺巧琳，四川大学

Abstract: We extend the algorithm presented by Han et al. to Navier-Stokes and Cahn-Hilliard equations in high dimension, which is an initial boundary value problem. The equation is reformulated using backward stochastic differential equations and the gradient of the unknown solution is approximated by neural networks. Numerical examples show the accuracy of the algorithm, which is quite effective in high dimension.

C4-4 DeLISA: Deep learning based iteration scheme approximation for solving PDEs

李颖，上海大学

Abstract: Solving the high dimensional partial differential equations (PDEs) with the classical numerical methods is a challenge task. As possessing the power of progressing high dimensional data, deep learning is naturally considered to solve PDEs. This paper proposes a deep learning framework based iteration scheme approximation, called DeLISA. First, we adopt the implicit multistep method and Runge-Kutta method for time iteration scheme. Then, such iteration scheme is approximated by a neural network. Due to

integrating the physical information of governing equation into time iteration schemes and introducing time-dependent input, our method achieves the continuous time prediction without a mass of interior points. Here, the activation function with adaptive variable adjusts itself during the iteration process. Finally, we present numerical experiments results for some benchmark PDEs, including Burgers, Allen-Cahn, Schrödinger, carburizing and Black-Scholes equations, and verify that the proposed approach is superior to the state-of-the-art techniques on accuracy and flexibility. Moreover, the Frequency Principle is also illustrated by the changes of prediction at different iterations in this paper.

二十一、分会报告 C5 专题：强化学习分会

C5-1 Fully Decentralized Multi-Agent RL

卢宗青，北京大学

摘要：独立学习/完全去中心化学习是合作场景多智能体强化学习中最简单的方法。但是由于没有收敛性保证，目前合作场景多智能体强化学习大多采用 CTDE 范式（集中式训练分布式执行）。本讲座将重新审视独立学习/完全去中心化学习，并介绍这方面最新的研究突破，包括有收敛保证的值函数、策略梯度以及基于模型的独立强化学习算法。

C5-2 多样性强化学习：不光要赢，还要赢得精彩

吴翼，清华大学

摘要：近年来，深度强化学习技术已经攻克了越来越多的人工智能挑战，也在许多任务中击败了无数顶尖人类选手。强化学习 AI 以得到最高的奖励分数为目标，并能以最简单高效的方式赢得比赛。虽然 AI 很强，但是 AI 的行为却往往千篇一律，与人相比缺少了些创造性。人类不光能够赢得比赛，常常还能发明各种各样有趣的创新策略并以此为乐。那为什么有些人类觉得好玩有意思的行为，强化学习 AI 却从来学不会呢？在这个报告中，我们会对这个问题进行分析，并提出新的强化学习范式，多样性强化学习——即，AI 不光要得高分，还要尽可能用不同的方式赢得高分。我们也会介绍两个多样性强化学习算法，并展示在不同场景中，这两个算法所发现的有趣的行为。

C5-3 合作博弈的通用求解框架

杨耀东，北京大学

摘要：基于 IGM 假设的值函数分开方法存在诸多弊端，本讲中，我会介绍基于优势函数分解的合作型多智能体强化学习框架以及镜像学习方法。该框架作为合作博弈的通用求解方案可以衍生出诸多算法，例如 HATRPO, HAPPO, Multi-Agent Transformer, Multi-agent Mirror Learning 等。

C5-4 探索强化学习大模型

张伟楠，上海交通大学

摘要：在本次报告中，我将首先介绍离线强化学习中的 Transformer 模型的使用。然后我将从自模仿学习这一类离线强化学习方法开始，讨论几个近期刚发布的面向决策智能的大模型方法，包括 decision transformer、trajectory transformer 和 Gato 等。进一步，我将讨论对比有监督学习的序列建模方法和强化学习方法之间的联系。最后我将讨论决策智能大模型的可行性和未来发展的趋势。

二十二、分会报告 C6 专题：机器学习与燃烧数值模拟（一）

C6-1 冲压发动机内部流场场信息智能重构方法初步探讨

常军涛，哈尔滨工业大学

摘要：针对冲压发动机内部流场，采用局部压力温度测量很难获得全局信息，如何获得全局信息为发动机控制系统决策服务是关键。本报告尝试利用深度学习感知及预测发动机内部流场信息，分别针对隔离段流场、燃烧室流场开展了超分辨重构，流场预测等方面的研究工作，可为控制系统提供高密度信息源，提高状态监测和故障诊断精度。

C6-2 机器学习在发动机喷雾燃烧模拟和优化中的应用

贾明，大连理工大学

摘要：发动机的工作过程复杂，影响因素众多，且相互限制和作用，传统研究局限于特定工况和指定参数范围，难以实现宽工况下燃烧的有效控制，亟需开发高效可靠的喷雾和燃烧模拟系统，为高效清洁发动机研发提供技术手段。在本报告中，利用高效机器学习工具，分别开展了实际燃油物性预测、燃料反应模型求解和发动机全参数优化三方面工作。首先，通过于柴油和生物柴油组分数据分析，分别选取馏程和烃类组成为输入特征，使用工人神经网络（ANN）构建了燃油物性预测模型，进一步开展 ANN 多因素敏感性分析，识别影响燃油特性的重要因素。其次，通过提出了改进的 ANN 算法，实现了燃料反应模型的准确求解，改进包括：基于监督学习的反应速率分区、基于 SiLU 梯度连续的激活函数、纳入条件扰动的训练数据，与直接数值积分法相比，计算速度提升显著。最后，将计算流体力学模拟与遗传算法和 ANN 相耦合，开展了双直喷发动机的优化设计，向 ANN 引入多模型权重预测和动态数据更新，实现了可变输入变量和优化目标的发动机优化设计，在精度、效率、扩展性和灵活性方面优势显著。

C6-3 基础燃烧数据与燃烧反应动力学模型优化

杨斌，清华大学

摘要：报告简要介绍下利用基础燃烧实验数据(着火、火焰传播、组分分布)进行详细化学反应动力学模型优化的面临的挑战和我们的发展的一些策略(实验归类、特征提取等)，展望机器学习对动力学模型优化的潜在推动。

C6-4 航空发动机燃烧振荡预测——机器学习图像分析方法

张弛，北京航空航天大学

摘要：由于航空发动机燃烧室的火焰图像与燃烧振荡之间有密切联系，可以采用火焰图像分析的方法预测燃烧振荡的发生，并提前采取相应措施主动控制燃烧振荡。介绍多种燃烧多物理场机器学习分析方法，据此发展了手动、无监督和有监督三种火焰低维特征提取模型架构，对 BASIS 燃烧器瞬态和时均图像数据集开展燃烧振荡预测，采用主成分分析法(PCA)将提取到的多个低维特征解耦，观察到不同特征对于燃烧振荡的表征能力。对于燃烧振荡预测，声压级监督神经网络体现出更高的效率和更好的鲁棒性及迁移性。

C6-5 深度学习赋能的多维燃烧诊断技术

蔡伟伟，上海交通大学

摘要：应用。最后报告将介绍光谱仪/光谱相机微型化、高速高光谱相机及火焰五维成像技术的发展思路。

二十三、分会报告 D1 专题：因果推断与统计学习中人工智能应用的新进展

D1-1 Random Matrix Methods for Machine Learning: "Lossless"Compression of Large Neural Networks

廖振宇，华中科技大学

Abstract: The advent of the Big Data era has triggered a renewed interest in large-dimensional machine learning (ML) and (deep) neural networks. These methods, being developed from small-dimensional intuitions, often behave dramatically different from their original designs and tend to be inefficient on large-dimensional datasets. By assuming both dimension and size of the datasets to be large, recent advances in random matrix theory (RMT) provide novel insights, allowing for a renewed understanding and the possibility to design more efficient machine learning approaches, thereby opening the door to completely new paradigms. In this talk, we will start with the "curse of dimensionality" phenomenon in high dimensions, and highlight many counterintuitive phenomena in ML that arise when large-dimensional data are considered. By focusing on the use case of neural network compression, and by considering the data dimension and/or the ML systems to be large, we discuss how RMT is able to provide a renewed understanding of modern ML.

D1-2 Paradoxes and solutions for semiparametric fusion learning with external summary statistics

苗旺，北京大学

Abstract: Suppose we have available individual-level data from an internal study and various types of summary statistics from relevant external studies. External summary statistics have been used as constraints on the internal data distribution, which promised to improve the statistical inference; however, paradoxical results arise in such data integration: efficiency loss may occur if the uncertainty of the summary statistics is not negligible and estimation bias can emerge if they are obtained from a different population from the internal study. We investigate these paradoxical results in a semiparametric framework. We establish the semiparametric efficiency bound for estimating a general functional of the internal data distribution, which is shown to be no larger than that using only internal data. We propose an efficient estimator that achieves this bound so that the efficiency paradox is resolved. This initial efficient estimator is further regularized with adaptive lasso penalty so that the resultant estimator can achieve the same asymptotic distribution as the oracle one that uses only unbiased summary statistics, which resolves the bias paradox. Simulations and an application to a Helicobacter pylori infection data are used to illustrate the proposed methods.

D1-3 Functional approximation perspective on neural networks and statistical models

杨朋昆，清华大学

Abstract: Modern machine learning has constantly presented puzzling empirical properties and surprised the classical statistical theory. Learning with overparametrized models is becoming a norm in data-analytic applications, and the tension of memorization rarely bothers practitioners. In this talk, I will discuss the training of overparametrized neural networks from both the neural tangent kernel and the mean-field perspectives, which guarantee the global convergence property despite the non-convexity of the optimization landscape. Time permitted, extensions to other overparametrized statistical questions will also be discussed.

D1-4 Nonparametric Estimation of the Continuous Treatment Effect with Measurement Error

张政，中国人民大学

Abstract: We identify the average dose-response function (ADRF) for a continuously valued error contaminated treatment by a weighted conditional expectation. We then estimate the weights nonparametrically by maximising a local generalised empirical likelihood subject to an expanding set of conditional moment equations incorporated into the deconvolution kernels. Thereafter, we construct a deconvolution kernel estimator of ADRF. We derive the asymptotic bias and variance of our ADRF estimator and provide its asymptotic linear expansion, which can help conduct statistical inference. To select our smoothing parameters, we adopt the simulation-extrapolation method and propose a new extrapolation procedure to stabilise the computation. Monte Carlo simulations and a real data study illustrate our method's practical performance.

二十四、分会报告 D2 专题：AI+复杂信息系统：机器学习在 ICT 领域的机遇与挑战

D2-1 Flow Neural Network and Beyond for AI-Native Network

程祥乐，华为

摘要：机器学习在数通网络领域的应用还处在学术探索及初步的工业应用尝试阶段。一方面，随着网络规模及流量的日益增长，网络的规划、运维、管控对智能化的诉求愈加强烈。而另一方面，当前 AI 技术/模型的成熟度及对网络场景的适配度与实际工业部署的要求仍存在一定差距。本次报告将介绍网络领域对设计特定机器学习模型的尝试，及未来实现 AI 原生的网络演进面临的机遇和挑战。

D2-2 机器学习方法和技术在光纤通信中的应用

范起瑞，华为香港综合技术创新中心

摘要：自 2011 年有研究提出用 Extreme Learning Machine 来做光纤非线性均衡以来，机器学习和光通信的"磨合"从未停止脚步。随着实际产品性能逐渐接近实验室离线水平，相对复杂的光纤非线性成为了阻碍了频谱效率的进一步提升的主要因素，而传统的理论办法收效甚微。与此同时，增加更多的带宽带来了额外的系统复杂性，具体表现在更加复杂的光纤信道和器件行为，传统的数值模型遇到了效率和准确性的挑战。这些未知数也为光网络层面的性能建模带来不确定性。除此之外，值得期待的是，这些年因深度学习的繁荣发展而来的硬件、软件基础有望推动光通信的领域知识和方法进一步前进。

D2-3 机器学习在光通信中的应用

罗龙，华为技术有限公司

摘要：在光纤信道中利用机器学习获得对多种光纤应力的检测、补偿以及器件的非理想因素进行补偿从而获得不同于传统信号处理的方法的探索及挑战

D2-4 AI+光感知

缪赟，华为理论研究部

摘要：分布式光纤传感器本身不带电，具有抗电磁干扰、电绝缘、耐腐蚀、灵敏度高、质量轻、体积小、可嵌入（物体）等优良特点，已在多领域广泛应用，市场总额接近百亿。目前分布式光纤传感器

技术已较为成熟，然而分布式光纤传感采集到的数据却对于环境噪声敏感，精度低，泛化能力差，在煤矿等场景中的应用受到极大的限制，如何利用传感器采集到的数据进行事件检测和声源定位成为了业界的一个重要课题。本报告介绍了光纤传感的物理原理并总结了理论研究部目前利用 AI+光感知技术取得的部分阶段性成果。

D2-5 AI+数据无损压缩

孙赵亿，华为香港理论研究部

摘要：信息时代下，海量数据需要被高效地传输和存储， 数据压缩（把大数据“变小”）是通信和存储领域的核心技术。而统计学习、建模优化以及人工智能的快速发展为数据缩减提供了新的前沿思路和方向。本次报告主要分享对当前业界 AI 无损压缩技术路线的分析思考，展望潜在的机遇和挑战。

D2-6 漫谈神经网络的统计物理：从 NTK、最优输运到非平衡态热力学

田洋，清华大学

摘要：神经网络并不是黑盒，这样一个共识的形成得益于近年来机器学习理论的发展。在本次报告中，我将从神经正切核（NTK）开始，分析神经网络初始化和在优化中偏移初始化状态的过程。通过最优输运、信息几何与非平衡态热力学间密切的联系，我们将能对神经网络初始化和优化过程中的非平衡态热力学过程进行一次漫谈，并指出其对于神经网络优化的作用。

D2-7 人工智能在算力优化领域的机遇和挑战

周李，上海华为技术有限公司

摘要：算力优化具有重要的现实意义，在东数西算工程的时代背景下，如何更好地分配和调度算力资源值得深入研究，从而有利于挖掘现有计算资源潜能。对于华为公司，能够助力提升 HPC 全栈解决方案竞争力。本报告将从 CESM 的算力资源分配出发， 分享相关优化问题的思考，展示部分结果，展望其中的机遇与挑战。

二十五、分会报告 D3 专题：材料科学中的机器学习方法

D3-1 Improving Symbolic Regression for Predicting Materials Properties with Iterative Variable Selection

欧阳润海，上海大学

Abstract: Symbolic regression offers a promising avenue for describing the structure−property relationships of materials with explicit mathematical expressions, yet it meets challenges when the key variables are unclear because of the high complexity of the problems. Here, we propose to solve the difficulty by automatically searching for important variables from a large pool of input features. A new algorithm that integrates symbolic regression with iterative variable selection (VS) was designed for optimization of the model with a large amount of input features. Using the recent method SISSO for symbolic regression and random search for variable selection, the VS-assisted SISSO (VS−SISSO) can effectively manage tens or hundreds of input features that the SISSO alone would be computationally hindered, and it fastly converges to (near) optimal solutions when the model complexity is not high. The efficiency of this approach for improving the accuracy of symbolic regression in materials science was demonstrated in the two showcase applications of learning approximate equations for the band gap of inorganic halide perovskites and the stability of single-atom alloy catalysts.

D3-2 DP 方法在材料研究中的应用

戴付志，北京科学智能研究院

摘要：原子模拟是揭示材料微观机制的重要手段。然而，原子模拟长期受限于效率和精度难以兼顾的矛盾，限制了其在材料研究中的应用。近些年，机器学习势函数（尤其是 Deep Potential, DP）方法克服了原子模拟效率和精度的矛盾，显著拓展了原子模拟的能力边界。此次报告将以有机无机杂化钙钛矿和超高温陶瓷为例，阐述 DP 方法是如何助力材料研究，提升对材料的认知，并帮助解释实验研究中的误解。

D3-3 极端条件材料的结构演化

戴佳钰，国防科技大学

摘要：材料在动态过程如高速碰撞、冲击、强激光作用使得其组成组分经过复杂加载路径进入极端高压和高温，且物质的热力学状态高度依赖路径和瞬态结构，非常依赖于高精度的大尺度动力学模拟。我们发展时间分辨、大尺度、非绝热等模型同时耦合的计算方法，利用和发展分子动力学并结合机器学习的多尺度研究手段，获得了强激光作用下电子-离子的非平衡动力学和熔化过程。结合机器学习发展了第一原理精度的大尺度模拟方法，并研究了冲击压缩下金属/CH 的相变动力学，为动态载荷下结构研究提供了手段和思路。

D3-4 主动学习在材料开发中的应用

薛德祯，西安交通大学

摘要：通过尽可能少的实验来寻找具有目标特性的新材料是材料研发的重要目标。材料研发面临着巨大复杂的高维搜索空间，如何从巨大的搜索空间中快速有效选择具有目标性能的新材料是主要挑战之一。机器学习可以基于已有的数据，通过算法建立变量和目标性能之间的映射关系，对材料的未知性能进行预测。但现实情况下，材料的数据比较少，模型的预测能力比较差，对实验或者计算的指导非常有限。针对这个问题，利用主动学习的方法，在机器学习预测的基础上，进行实验设计，选择对目标提升最有帮助、带有更大信息量的实验，获得数据反馈回数据库，反复迭代后达到更快的优化材料性能的目的。本文主要从材料性能优化、曲线优化、材料分类问题三个方面，介绍主动学习在材料开发中的一些应用。

二十六、分会报告 D4 专题：机器学习与燃烧数值模拟（二）

D4-1 基于数据驱动的可压缩湍流的反卷积模型

王建春，南方科技大学

摘要：大涡模拟方法通过对湍流做滤波，大幅度降低了流场的自由度，在各类湍流问题中得到了越来越广泛的应用。常用的大涡模拟模型存在着精度不高、耗散过强等问题。近几年来，机器学习方法为湍流模型提供了新的发展方向。在可压缩湍流的大涡模拟中，我们使用人工神经网络方法，构造近似反卷积算子，用滤波后的大尺度流场还原滤波前的密度、动量和压力。在此基础上，对各个不封闭的亚格子项构造了人工神经网络反卷积模型：DANN 模型。我们还对简单化学反应情况下的可压缩湍流发展了相应的人工神经网络反卷积模型。进一步，我们发展了动态迭代反卷积模型：DIAD 模型。该模型使用滤波后的湍流场数据，多次迭代逼近最优的反卷积权重系数，进而得到滤波前的物理场，最后采用高精度的相似模型对亚格子未封闭项进行建模。动态更新模型系数的过程只需要大涡模拟的流场数据，不依赖于直接数值模拟的数据，因此，该模型具有很好的泛化能力。数值模拟的结果表明，DANN 模型和 DIAD 模型比传统的大涡模拟模型能更精确地预测湍流的各类统计量和瞬态流动结构。

D4-2 机器学习在湍流燃烧中的应用

王海鸥，浙江大学

摘要：在湍流燃烧的大涡模拟（LES）中，亚网格标量通量和过滤反应源项的精确建模是一个挑战。我们提出了一种卷积神经网络（CNN），对过滤物理量来重建高精度的过程变量、速度和反应速率，并用该 CNN 模型和近似反褶积方法（ADM）对滤波过程变量传输方程中的未闭项进行建模。通过比较 ADM 模型、CNN 模型与 DNS 的结果，发现 CNN 模型能够准确预测 DNS 各种量的分布。建立了亚网格标量通量项和过滤反应源项的预测模型，发现 CNN 模型的性能优于 ADM。总体而言，CNN 模型在湍流燃烧的数据重建和模型开发方面具有广阔的应用前景。

D4-3 基于卷积神经网络反应模型的湍流燃烧数值模拟

万凯迪，北京航空航天大学宁波创新研究院

摘要：提出了一个新的基于卷积神经网络（CNN）的化学反应简化策略，并应用于研究与水冷壁面相互作用的合成气/氧气湍流非预混火焰的直接数值模拟。CNN 的训练和应用依赖于对组分质量分数和温度的二维图像化处理（CNN 输入量），用于预测图像中心位置的化学反应源项值（CNN 输出量）。该图像化处理方法能够准确预测对当地火焰结构高度敏感的燃烧中间产物自由基组分。同时，CNN 反应简化模型比传统阿伦尼乌斯形式的简化机理计算速度更快。

D4-4 基于深度学习实现碳氢燃料燃烧的反应分子动力学模拟

朱通，华东师范大学

摘要：反应分子动力学模拟(Reactive MD simulation)是探究燃料氧化反应机理的主要工具。然而，受限于计算能力和采样效率，传统的反应动力学模拟方法通常在较高的温度和密度下进行，无法用其探究燃料的低温氧化机理。最近提出的 Collective Variable-driven HyperDynamics (CVHD) 方法通过在反应势能面中添加偏置势(bias potential)的方式，有效地加速了 MD 模拟的采样效率，实现了更低温度下的燃烧动力学模拟。与此同时，越来越多的工作发现在燃烧动力学模拟中广泛使用的 ReaxFF 力场的精度仍有较大的提升空间。最近，我们结合深度学习方法，成功在 DFT 精度上实现了燃料燃烧的高精度模拟。在本工作中，我们结合 CVHD 方法和基于神经网络的精确势能面模型，探索了 600K-2400K 宽温度工况下正十二烷及 DME 的氧化机理，构建了完整的反应网络并详细阐明了温度对于氧化机理的影响。精确的势能面模型和增强采样算法的结合有望为燃料低温氧化机理的研究提供更加精确有效的工具。

D4-5 机器学习在含能材料反应动力学中的应用

陈东平，北京理工大学

摘要：含能材料反应动力学是决定含能材料能量释放性能的化学基础，在辅助设计人员优化含能材料配方中起关键作用。本次报告将围绕报告人过去 4 年在含能材料反应动力学方面的工作。首先介绍新一代分子动力学力场（DeepMD 力场）在含能材料体系中的应用。报告人将以含能材料（RDX、ICM-102 等）的 DeepMD 力场开发为例，展开讨论 DeepMD 力场未来在含能材料领域的优势。其次，报告人将介绍一种全新的机器学习建模技术在含能材料反应动力学模型开发中的应用，着重讨论如何合理的利用热重实验数据，建立 RDX、含铝复合体系（PTFE/Al 和 FPU/Al）等动力学模型以及相应的反应机理。最后，报告人将对于未来含能材料反应动力学研究的发展展开简要探讨。

二十七、分会报告 D5 专题：机器学习辅助分子模拟的新进展

D5-1 The Electrolyte Project

陈翔，清华大学

Abstract: Lithium metal batteries are considered as promising next-generation energy storage devices due to their ultrahigh energy densities. However, the high reactivity of lithium metal toward organic solvents and salts renders inevitable side reactions, which further leads to undesirable electrolyte depletion, cell failure, and the evolution of flammable gas. Tremendous efforts from experiments have been devoted to the exploration of advanced electrolytes mainly through trial-and-error approaches, which are both time-consuming and cost-expensive. Herein, we prosed the Electrolyte Project to accelerate the rational design of advanced electrolytes through data-driven and machine-learning approaches. Specifically, a big dataset of electrolyte molecules and corresponding physicochemical properties is constructed through graph theory and high-throughput calculations. Machine-learning models are further training to predict electrolyte properties. At last, a demo of the high-throughput design of advanced electrolytes is provided.

D5-2 DeepH: 深度学习 DFT 哈密顿量

李贺，清华大学

摘要：第一性原理计算已被广泛应用于物理、材料、化学、生物相关的科学研究。然而，受限于计算效率和精度，如何实现大尺度材料体系的第一性原理研究是该领域的一个重大挑战。近期，深度学习已经成功应用于精确预测原子间相互作用，并加速分子动力学模拟。相比之下，理解、预测材料物性离不开电子结构计算，其深度学习的方法实现更具挑战性，研究进展有限。因此，发展深度学习方法、解决第一性原理电子结构计算的效率-精度两难困境是一个关键的科学问题。我们提出了一种深度学习方法 DeepH，能实现由材料构型快速、高精度预测任意原子构型的密度泛函理论哈密顿量，可极大加速电子结构计算。通过对多种代表性材料的态密度、能带结构、非线性光学响应多种物理性质的精确预测，我们展示了 DeepH 方法的高精度（预测误差为毫电子伏特级别）和良好的可迁移性。

D5-3 DeePKS+ABACUS：AI 辅助的电子结构方法

李文菲，北京人工智能研究院（AISI）

摘要：近些年来，机器学习势函数的发展极大推动了分子模拟领域的研究，使得针对大体系的高精度性质预测成为可能。然而，机器学习势函数的训练首先需要生成大量的第一性原理计算数据，对于精度要求很高的场景（如 Quantum Monte-Carlo、杂化密度泛函等），生成这些数据会耗费极大甚至难以负担的计算资源。DeePKS 泛函模型的研发能够有效解决这一计算瓶颈问题，构建第一性原理计算与机器学习势函数之间的桥梁，显著提升动力学模拟在规模、精度等方面的天花板。

DeePKS+ABACUS 的工作依托周期性边界条件的 DFT 软件 ABACUS，实现了在周期性体系下的 DeePKS 训练，并以纯水为例测试了 DeePKS 模型应用于 DeePMD 势函数训练及动力学模拟的效果，在相同的模拟条件下，DeePKS-DeePMD 模拟完美复现了此前 SCAN0-DeePMD 模拟给出的结构性质，且与 SCAN0-AIMD 的结果高度吻合。

D5-4 Mg-Y 合金深度学习势构建及锥面刃位错分解滑移反应研究

王亦楠，北京大学

摘要：镁作为密度最低的结构材料，具有比强度、比刚度高、质量轻、储量丰富、生物相容性好等多种优点，未来将广泛应用于航天、汽车、太空领域，实现节能减排、节约能源的目的。然而镁的室温塑性差，加工性能有限，对镁基材料的研发和大规模应用造成了较大的阻碍。镁室温塑性差的原因是室温可开动的滑移系有限，而稀土材料如 Y 的加入对镁的室温塑性变形能力有很大的提升。这种提升

是否源于 Y 元素稳固了锥面位错的存在，并促进了锥面位错滑移，需要分子动力学方法辅助研究。本工作构建了 Mg-Y 合金深度学习势函数（DP），对 DP 势进行了系统的测试工作，包括点阵常数、弹性模量、空位形成能、表面能等基本性质，也包括层错、位错、孪晶、晶界等缺陷性质。针对镁的锥面刃位错，模拟了锥面-基面转化过程和锥面位错滑移行为，并计算了不同浓度 Y 富集于位错核心后转化和滑移两个过程的势垒变化，讨论了应力和温度影响下两种机制的竞争关系。结果表明，尽管 Y 富集使得锥面位错存在更稳定，但是位错滑移受到的阻力更强，在温度和应力作用下转化事件优先出现。这说明合金富集方式稳定锥面位错、对塑性变形产生贡献并不是一个有效的方式，之后的研究将更多针对于合金元素对其他变形机制的影响，如锥面螺位错的交滑移行为，及其它缺陷如孪晶、晶界参与的塑性变形过程。