目录

一、会务指南

2025 年中国机器学习与科学应用大会（CSML2025）由北京大学国际机器学习研究中心、数学科学学院和北京国际数学研究中心举办，上海交通大学自然科学研究院、数学科学学院和人工智能学院协办。

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：2025 年 8 月 9 日（周六）-8 月 10 日（周日）

签到时间：8 月 8 日 15:00-18:00；8 月 9 日-8 月 10 日：7:30 - 17:00

会议地点：北京友谊宾馆 • 友谊宫（北京市海淀区中关村南大街 1 号）

会议网站：https://c2sml.cn/conference.html

会务组联系方式：吴磊：leiwu@math.pku.edu.cn 袁坤：kunyuan@pku.edu.cn

王铭泽：mingzewang@stu.pku.edu.cn （学生分会联系人）

二、会议组织

大会主席：

鄂维南 （中国科学院院士，北京大学）

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，武汉大学）

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）：

董彬 (北京大学)

文再文 (北京大学)

谢伟迪 (上海交通大学)

张林峰 (上海交通大学)

张文涛 (北京大学)

张午阳 (中国科学技术大学)

邀请大会报告人（按姓氏拼音排序）：

董彬　(北京大学)

林洲汉 (上海交通大学)

孙若愚 (香港中文大学(深圳))

吴磊　(北京大学)

会议组织委员会 (按姓氏拼音排序)

组织委员会主席:

吴磊　(北京大学)

袁坤　(北京大学)

组织委员会成员:

董彬　(北京大学)

黄政宇 (北京大学)

许志钦 (上海交通大学)

张文涛 (北京大学)

张耀宇 (上海交通大学)

周沛勘 (北京大学)

主办单位

北京大学国际机器学习研究中心

北京大学数学科学学院

北京国际数学研究中心

协办单位

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

上海交通大学人工智能学院

三、会议日程

2025 年 8 月 9 日 - 10 日, 北京友谊宾馆 • 友谊宫

【签到时间：8 月 8 日 15:00-18:00；8 月 9 日-8 月 10 日：7:30 - 17:00】

时间	活动内容	主持人	地点
8 月 9 日			
8:20-8:40	**致辞**	吴磊	友谊宫 聚英厅
8:40-9:25	**邀请报告一** AI for Mathematics：数学的数字化与智能化 董彬 (北京大学)	许志钦	
9:25-10:10	**邀请报告二** Modeling LLM Pre-Training Dynamics with Functional Scaling Laws 吴磊 (北京大学)		
10:10-10:30	**茶歇**		
10:30-12:00	**圆桌论坛** 董彬（北京大学）、文再文（北京大学）、谢伟迪（上海交通大学）、张午阳（中国科学技术大学）、张文涛（北京大学）	张林峰	
12:00-13:30	**午餐**		见餐券
13:30-15:30	**A1 专题：深度学习理论** **邹荻凡：** Towards understanding the representation learning of diffusion models **曹原：**自注意力机制中的变量选择：案例研究与理论理解 **刘方辉：**Bridge theory to practice at scale: One-step gradient suffices for fine-tuning LLMs, provably and efficiently **凌舒扬：**Beyond Unconstrained Features: Neural Collapse for Shallow Neural Networks with General Data	罗涛	友谊宫 5 号 会议室
	A2 专题：神经科学和人工智能（I） **唐华锦：**高效脉冲神经网络模型训练算法研究	周栋焯	友谊宫

	徐齐：脉冲神经网络结构学习方法研究 **杨冬平：**从感知到认知的神经动力学机制 **田永鸿：**大规模脉冲神经网络理论与方法	陈国璋	11 号 会议室
	A3 专题：Optimization for AI **林涛：**面向生成模型的高效训练与推理算法 **凌青：**Physics-Assisted and Topology-Informed Deep Learning for Weather Prediction **罗珞：**On the Complexity of Distributed Nonconvex Optimization **濮实：**Distributed Learning over Arbitrary Topology: Linear Speed-Up with Polynomial Transient Time	袁坤	友谊宫 1 号 会议室
	A4 专题：大模型数据智能 **周号益：**从异构科学数据到统一模型输入：面向科学智算大模型的数据基础设施构建 **张林峰：**数据视角下的模型压缩加速 **周煊赫：**大模型数据准备的“IaaS”原理 **王斌：**MinerU: 精准解析文档，驱动 AI 应用	何聪辉	友谊宫 4 号 会议室
	A5 专题：科学机器学习 **郭孟武：**Bayesian Learning for Compact Dynamical Representations of Nonlinear Systems **龚禾林：**基于物理信息机器学习的反应堆多物理场耦合建模与数字孪生构建 **周元诚：**DeepSPoC: a deep learning based sequential propagation of chaos **于腾超：**基于混合神经网络的多保真度不确定度量化方法	郭玲 周元诚 冯晓东	友谊宫 2 号 会议室
	A6 专题：机器学习与数值方法 **宋学官：**算测融合的装备形性一体化数字孪生技术 **张伟伟：**数据驱动的流体力学知识发现与 AI4E 的应用 **蔡伟伟：**基于张量分解的高速背景纹影层析用于 4D 流场密度重建 **巴顿：**数智赋能航空发动机关键技术思考与研究进展 **邬晓敬：**AI 赋能工业 3-D 空气动力学设计与智能优化	陈景润	友谊宫 8 号 会议室

	A7 专题：机器学习与科学计算（I） **杨云斐：**Rates for least squares using over-parameterized neural networks **张仕俊：**Fourier Multi-Component and Multi-Layer Neural Networks: Unlocking High-Frequency Potential **马文森：**Distribution Matching for Self-Supervised Transfer Learning **袁成：**Score-Based Sequential Langevin Sampling for Data Assimilation **吴佩颖：**DRM Revisited: A Complete Error Analysis	焦雨领	友谊宫 10 号会议室
	A8 专题：大模型训练 **赵鑫：**大模型复杂推理技术 **刘知远：**大模型密度法则与高密度大模型关键技术 **束俊：**初探大模型“智能涌现”现象：从线性到非线性 **杭良慨：**Scalable Complexity Control Facilitates Reasoning Ability of LLMs	许志钦	友谊宫 7 号会议室
15:30-16:00	**茶歇**		
16:00-18:00	**B1 专题：大模型相关理论** **贺笛：**大模型表达能力理论 **吕凯风：**大模型训练的最优学习率衰减与扩展定律 **邱凯：**通过强化学习提升大语言模型的逻辑推理能力 **刘勇：**大模型推理机制分析	李建	友谊宫 5 号会议室
	B2 专题：机器学习与科学计算（II） **段晨光：**Solving Bayesian Inverse Problems via Diffusion-based Sampling **康利灿：**Schrodinger-Follmer Diffusion: Sampling, Optimization, Generative Learning **谢琦：**几何等变先验嵌入的深度网络模块设计 **丁钊：**Flow-based Sampling Method	焦雨领	友谊宫 10 号会议室
	B3 专题：神经科学和人工智能（II） **钟毅：**Identification of an engram ensemble encoding memory flexibility in the dentate gyrus	周栋焯	友谊宫 11 号

	王立元：脑启发的持续学习方法与生理健康应用 **郭尚岐：**基于脑启发的类脑决策模型 **黄子罡：**神经递质调控效应启发的类脑算法研究	陈国璋	会议室
	B4 专题：AI for Math **郑楚杰：**Beyond the 80/20 Rule: High-Entropy Minority Tokens Drive Effective Reinforcement Learning for LLM Reasoning **支丽红：**面向组合数学的定理自动生成和证明 **王海明：**Kimina-Prover: 一种推理驱动的形式化定理证明探索范式 **李嘉：**数据集的规模形式化	董彬 张文涛 严骏驰	友谊宫 7 号 会议室
	B5 专题：AI for Optimization **高斌：**A space-decoupling framework for optimization on bounded-rank matrices with orthogonally invariant constraints **丁添：**基于图同构判定的大模型优化建模评测体系 **孙建永：**人工智能驱动的大规模组合优化算法、平台与应用 **李天佑：**LMask: Learn to Solve Constrained Routing Problems with Lazy Masking	文再文	友谊宫 8 号 会议室
	B6 专题：量子计算理论与方法 **刘锦鹏：**Quantum for Science: Efficient Quantum Algorithms for Nonlinear Dynamics and Artificial Intelligence Models **李震宇：**Achieving Chemical Accuracy with Quantum Computing Enforced Language Model **史良良：**面向隐优化视角的可约束神经网络 **王鑫：**From Parameterized Quantum Comb to Quantum Unitary Time-Reversal	安冬	友谊宫 1 号 会议室
	B7 专题：AI for Physics and Chemistry（I） **朱通：**AI 物理双驱动的化学反应路径搜索 **朱戎：**AI 和自动化加速功能分子和反应发现 **张颖：**通用深度学习密度泛函框架：DL-xDH **任维络：**神经网络赋能量子蒙特卡洛	王涵	友谊宫 4 号 会议室
	B8 专题：Optimization for LLM		友谊宫

	李肖：Memory-Efficient Block Coordinate Descent and Backpropagation for LLM Training **沈力**：Fine-Tuning Large Language Models with Forward-only Optimizers **袁雁城**：Accelerating RLHF Training with Reward Variance Increase **袁坤**：Subspace Optimization for Large Language Models with Convergence Guarantees	袁坤	2 号 会议室
8 月 10 日			
8:20-9:05	**邀请报告三** 大模型新架构的初步探索与思考 林洲汉(上海交通大学)	袁坤	友谊宫聚英厅
9:05-9:50	**邀请报告四** 理解和改进大模型的训练: 一些新进展 孙若愚(香港中文大学(深圳))		
9:50-10:10	**茶歇**		
10:10-12:10	**C1 专题：强化学习理论与算法（I）** **陈卫**：Offline learning for combinatorial optimization **黄隆波**：uniINF: Best-of-Both-Worlds Algorithm for Parameter-Free Heavy-Tailed MABs **俞扬**：大模型背景下的强化学习 **陈昱鑫**：Settling the Sample Complexity of Online Reinforcement Learning	魏轲 李帅	友谊宫 5 号 会议室
	C2 专题：机器学习和逼近理论（I） **林绍波**：Learning performance of Off-line Q-learning algorithms **石磊**：Learning Theory of Classification with Deep Neural Networks **龙吉昊**：随机特征模型与两层神经网络分析的对偶框架 **刘皓**：Operator Learning and Neural Scaling Laws	吴磊 何俊材	友谊宫 1 号 会议室
	C3 专题：生成模型算法 **邓志杰**：高效多模态生成：方法与应用 **刘勇**：基于 LLM 的合成数据有效吗？	李崇轩	友谊宫 7 号

	贺笛：Diffusion vs. Autoregression: Which is the Key to Next-Generation LLMs **李崇轩**：LLaDA：大语言模型新范式		会议室
	C4 专题：大模型系统 **袁彬航**：大语言模型在异构算力环境中的部署 **章明星**：从同构走向分离的大模型推理系统 **赵汉宇**：PAI-Llumnix：动态、弹性、可扩展的分布式推理 **符芳诚**：复杂、动态负载下的分布式大模型训练	符芳诚	友谊宫 2 号 会议室
	C5 专题：机器学习与材料 **张露婵**：Modeling Randomness Effects in High-Entropy Alloys **袁成**：A Stabilized Physics Informed Neural Networks Method for Wave Equations **干则成**：Data-driven approaches for numerical PDEs: reduced order modeling & operator learning **黄记祖**：Frequency-adaptive Multi-scale Deep Neural Networks **项阳**：A Generative Model for Composition Engineering in Multi-Principal Element Alloys	戴书洋	友谊宫 4 号 会议室
	C6 专题：图像处理与人工智能 **张立**：面向肿瘤疗效预测的多模态分析方法 **邱凌云**：Enhancing Full Waveform Inversion via Learned and Regularized Source Wavelet Manipulation **段玉萍**：Parametric Neural Operator for Non-Line-of-Sight Imaging	包承龙	友谊宫 11 号 会议室
	C7 专题：算子学习 **魏华祎**：FEALPy: A Cross-Platform Intelligent CAX Engine with Scalable Tensor Computation for Multi-Method Simulations **毛志平**：Solving PDEs using deep neural networks with error control **刘新亮**：Multigrid Neural Operator and Preconditioner: Operator Learning and Fast Helmholtz Solver **金鹏展**：A deformation-based framework for learning solution mappings of PDEs defined on varying domains	郭汝驰 黄政宇	友谊宫 8 号 会议室

	C8 专题：AI for Physics and Chemistry（II） **杨斌：**人工智能赋能的燃烧反应动力学模型发展 **王兴建：**复杂流动与燃烧过程的数据驱动降阶代理模型研究 **张天汉：**结合领域结构化知识的流体数值仿真智能体方法 **王柏森：**基于混合机器学习架构的复杂流场长期高保真预测方法研究	陈正	友谊宫 10 号 会议室
12:10-13:30	**午餐**		见餐券
13:30-15:30	**D1 专题：强化学习理论与算法（II）** **杨耀东：**欺骗性对齐机理与方法 **温颖：**面向大模型智能体的强化学习 **陈志平：**A Normalizing Flows-based Deep Reinforcement Learning Algorithm for Mean-Field Games **杨天培：**多智能体强化学习与 AI Agent 研究	魏轲 李帅	友谊宫 5 号 会议室
	D2 专题：机器学习和逼近理论（II） **谢和虎：**从有限元到机器学习 **蔡永强：**Neural Networks, Dynamical Systems, Control Families, and Formal Languages **陆帅：**Norm spaces rooted in neural networks and their applications **郭正初：**Learning theory of spectral algorithms under covariate shift **张耀宇：**The Condensation Phenomenon of Deep Learning	吴磊 何俊材	友谊宫 1 号 会议室
	D3 专题：机器学习与统计 **赵俊龙：**Approximation error from discretizations and its applications **胡天阳：**Connections between context data and model weights in transformers **程宇骞：**Uniform Inference for Kernel Gradient Flow Regression **丁嘉麟：**Over-parameterization Leads to Adaptivity in High Dimensional Gaussian Sequence	林乾	友谊宫 11 号 会议室
	D4 专题：机器学习与优化理论 **张景昭：**Progress and open problems in structured optimization	方聪	友谊宫 2 号

	江如俊：Accelerated Gradient Descent by Concatenation of Stepsize Schedules **方聪：**随机梯度下降算法在高维回归问题中正则效应与泛化性能分析 **叶海山：**FZOO: Fast Zeroth-Order Optimizer for Fine-Tuning Large Language Models towards Adam-Scale Speed		会议室
	D5 专题：深度学习与科学计算 **张瑞：**OmniFluids: Unified Physics Pre-trained Modeling of Fluid Dynamics **蔡声泽：**复杂流场环境的智能感知与控制 **孙赫：**基于神经 PDE 求解器的生物医学反散射成像 **陈云天：**基于人工智能的科学知识自动发现	孙浩	友谊宫 7 号 会议室
	D6 专题：机器学习与复杂系统 **樊京芳：**地球系统复杂性及 AI **胡延庆：**Restoring Network Evolution with Transferable Graph-Based Machine Learning **赖志路：**凸优化框架下的物理信息机器学习 **彭昊：**基于时空信息转换的高维复杂系统预测与表征算法研究 **高婷：**Latent Iterative Refinement Flow: A Geometric-Constrained Approach for Few-Shot Generation	冷思阳	友谊宫 10 号 会议室
	D7 专题：大模型数据准备 **张文涛：**Data-centric AI 基础设施 **陆鸣：**具身智能 VLA 大模型的数据研究 **张远行：**视频生成背后的多模态理解技术 **陈冲：**多源多模态的下一代 rag	张文涛	友谊宫 4 号 会议室
	D8 专题：生成式 AI 交叉研究 **陈阳：**智能医学成像和处理 **符天凡：**深度学习赋能的药物发现与开发 **李秋熠：**Towards AI for Genomics: GENERator & GENERanno **张强：**语言与知识驱动的科学智能体	张强 李柱	友谊宫 8 号 会议室

16:00-18:00	**学生分会 E1 专题：AI 理论（I）** **陈宗昊：**(De)-regularized Maximum Mean Discrepancy Gradient Flow **刘雨濠：**Context-Size Scaling for Operator and In-Context Learning **王梓麟：**SGD Achieves Optimality for Least Squares via Power-Decay Learning Rates **赵佳杰：**Architecture induces invariant manifolds of neural network training dynamics	王梓麟	友谊宫 7 号会议室
	学生分会 E2 专题：AI 理论（II） **朱同天：**DICE: Data Influence Cascade in Decentralized Learning **周展鹏：**The Underlying Mechanism behind Deep Learning: From Empirical Discoveries to Theoretical Attempts **杨若峰：**Why Rectified Flow is Better? Elucidating VP, VE and RF-based diffusion models **陈焕然：**Diffusion for Discriminative Modeling and Certification	周展鹏	友谊宫 5 号会议室
	学生分会 E3 专题：AI 算法 **李艺康：**Affine Equivariant Networks Based on Differential Invariants **游泽彬：**LLaDA-V：对扩散语言模型进行视觉指令微调 **王锦波：**The Sharpness Disparity Principle in Transformers for Accelerating Language Model Pre-Training	王锦波	友谊宫 4 号会议室
	学生分会 E4 专题：优化 **杨俨：**Bilevel Reinforcement Learning via the Development of Hyper-gradient without Lower-Level Convexity **孔博傲：**SPARKLE: A Unified Single-Loop Primal-Dual Framework for Decentralized Bilevel Optimization **姜林硕：**Stochastic optimization over expectation-formulated generalized Stiefel manifold **何雨桐：**Subspace Optimization for Large Language Models with Convergence Guarantees	何雨桐	友谊宫 8 号会议室
	学生分会 E5 专题：AI and PDE (I) **曾祉竣：**In vivo 3D ultrasound computed tomography of musculoskeletal tissues with generative neural PDE solvers **宋昊泽：**Redefining Neural Operators in ξd + 1ξ Dimensions	曾晨宇	友谊宫 2 号会议室

	李志豪：Harnessing Scale and Physics: A Multi-Graph Neural Operator Framework for PDEs on Arbitrary Geometries **曾晨宇：**Point Cloud Neural Operator for Parametric PDEs on Complex and Variable Geometries		
	学生分会 E6 专题：AI and PDE (II) **蔡志强：**Weak Generative Sampler to Sample Invariant Distribution of Stochastic Differential Equation **周金蕊：**Data-driven optimized high-order WENO schemes with low-dissipation and low-dispersion **卜凡：**A physics-informed deep learning method for solving hydrate dissociation problems in sediment **张晨晔：**High Order Integrated Reconstruction for Finite Volume Scheme **苏华：**SPIKE: stable physics-informed kernel evolution method for solving hyperbolic conservation laws	张晨晔	友谊宫 10 号会议室
	学生分会 E7 专题：AI for Science (I) **刘鹏伟：**AeroGTO: An Efficient Graph-Transformer Operator for Learning Large-Scale Aerodynamics of 3D Vehicle Geometries **张振毅：**Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport **李瑞堃：**Predicting Dynamical Systems across Environments via Diffusive Model Weight Generation **李沛函：**基于混合机器学习架构的复杂流场长期高保真预测方法研究	张振毅	友谊宫 1 号会议室
	学生分会 E8 专题：AI for Science (II) **左维：**面向机器学习的材料数据质量评价体系与指标构建方法 **李瑞凤：**统一匹配框架：少样本场景下的分子性质预测任务新解 **周倩：**面向高比能固态锂电池的聚合物电解质——从分子设计到智能预测	李瑞凤	友谊宫 11 号会议室

四、大会邀请报告摘要及报告人简介

大会邀请报告一

AI for Mathematics：数学的数字化与智能化

（董彬，北京大学）

报告摘要：数学研究中存在诸多限制研究效率的瓶颈问题，亟需人工智能技术的赋能。“AI for Mathematics”（AI4M）正是在这样的背景下兴起的一个新兴交叉研究领域，其核心任务包括定理的自动证明与证伪、数学语义搜索，以及数学知识的自动形式化等。本报告将首先从数学研究自身面临的挑战与需求出发，探讨数学研究为什么需要 AI 技术的深度赋能；随后，报告将介绍近几年 AI4M 领域的一些代表性工作，并分析不同技术路线各自的优势与局限性。在此基础上，报告将进一步指出，若要显著提升 AI 的数学推理能力，关键在于推进数学知识的形式化，即数学的“数字化”。接下来，报告将详细介绍北京大学 AI4M 团队的整体研究规划，并分享团队近期取得的一些阶段性研究成果。最后，报告将展望未来，表达对 AI 的期许：不仅能辅助处理数学研究中机械繁琐的证明过程，更有潜力推动数学不同领域知识的深度融合，助力数学家聚焦于探索数学本质，开启更富创造力的研究时代。

报告人简介：董彬，北京大学博雅特聘教授，任职北京大学北京国际数学研究中心，兼任北京大学国际机器学习研究中心副主任、北京中关村学院常务副院长。主要研究领域包括机器学习、科学计算、计算成像。于 2014 年获得求是杰出青年学者奖，2022 年受邀在世界数学家大会（ICM）做 45 分钟报告，2023 年入选新基石研究员项目，同年获得王选杰出青年学者奖，并受邀在 2027 年国际工业与应用数学大会（ICIAM）做邀请报告。

大会邀请报告二

Modeling LLM Pre-Training Dynamics with Functional Scaling Laws

（吴磊，北京大学）

报告摘要：Understanding the pre-training dynamics of large language models (LLMs) is critically important. However, traditional optimization theory fails to account for many intriguing phenomena observed in LLM pre-training, including the emergence of scaling laws and the widespread adoption of warmup–stable–decay (WSD) learning rate schedules. In this talk, we reveal a surprising alignment between the loss curves of LLM pre-training and those of power-law kernel regression. Motivated by this observation, we develop a theoretical framework of Functional Scaling Laws (FSL), which accurately captures the loss dynamics through the central notion of intrinsic time. Remarkably, FSL not only exhibits strong predictive power for LLM pre-training but also provides a principled approach for explaining why certain learning rate schedules work so well in practice.

报告人简介：吴磊，北京大学数学科学学院与国际机器学习研究中心助理教授，主要研究方向为深度学习的数理基础。2012 年毕业于南开大学，获数学与应用数学学士学位；2018 年毕业于北京大学，获得计算数学博士学位。2018 年 11 月至 2021 年 10 月，先后在美国普林斯顿大学与宾夕法尼亚大学从事博士后研究工作。 相关成果发表于 NeurIPS、ICML、AoS, JMLR 等国际顶级会议与期刊。

大会邀请报告三

大模型新架构的初步探索与思考

（林洲汉，上海交通大学）

报告摘要：Decoder-only 的 Transformer 是目前几乎所有预训练大语言模型所采用的默认模型架构。对于不同大小的模型，Scaling Law 也成为这些模型所能达到的基础能力的可靠指引。与此同时，学术界的探索性研究已经出现了一些不同于经典 Transformer 的新的模型架构，这些架构能在不同方面优于经典 Transformer。在这一讲座中，我们将从 Scaling Law 入手，结合相关公开发表的工作，介绍 3 个不同方向上的新模型架构探索。

报告人简介：林洲汉，理学博士、上海交通大学人工智能学院副教授、John Hopcroft 计算机科学中心副主任，国家海外高层次青年人才、上海市浦江学者。博士师从于深度学习领域图灵奖得主 Yoshua Bengio，目前主要从事机器学习与自然语言处理方向的研究，已发表学术论文 70 余篇，谷歌学术总引用量 10000 余次。他曾于 Facebook AI Research（FAIR）、Google AI、Microsoft Research、IBM Watson 等核心 AI 研究部门实习或工作。担任 Journal of Machine Learning Research（JMLR）、IEEE Transactions on Audio, Speech and Language Processing（TASLP）、 IEEE Transactions on Neural Networks and Learning Systems（TNNLS）等国际权威期刊的审稿人，ICLR、NeurIPS、ICML、AAAI、ACL、EMNLP、NAACL、AACL 等国际顶级会议的审稿人，以及 EMNLP、AAAI、AACL、COLING 会议的领域主席。

大会邀请报告四

理解和改进大模型的训练: 一些新进展

（孙若愚，香港中文大学（深圳））

报告摘要：本次报告讨论对大型语言模型(LLMs)训练算法的理解和提升。在第一部分 中，我们分析为什么 Adam 在 Transformer 上优于 SGD，并提出一种轻量级的替代方法— —Adam-mini。我们解释了 SGD 在 Transformer 上的失败原因:(i) Transformer 是“异质性 的”:参数块之间的 Hessian 谱差异显著;(ii) 异质性阻碍了 SGD:SGD 在存在块异质性的 问题上表现不佳。受此发现启发，我们引入了 Adam-mini，它为每个块中的所有权重分配 了一个单一的二次动量项。我们通过实验证明，Adam-mini 在不牺牲性能的情况下，相比 Adam 节省了 35-50%的内存，在包括 8B 规模的语言模型和 ViT 在内的多种模型上表现出 色。在第二部分中，我们介绍 MoFO，一种在 LLMs 的 SFT 阶段减轻遗忘效应的新算法。 我们观察到遗忘的一个原因是 SFT 后权重偏离了预训练权重，因此提出一种结合了 Adam 和贪心坐标下降法的新算法 MoFO，并给出收敛分析。和传统的混合预训练数据和 sft 数 据的算法相比，我们的方法不需要使用预训练数据，且在 7B 模型实验中效果显著更好。 在第三部分中，我们介绍一种新的 RLHF(基于人类反馈的强化学习)算法 ReMax。我们 指出未被经典 PPO 方法充分利用的 RLHF 任务的三个特性，并提出一个新方法 ReMax。 ReMax 减少了 50%的 GPU 内存使用，并将训练加速了 1.6 倍，在 Mistral-7B 模型测试中表 现优于 PPO 和 DPO。

报告人简介：孙若愚，香港中文大学（深圳）数据科学学院副教授、博士生导师。此前他于 2017 年至 2022 年任伊利诺伊大学香槟分校（UIUC）助理教授、博士生导师，2016 年任脸书人工智能研究所（由 LeCun 领导）全职访问科学家，2015-2016 年任斯坦福大学博士后研究员。他 2015 年在美国明尼苏达大学电子与计算机工程系获得博士学位，2009 年在北京大学数学科学学院基础数学系获得本科学位。他的主要研究领域为人工智能和机器学习、数学优化理论与算法、无线通信和信号处理等，具体研究方向包括神经网络理论和算法、生成模型、大数据优化算法、学习优化、通信网络容量理论与优化算法等。他曾获得 INFORMS(国际运筹与管理协会) George Nicolson 学生论文竞赛第二名，以及 INFORMS 优化协会学生论文竞争荣誉奖。在人工智能与机器学习会议 NeurIPS, ICML, ICLR, AISTATS，顶尖信息论与通信杂志 IEEE transaction on information theory, IEEE Signal Processing Magazine, Journal of Selected Areas in Communications，顶尖数学优化与运筹杂志 Mathematical Programming, SIAM Journal on Optimization, Math of Operations Research 等会议与杂志发表数十篇文章。目前担任 NeurIPS, ICML, ICLR, AISTATS 等人工智能会议的领域主席。

五、分会报告摘要

分会报告 A1 专题：深度学习理论

A1-1 Towards understanding the representation learning of diffusion models

邹荻凡, 香港大学

摘要： Diffusion models (DMs) excel in generative modeling, but their theoretical foundations and limitations remain underexplored. This talk addresses two key aspects: their feature learning dynamics and their ability to capture hidden inter-feature rules. First, I show that the denoising objective encourages DMs to learn balanced and comprehensive data representations, unlike classification models that prioritize easy-to-learn patterns. Theoretical analysis and experiments on synthetic and real-world datasets highlight this distinction. Next, I explore a critical limitation: DMs often fail to learn fine-grained hidden rules between dependent features, such as the relationship between the height of the sun and shadow length in images. Empirical evaluations on models like Stable Diffusion reveal consistent failures, supported by synthetic tasks and theoretical insights showing that denoising score matching (DSM) is incompatible with enforcing rule conformity. I discuss potential solutions, such as classifier-guided sampling, and their limitations. This talk provides a deeper understanding of DMs' strengths and weaknesses, offering insights for building more robust and interpretable generative models.

A1-2 自注意力机制中的变量选择：案例研究与理论理解

曹原, 香港大学

摘要： Transformer 已成为机器学习领域的主流力量，在各类应用中展现出前所未有的成功。其以自注意力机制为核心的独特架构，革新了模型处理数据的方式。本报告将探讨自注意力机制中变量选择的若干理论案例。我们首先展示单层 Transformer 模型如何通过梯度下降得以成功训练，实现上下文中的最近邻预测。随后，我们将讨论单层 Transformer 在学习变量选择以及解决具有组稀疏性的线性回归问题上的能力。此外，我们还将研究简单 Transformer 模型在学习随机游走过程中的表现。这些研究的核心在于分析 softmax 自注意力机制如何通过训练实现合理的变量选择。相关结论展示了 Transformer 模型对多种经典统计模型的良好的兼容性与适应性。

A1-3 Bridge theory to practice at scale: One-step gradient suffices for fine-tuning LLMs, provably and efficiently

刘方辉, University of Warwick

摘要： In this talk, I will illustrate how theory can guide practice, through the lens of low-rank adaptation (LoRA) for fine-tuning large language models. The results include three aspects: 1) Our theoretical results show that LoRA will align to the certain singular subspace of one-step gradient of full fine-tuning. Hence, the subspace alignment and generalization guarantees can be directly achieved by a well-designed spectral initialization strategy for both linear and nonlinear models. 2) Our analysis leads to the LoRA-One algorithm, a theoretically grounded algorithm that achieves significant empirical improvement over vanilla LoRA and its

variants on several benchmarks by fine-tuning Llama 2. Additionally, our results also clarify some misconceptions in previous algorithm design. Talk is based on https://arxiv.org/abs/2502.01235 (ICML'25 oral)

A1-4 Beyond Unconstrained Features: Neural Collapse for Shallow Neural Networks with General Data

凌舒扬, 上海纽约大学

摘要：Neural collapse (NC) is a phenomenon that emerges at the terminal phase of the training (TPT) of deep neural networks (DNNs). The features of the data in the same class collapse to their respective sample means and the sample means exhibit a simplex equiangular tight frame (ETF). In the past few years, there has been a surge of works that focus on explaining why the NC occurs and how it affects generalization. Since the DNNs are notoriously difficult to analyze, most works mainly focus on the unconstrained feature model (UFM). In this work, we focus on shallow ReLU neural networks and try to understand how the width, depth, data dimension, and statistical property of the training dataset influence the neural collapse. We provide a complete characterization of when the NC occurs for two or three-layer neural networks. For two-layer ReLU neural networks, a sufficient condition on when the global minimizer of the regularized empirical risk function exhibits the NC configuration depends on the data dimension, sample size, and the signal-to-noise ratio in the data instead of the network width. For three-layer neural networks, we show that the NC occurs as long as the first layer is sufficiently wide. Regarding the connection between NC and generalization, we show the generalization heavily depends on the SNR (signal-to-noise ratio) in the data. Our results significantly extend the state-of-the-art theoretical analysis of the NC under the UFM by characterizing the emergence of the NC under shallow nonlinear networks and showing how it depends on data properties and network architecture.

分会报告 A2 专题：神经科学和人工智能（I）

A2-1 高效脉冲神经网络模型训练算法研究

唐华锦, 浙江大学

摘要：脉冲神经网络（SNN）以其事件驱动与异步计算特性，展现出低功耗和高效率的计算优势。然而，由于脉冲发放过程具有非连续性和不可微性，传统误差反向传播算法难以直接应用于 SNN 训练。当前主流的代理梯度方法采用固定平滑系数的近似函数进行端到端训练，但这种方法导致梯度失配问题，限制了网络的收敛效率与泛化性能。针对上述问题，本报告介绍自适应平滑梯度学习算法，该算法构建了一个与目标 SNN 共享动力学特性的带脉冲噪声的混合代理网络，使平滑系数在混合模式训练过程中自适应地优化，显著提升了 SNN 的泛化性能与稳定性。

A2-2 脉冲神经网络结构学习方法研究

徐齐, 大连理工大学

摘要：脉冲神经网络（SNN）是基于生物神经机制的大脑启发模型，被誉为第三代神经网络模型，具有功耗低、实时性强、生物可解释性高等优点。然而，大多数 SNN 采用人工神经网络（ANN）的固定结构，未能充分发挥 SNN 通过脉冲序列表示时空特征的特点和优势。受生物神经突触生长、修剪和再生的突触可塑性机制的启发，本次报告将介绍结构可学习的脉冲神经网络模型与有监督学习算法，通过利

用神经兴奋-抑制机制来动态调整网络的连接状态，促进了参数局部和全局学习方法的高效混合，形成拓扑结构优化的脉冲神经网络模型。通过融合脉冲神经网络的神经机制和深度神经网络的计算优势，为高效类脑脉冲模型的开发和应用提供一种新的思路。

A2-3 从感知到认知的神经动力学机制

杨冬平, 之江实验室

摘要：活的大脑具有复杂多变的内部活动，却可以可靠地产生适应性响应来处理外部信息。因此，一个长期存在的基本问题是:神经网络如何在具有复杂多变的内部活动下产生对外界信息的可靠表征及编码?我们的研究揭示了神经异质性在可靠信息处理中的关键作用，并提出了一个统一的内在动力学机制来阐明各种神经异质性的作用。我们的研究发现神经异质性可以打破神经网络的内在活动模式，使得神经元的活动更独立灵活、放电率更异质，增强了神经元对外界信号的敏感性，在外界输入的诱导下形成与输入信号对齐的稳定暂态活动模式，进而获得可靠的信息表征。而大脑的认知需要进一步将感知信息映射到高维活动中的低维流形空间进行处理，我们的研究进一步发现这种映射可以处理任务的迁移，实现认知的泛化。

A2-4 大规模脉冲神经网络理论与方法

田永鸿, 北京大学

摘要：纵观科学技术发展史，借鉴和模仿大脑一直是智能技术创新的源泉。人类大脑是一个高效智能的超大规模生物脉冲网络。作为脑网络启发的第三代神经网络，目前脉冲神经网络（SNN）在规模和性能上均不及深度神经网络，面对复杂任务的泛化性和灵活性存在严重不足，因此迫切需要借鉴大脑机制发展大规模脉冲学习理论与方法。本报告将围绕“兼具生物合理和计算有效的大规模脉冲神经网络模型与学习”这一主题，重点分享在神经元建模、脉冲网络生长与剪枝、深度脉冲神经网络学习等三方面的挑战问题与研究进展，并探讨在神经形态视觉领域的创新应用。

分会报告 A3 专题：Optimization for AI

A3-1 面向生成模型的高效训练与推理算法

林涛, 西湖大学

摘要：深度生成模型在提升性能的同时，通常也伴随着巨大的计算开销，这在训练和推理阶段都构成了实际的挑战。如何提高生成模型的效率和性能，是当前领域关注的重点。 本次报告将介绍两种新的技术路径。 首先是 GMem，一种模块化的生成模型方法。它通过解耦记忆与泛化，将关键的语义信息存储于一个独立的外部模块中，从而降低了模型对网络自身规模的依赖。实验结果表明该方法能显著提升训练效率。 其次，我们将讨论一个统一的连续生成模型框架 (UCGM)。该框架旨在整合不同的采样方法，如多步扩散和少步一致性模型，提供统一的训练和采样流程。应用该框架，不仅可以训练出在极少步数（例如 2 步）内就达到高质量的模型，也能用于优化现有的预训练模型，在大幅减少采样步数（减少 84%）的情况下获得性能提升。

A3-2 Physics-Assisted and Topology-Informed Deep Learning for Weather Prediction

凌青, 中山大学

摘要： Although deep learning models have demonstrated remarkable potential in weather prediction, most of them overlook either the physics of the underlying weather evolution or the topology of the Earth's surface. In light of these disadvantages, we develop PASSAT, a novel Physics-ASSisted And Topology-informed deep learning model for weather prediction. PASSAT attributes the weather evolution to two key factors: (i) the advection process that can be characterized by the advection equation and the Navier-Stokes equation; (ii) the Earth-atmosphere interaction that is difficult to both model and calculate. PASSAT also takes the topology of the Earth's surface into consideration, other than simply treating it as a plane. With these considerations, PASSAT numerically solves the advection equation and the Navier-Stokes equation on the spherical manifold, utilizes a spherical graph neural network to capture the Earth-atmosphere interaction, and generates the initial velocity fields that are critical to solving the advection equation from the same spherical graph neural network. In the 5.625-resolution ERA5 data set, PASSAT outperforms both the state-of-the-art deep learning-based weather prediction models and the operational numerical weather prediction model IFS T42.

A3-3 On the Complexity of Distributed Nonconvex Optimization

罗珞, 复旦大学

摘要： We consider the incremental first-order optimization (IFO) for distributed nonconvex optimization. We first revisit the problem setting in single machine scenario by distinguishing the difference between global and mean-squared smoothness parameters. The key observation is that the optimal IFO complexity is indeed achieved by the trade-off between variance reduction methods and classical gradient descent. We then design the distributed algorithm by introducing the new sampling strategy that allows different mini-batch sizes on different nodes. The theoretical analysis shows the IFO calls, the computational rounds, and the communication rounds of our algorithms are near-optimal. We can extend our results to the problem with PL condition, which also achieve the near-optimal upper complexity bounds.

A3-4 Distributed Learning over Arbitrary Topology: Linear Speed-Up with Polynomial Transient Time

濮实, 香港中文大学（深圳）

摘要： We study a distributed learning problem in which n agents, each with potentially heterogeneous local data, collaboratively minimize the sum of their local cost functions via peer-to-peer communication. We propose a novel algorithm, Spanning Tree Push-Pull (STPP), which employs two spanning trees extracted from a general communication graph to distribute both model parameters and stochastic gradients. Unlike prior approaches that rely heavily on spectral gap properties, STPP leverages a more flexible topological characterization, enabling robust information flow and efficient updates. Theoretically, we prove that STPP achieves linear speedup and polynomial transient iteration complexity under arbitrary network topologies.

分会报告 A4 专题：大模型数据智能

A4-1 从异构科学数据到统一模型输入：面向科学智算大模型的数据基础设施构建

周号益, 北京航空航天大学

摘要：科学智算大模型的成功离不开高质量、大规模的训练与评估数据，然而科学数据天然具有异构性，例如 PDE 数据跨越不同维度、天文数据包含不同几何结构、电磁数据耦合不同时间周期尺度，如何将这些异构数据转化为统一的模型输入是当前面临的核心挑战。本报告将系统介绍在科学数据基础设施构建方面的探索与实践。首先，设计了科学数据的跨维度、跨几何、跨尺度统一处理方法，形成了流体、天文、电磁等领域科学智算大模型；其次，构建了 CNAI4S 科学智算共性平台，涵盖流体、材料、气象、海洋等 12 个科学领域的大规模基准数据库（≥482 万条结构化数据），并通过 PDENNEval、SuperCon3D、CFDNNEval 等评测基准体系验证了数据质量与模型性能。通过科学智算大模型提供数据统一处理方法和标准化基础设施的探索，有望支撑下一代科学智算大模型的训练与评估，进一步发现科学领域的大模型 Scaling Law。

A4-2 数据视角下的模型压缩加速

张林峰, 上海交通大学

摘要：大模型的计算成本严重制约了其落地应用。一般来说，模型的计算成本由其参数量与数据量共同决定。已有压缩研究主要关注如何减少模型的参数量而忽视了数据维度的压缩。随着强推理模型和视频生成模型的出现，我们发现数据规模（Token 数量）的增加已经成为了计算成本居高不下的首要因素。在本报告中，我们将介绍数据中心的模型压缩加速在大模型、多模态大模型、图像视频生成模型上的几个典型案例。

A4-3 大模型数据准备的“IaaS”原理

周煊赫, 上海交通大学

摘要：如今，大模型在通用和专用领域应用中都取得了显著进展。然而，其成功离不开高质量数据的“喂养”。本报告首先介绍大模型数据的"IaaS"概念，即高质量的大模型数据应具备四个关键特性：（1）包含性（Inclusiveness）：确保数据覆盖广泛的领域和类型；（2）冗余性（Abundance）：通过适度的数据重复增强模型的学习效果；（3）高质量（Articulation）：保证数据的准确性、相关性和有用性;(4)无害化（Sanitization）：确保数据经过伦理审查，不包含有害内容或隐私信息。这一框架贯穿于 LLM 的各个阶段，包括预训练、持续预训练、微调、强化学习、检索增强生成（RAG）、LLM 代理和评估等。围绕“IaaS”框架，报告还将介绍 LLM 全生命周期中的数据处理技术，包括数据去重、数据过滤、数据混合与选择、数据成与标注等（https://github.com/weAIDB/awesome-data-llm）。

A4-4 MinerU: 精准解析文档，驱动 AI 应用

王斌, 上海人工智能实验室

摘要：在大模型时代，精准的文档解析是实现高质量文档问答、财报分析、论文研究等智能应用（RAG）的先决条件。面对多样化文档带来的解析挑战，我们推出了开源、高效的文档解析器

MinerU。本报告将阐述 MinerU 如何攻克复杂文档解析难题，为大模型应用提供坚实的数据基础，并展望该领域未来的发展方向。

分会报告 A5 专题：科学机器学习

A5-2 Bayesian Learning for Compact Dynamical Representations of Nonlinear Systems

郭孟武, Lund University (Sweden)

摘要：Credible real-time simulation is a crucial enabler for digital twin technology, and data-driven model reduction is a key approach to achieving it. In this talk, we will discuss non-intrusive Bayesian methods for learning reduced-order representations of high-dimensional dynamical systems, with built-in probabilistic quantification of modeling uncertainties to certify computational reliability. The core strategy involves using Bayesian inference for the parametrization inspired by projection-based model reduction. Particularly, Gaussian process approximations are leveraged to formulate differential-equation-constrained likelihood functions and hence improve predictive performance, especially when training data are noisy and/or scarce. These techniques have demonstrated their effectiveness in data-driven reduced-order modeling by delivering accurate temporal predictions along with robust uncertainty quantification.

A5-3 基于物理信息机器学习的反应堆多物理场耦合建模与数字孪生构建

龚禾林, 上海交通大学

摘要：针对核反应堆多物理场耦合模拟中存在的计算效率低、参数不确定性大等关键科学问题，本研究提出了一种融合物理机理与实验数据的机器学习建模方法。通过将反应堆控制方程等物理约束嵌入神经网络损失函数，构建了具有明确物理意义的替代模型，实现了多物理场的高效精确预测。在此基础上，发展了基于替代模型的反应堆数字孪生系统构建方法，解决了传统数值模拟方法难以满足实时性要求的难题。研究重点突破了物理信息神经网络架构设计、多物理场耦合机制嵌入、模型不确定性量化等关键技术，形成了完整的理论方法体系。通过典型反应堆工况下的数值实验验证，所提方法在保证计算精度的同时，将计算效率提升了 2 个数量级，为反应堆智能化设计、运行和安全评估提供了新的技术途径。

A5-4 DeepSPoC: a deep learning based sequential propagation of chaos

周元诚, 上海师范大学

摘要： Sequential propagation of chaos (SPoC) is a recently developed tool for solving mean-field stochastic differential equations and their related nonlinear Fokker-Planck equations. Based on the theory of SPoC, we present a new method (DeepSPoC) that combines the interacting particle system of SPoC with deep learning. A recently developed deep generative model called KRnet is used to store the empirical measure of particles in our algorithm. Our method has computational complexity O(N) with respect to particle number and can also significantly reduce the memory used to store particle trajectories. These two features make our method applicable to the computation of complex high-dimensional problems that require simulations of large particle systems. We apply our method to a wide range of different types of mean-field equation and verify its effectiveness and advantages.

A5-5 基于混合神经网络的多保真度不确定度量化方法

于腾超, 北京应用物理与计算数学研究所

摘要：基于混合神经网络的多保真度不确定度量化方法 基于贝叶斯准则的不确定度量化方法对数据的数量和质量均有较高的要求。然而对于实际工程问题，实验数据获取困难且昂贵，高精度数值模拟计算耗时且计算消耗大，会面临小样本问题。决小样本问题下复杂系统的不确定度量化是适应工程需求的重要问题。近年来，神经网络的迅速发展为复杂系统的代理提供了强有力的支持，贝叶斯神经网络的应用可以对复杂系统进行概率建模。与此同时，使用多保真度方法，通过利用低成本的低保真度数据对高保真度据进行补充可以减少不确定度量化对高保真度数据的需求。本工作提出一种基于混合神经网络的多保真度不确定度量化方法，并通过相关数值算例对方法进行了验证。

分会报告 A6 专题：机器学习与数值方法

A6-1 算测融合的装备形性一体化数字孪生技术

宋学官, 大连理工大学

摘要：数字孪生是近些年提出的一种新的数字化概念，在智慧城市、黑灯工厂等中应用广泛，但是在重大装备中的研究和应用还十分匮乏，主要是因为孪生模型在计算精度、计算速度、数据融合等方面存在诸多难点。那么，数字孪生是否能够应用到装备的设计、制造以及维护上呢？或者，到底什么是装备级的数字孪生？其难点与挑战是什么？该如何攻破呢？本报告将针对以上问题，进行简单的阐述与探讨。本次报告分为四个部分：现状、挑战、途径与感悟，分别介绍数字孪生的研究现状与产品级数字孪生的根本目的——形性一体化；面向形性一体化数字孪生时，物理体与数字体面临的“算不了”“算不快”“算不准”“测不了”“测不准”和“测不全”六个主要难题；针对这些难题的可行解决措施与技术途径（算测融合），并通过起重装备、高压系统、数字骨骼、光机系统等几个典型演示案例进行针对性讲解；最后，分享本人对数字孪生与重大装备结合的一些思考。

A6-2 数据驱动的流体力学知识发现与 AI4E 的应用

张伟伟, 西北工业大学

摘要：人工智能为流体力学的发展提供了新的研究范式。基于数据，不仅可以通过深度学习提供一个复杂流体系统的黑箱表达，而且利用符号主义的若干方法，为工业知识发现提供了新的契机。报告将展示团队近年在含噪数据驱动的流体力学偏微分方程识别，数据驱动的精准混合长公式发现及其新型代数湍流模型的构建，适用于高 Re 分离流湍流输运方程的构造及其复杂流动模拟，气动载荷关联参数的符号回归与流体动力学系统的流形降维等方面的初步进展。研究表明，相比于黑箱表达的深度学习方法，数据驱动的白箱机器学习方法，以及数据和知识的融合建模方法，具备天生的可解释性和信任性，也具有很强的小样本学习能力，模型的泛化性强。相关研究不局限于流体力学领域，为小样本约束下的工业人工智能研究提供了很好的示范。

A6-3 基于张量分解的高速背景纹影层析用于 4D 流场密度重建

蔡伟伟, 上海交通大学

摘要：背景纹影法（Background Oriented Schlieren, BOS）是一种低成本的流场测量方法，能够获取流场中的折射率、密度和温度信息。提高其测量精度和效率以实现三维动态测量一直是研究的重点。这项研究提出了一种基于混合张量分解和神经网络的高效背景导向纹影层析成像（Background Oriented Schlieren Tomography, BOST）算法，该算法通过融合张量降维和深度学习技术，优化了传统方法的计算效率和精度，实现 4D 折射率场和密度场的重建。该算法构建了一个四维时空张量 XYZT，并采用了一种使用三组独立分量空间矩阵和时空矩阵的分解策略。结合特征空间分析和轻量级网络，实现了对复杂流场的高效表征。同时，研究中引入了一个专用的校正神经网络来增强动态畸变校正能力，通过对背景板上光线位移的迭代优化来补偿动态畸变。该框架采用梯度下降优化策略，在经过双线性插值、神经网络和光线追踪后进行，以更新张量分量和神经网络参数，并采用混合精度加速来加快收敛速度。研究中使用不同算法对计算流体动力学（CFD）基准算例进行三维重建的结果对比。实验结果表明，该框架在保持重建精度的同时显著降低了计算复杂度。在无需任何预训练的情况下，相比于当前最先进的神经辐射场算法误差降低了 15%以上、同时重建速度提高了两个数量级，实现了每帧 12.6 秒的重建速度（还有进一步加速的潜力）。这项工作为复杂流场的实时可视化提供了一种新方法。

A6-4 数智赋能航空发动机关键技术思考与研究进展

巴顿, 中国科学院工程热物理研究所

摘要：面向下一代先进航空发动机技术发展的重大需求，围绕复杂边界与多物理场耦合愈为紧密、设计工作愈加繁重、复杂环境试验成本/风险愈加巨大的挑战，探索数智化赋能航空发动机设计的创新解决思路，重点介绍 AI 赋能仿真、AI 赋能设计与 AI 赋能试验的思考，进一步介绍团队在航空发动机数值仿真、航空压气机/涡轮气动设计、航空发动机虚实交互显示和航空发动机智能流动控制方面的研究进展，并共同研讨数智化赋能航空发动机技术实施路线。

A6-5 AI 赋能工业 3-D 空气动力学设计与智能优化

邬晓敬, 中科工业人工智能研究院

摘要：三维构型气动优化设计正面临着传统参数化方法所致高维参数空间以及高昂 CFD 数值模拟代价的高维昂贵问题，这阻碍了工程研发提质增效与复杂布局创新设计。随着数据驱动的人工智能技术的发展，为复杂 3-D 构型的几何高效优化带来了新的思路。本报告特征-气动力知识关联挖掘、知识迁移学习以及几何点云深度学习等三个方面发展了支撑 3-D 构型高效高精度气动优化设计新方法。以机翼、螺旋桨、汽车等 3-D 典型构型设计作为案例，验证了所发展方法的优势和效果。本报告研究工作为三维构型气动设计所面临的高维昂贵问题提供了崭新的研究思路和技术支撑。

分会报告 A7 专题：机器学习与科学计算（I）

A7-1 Rates for least squares using over-parameterized neural networks

杨云斐, 中山大学

摘要： Recent studies showed that deep neural networks can achieve minimax optimal rates for learning smooth function classes. However, most of these results require that the neural networks in use are under-parameterized, which cannot explain the successes of over-parameterized models used in practice. In this talk, we will discuss how to derive convergence rates for neural networks in the over-parameterized regime. We will begin with a discussion on the approximation capacity of ReLU neural networks with certain norm constraints on the weights. By using this result, we are able to prove nearly optimal learning rates for least squares estimations based on over-parameterized (deep or shallow) neural networks if the weights are properly constrained. Finally, we will also show how to obtain minimax optimal rates for shallow neural networks by using localization technique and generalize the results to regularized least squares.

A7-2 Fourier Multi-Component and Multi-Layer Neural Networks: Unlocking High-Frequency Potential

张仕俊, 香港理工大学

摘要： The architecture of a neural network and the selection of its activation function are both fundamental to its performance. Equally vital is ensuring these two elements are well-matched, as their alignment is key to achieving effective representation and learning. In this paper, we introduce the Fourier Multi-Component and Multi-Layer Neural Network (FMMNN), a novel model that creates a strong synergy between them. We demonstrate that FMMNNs are highly effective and flexible in modeling high-frequency components. Our theoretical results demonstrate that FMMNNs have exponential expressive power for function approximation. We also analyze the optimization landscape of FMMNNs and find it to be much more favorable than that of standard fully connected neural networks, especially when dealing with high-frequency features. In addition, we propose a scaled random initialization method for the first layer's weights in FMMNNs, which significantly speeds up training and enhances overall performance. Extensive numerical experiments support our theoretical insights, showing that FMMNNs consistently outperform traditional approaches in accuracy and efficiency across various tasks.

A7-3 Distribution Matching for Self-Supervised Transfer Learning

马文森, 武汉大学

摘要： In this paper, we propose a novel self-supervised transfer learning method called Distribution Matching (DM), which drives the representation distribution toward a predefined reference distribution while preserving augmentation invariance. DM results in a learned representation space that is intuitively structured and therefore easy to interpret. Experimental results across multiple real-world datasets and evaluation metrics demonstrate that DM performs competitively on target classification tasks compared to existing self-supervised transfer learning methods. Additionally, we provide robust theoretical guarantees for DM, including a population theorem and an end-to-end sample theorem. The population theorem bridges the gap between the self-supervised learning task and target classification accuracy, while the sample theorem shows that, even

with a limited number of samples from the target domain, DM can deliver ex ceptional classification performance, provided the unlabeled sample size is sufficiently large.

A7-4 Score-Based Sequential Langevin Sampling for Data Assimilation

袁成, 华中师范大学

摘要：In this presentation, we will introduce a deep learning approach for data assimilation. Grounded in the Bayesian inference framework, we leverage denoising score matching to learn the prior distribution. This method enables the explicit expression of the gradient of the log posterior, facilitating accurate predictions of the new state by sampling from the posterior distribution. Through a series of numerical illustrations and theoretical analyses, we will showcase the efficacy, superiority, and constraints of our innovative methodology.

A7-5 DRM Revisited: A Complete Error Analysis

吴佩颖, 武汉大学

摘要：It is widely known that the error analysis for deep learning involves approximation, statistical, and optimization errors. However, it is challenging to combine them together due to overparameterization. In this paper, we address this gap by providing a comprehensive error analysis of the Deep Ritz Method (DRM). Specifically, we investigate a foundational question in the theoretical analysis of DRM under the overparameterized regime: given a target precision level, how can one determine the appropriate number of training samples, the key architectural parameters of the neural networks, the step size for the projected gradient descent optimization procedure, and the requisite number of iterations, such that the output of the gradient descent process closely approximates the true solution of the underlying partial differential equation to the specified precision?

十二、分会报告 A8 专题：大模型训练

A8-1 大模型复杂推理技术

赵鑫, 中国人民大学

摘要：最近以 DeepSeek-R1 为代表的深度推理模型受到了较大关注，这种通过生成更长的思考过程来解决更具挑战性的问题，在多个科学场景和应用领域都取得了重要突破。本次报告将聚焦深度推理模型的基础技术与实现方法，对于其中可能涉及到的技术路径进行探索和讲解，主要介绍以强化学习为主线的关键技术，并结合自身实践经验讨论其中的技术挑战，然后探讨推理模型在智能信息获取方面的应用，并且总结现阶段推理模型的局限以及未来的技术发展趋势。

A8-2 大模型密度法则与高密度大模型关键技术

刘知远, 清华大学

摘要：2018 年以来大模型规模不断增大、产生智能涌现，验证了 OpenAI 提出的模型规模法则（Scaling Law），特别是 ChatGPT 的推出引发全世界对大模型技术的关注。面向未来，大模型的发展趋势是什么，就是不断增加模型参数规模以追求更多能力涌现么？本报告发现，大模型在印证规模法则的同时，

还呈现能力密度持续增强的规律，我们称为大模型的密度法则（Densing Law），这揭示了端侧智能的巨大潜力，并指出未来应持续探索大模型科学化建设路径，不断改进模型制造工艺，实现人工智能的高质量、可持续发展。本报告将介绍大模型的密度法则和实现高密度大模型的关键技术。

A8-3 初探大模型“智能涌现”现象：从线性到非线性

束俊, 西安交通大学

摘要：近年来，以 ChatGPT 和 DeepSeek 为代表的大模型技术取得显著发展。按照主流的认识，大模型的能力之所以强大源自它可能存在的智能涌现(Intelligent Emergence)。然而,什么是智能涌现? 是什么要素摧生了智能涌现?大模型在什么情况下才会出现智能涌现?本报告主要针对这些问题展开，介绍大模型的智能涌现及尺度率等基本知识，并提出一个数学框架和相应的数学理论来对智能涌现及尺度率加以解析，并揭示大模型完全不同于小模型的统计学习规律。

A8-4 Scalable Complexity Control Facilitates Reasoning Ability of LLMs

杭良慨, 上海交通大学

摘要：近年来，大型语言模型（LLMs）的推理能力迅速提升，这引发了人们对能够可靠增强其泛化能力的更为根本性方法的关注。本研究展示了通过调整初始化率和权重衰减系数来控制模型复杂度，这一方法可在不同模型规模和数据规模下持续改进 LLMs 的 Scaling Law。我们以在 1T token 上预训练的 2.4B 模型为例，比较了不同复杂度超参数设置下的基准性能，进一步说明了该方法的效果。研究发现，相较于固定初始化标准差，采用恒定的初始化率（即标准差的指数）能够使得 Scaling Law 在模型规模和数据规模两方面更快地下降。上述结果表明，复杂度控制是推动 LLMs 持续进步的一条有前景的路径。

分会报告 B1 专题：大模型相关理论

B1-1 大模型表达能力理论

贺笛, 北京大学

摘要：无

B1-2 大模型训练的最优学习率衰减与扩展定律

吕凯风, 清华大学

摘要：大模型训练成本极其高昂，难以在大规模训练中直接调整超参数，尤其是学习率及其衰减策略。本报告将介绍我们近期的研究工作：我们提出了一条基于多重幂律的扩展定律，仅需不超过三次训练，即可在包括常数学习率、余弦衰减、阶梯衰减等多种调度策略下，精确预测大模型预训练的损失曲线。更进一步，我们通过最小化最终训练损失的预测值，自动搜索出一种优于传统余弦衰减的学习率衰减策略。其形状与近期提出的 Warmup-Stable-Decay（WSD）策略类似，但在最终损失上表现更加出色。

B1-3 通过强化学习提升大语言模型的逻辑推理能力

邱凯, 微软亚洲亚洲研究院

摘要：我们探讨基于规则的强化学习（RL）在大规模推理模型中的潜力。受 DeepSeek-R1 成功的启发，我们通过合成逻辑谜题作为训练数据，分析推理动态。这些逻辑谜题因其可控的复杂性和简单的答案验证过程而成为理想的训练数据。本研究提出了几项关键技术贡献，包括强调思维和回答过程的系统提示、惩罚捷径输出的严格格式奖励函数，以及实现稳定收敛的简单训练方案。我们的 7B 模型在训练仅 5,000 个逻辑问题后，展示了在挑战性数学基准测试 AIME 和 AMC 上的泛化能力。

B1-4 大模型推理机制分析

刘勇, 中国人民大学

摘要：近年来，大模型推理算法在效率与性能方面实现显著突破，推理速度与准确率得以大幅提升。但算法创新的热潮背后，对大模型推理内在机制的系统性探究仍显不足，致使其推理能力的认知存在诸多盲区。本报告从 “外部慢思考” 与 “内部慢思考” 双维度切入，着重剖析大模型外部推理的能力边界，以及长思维链对内部推理机制的影响，旨在为后续推理算法的优化设计夯实理论基础，突破技术桎梏，推动大模型推理能力实现新的跨越。

分会报告 B2 专题：机器学习与科学计算（II）

B2-1 Solving Bayesian Inverse Problems via Diffusion-based Sampling

段晨光, 武汉大学

摘要：Bayesian inverse problems are fundamental to image science and scientific computing, with posterior sampling serving as the primary mechanism for quantifying solution uncertainty. Despite its importance, sampling from complex posterior distributions in high-dimensional spaces remains computationally challenging. This paper introduces a novel diffusion model-based algorithm for tackling high-dimensional nonlinear Bayesian inverse problems, where the posterior score is estimated through denoising Langevin dynamics. We provide theoretical guarantees for our method and demonstrate its effectiveness in sampling from even non-log-concave posterior distributions. Experimental validation across multiple image reconstruction tasks confirms our algorithm's superior performance compared to existing approaches.

B2-2 Schrodinger-Follmer Diffusion: Sampling, Optimization, Generative Learning

康利灿, 武汉大学

摘要：这个工作基于 Schrodinger-Follmer 扩散过程研究随机抽样，优化，和生成模型。

B2-3 几何等变先验嵌入的深度网络模块设计

谢琦, 西安交通大学

摘要：本报告以图像处理为例，探讨几何等变先验在深度网络设计中的重要性，重点介绍高精度旋转/尺度等变卷积、旋转等变隐式神经表示、旋转等变 Vision Transformer（ViT）、变换可学习等变卷积、

等新型网络基础模块的构建方法与基础理论；进一步地，本报告将通过医学自然图像处理、图像重建、多帧图像匹配等实际应用展示先验几何对称性的嵌入将显著提升模型的性能与泛化能力。

B2-4 Flow-based Sampling Method

丁钊, 武汉大学

摘要：Diffusion models have recently achieved remarkable success in generative learning. In this talk, we adapt similar ideas to address classical distributional sampling problems. Specifically, we model the sampling process as solving an initial value problem defined by a probability flow ODE. We propose several schemes for computing the associated velocity field, and demonstrate that the method is applicable to any unnormalized target distribution. In terms of Wasserstein distance, we provide theoretical guarantees on the sampling error and show that the method avoids the curse of dimensionality. We evaluate the proposed sampling approach on various mixture distributions. Experimental results show that it outperforms mainstream MCMC methods, particularly in challenging multi-modal scenarios.

分会报告 B3 专题：神经科学和人工智能（II）

B3-1 Identification of an engram ensemble encoding memory flexibility in the dentate gyrus

钟毅, 清华大学

摘要：How the memory engram is organized at the cell-assembly level to support not only encoding of learned information but also memory flexibility remains elusive. Here, we propose a novel engram model encoded by two orthogonal learning-recruited neuronal ensembles in the mouse dentate gyrus. One ensemble encodes memory, while the other ensemble encodes forgetting. These two ensembles compete for retrieval-evoked reactivation, where altering the reactivation of one ensemble shifts the other oppositely. Such encoding enables flexibility in recall outcomes, ranging from full-scale memory expression to complete forgetting. Meanwhile, learned information remains unperturbed, as reactivation modifications specifically target the forgetting ensemble by regulating Rac1 activity, which is sensitive to cognitive and emotional events. Notably, memory phenotypes observed in mouse models of Alzheimer's disease and autism are primarily linked to dysfunctions of the forgetting ensemble, suggesting that the encoding of memory flexibility, rather than memory itself, is a major target of cognitive disorders.

B3-2 脑启发的持续学习方法与生理健康应用

王立元, 清华大学

摘要：持续学习是生物智能在“适者生存”的长期自然选择中形成的基本学习范式。然而，这种范式对于目前的人工智能来说却极其困难，在处理动态数据分布时易导致灾难性遗忘。如何深入挖掘并有效利用生物智能的持续学习机制，是实现通用人工智能的重要路径，也为理解生物智能的形成与演化提供了新视角。在本次报告中，我将介绍课题组利用生物学习记忆机制启发的持续学习理论与方法，以及持续学习模型启发的生物学习记忆机制探索与验证。此外，我还将介绍持续学习在多模态生理信号生成中的最新应用，以实现生理健康的实时监测与智能评估。

B3-3 基于脑启发的类脑决策模型

郭尚岐, 清华大学

摘要：强化学习是一种基于环境交互与奖赏反馈的智能决策范式，其机制与生物大脑在奖赏驱动下进行行为调节与目标选择高度契合。尽管该方法已在多个领域取得显著成果，但在面对高维状态、长时序依赖及资源受限的实际环境时，仍面临样本效率低、泛化性差与推理计算高耗能等核心挑战。为突破这些瓶颈，我们从人脑决策智能中汲取灵感，构建脑启发的类脑强化学习模型，融合抽象推理、分层任务规划与脉冲神经动力学等关键神经机制，以提升智能体在复杂动态环境中的决策效率、鲁棒性与能效水平。

B3-4 神经递质调控效应启发的类脑算法研究

黄子罡, 西安交通大学

摘要：(TBD)

分会报告 B4 专题：AI for Math

B4-1 High-Entropy Minority Tokens Drive Effective Reinforcement Learning for LLM Reasoning

郑楚杰, 阿里集团

摘要：（TBD）

B4-2 面向组合数学的定理自动生成和证明

支丽红, 中国科学院数学与系统科学研究院

摘要：自动化定理证明（ATP）传统上依赖于证明搜索，近年来，随着大语言模型的快速发展，AI 赋能的定理自动证明已成为一种新范式。然而，由于现有证明数据的匮乏，基于人工智能的定理自动证明仍面临诸多挑战。针对组合恒等式定理自动证明问题，我们提出了一种结合大语言模型与强化学习的定理自动生成方法；通过融合人工形式化与定理自动生成，开发了一个基于 Lean 的组合恒等式形式化数据集 LeanComb，及其相应的定理自动证明器；实验结果表明，针对组合恒等式的定理自动证明问题，我们开发的证明器在准确率上优于现有的证明器，且提出的定理自动生成方法显著提高了自动证明的效率和准确率。 共同完成人：华东师范大学：熊贝贝、单好佳、杨争峰　　河南大学：吕航宇、王建林

B4-3 Kimina-Prover: 一种推理驱动的形式化定理证明探索范式

王海明, Moonshot AI

摘要：本报告介绍 Kimina-Prover，一个为形式化定理证明设计的大语言模型。该模型的核心是一种“推理驱动的探索范式”，旨在让模型模拟人类在 Lean 4 等证明环境中解决问题的策略。Kimina-Prover 通过大规模强化学习（RL）进行训练。它采用了一种被称为形式化推理模式的结构化方法，通过迭代式地生成和优化证明步骤来构建证明。该模型在 miniF2F 基准测试中取得了 92.2% 的证明成功率。其独特的推理模式和 RL 训练也带来了高样本效率，且性能随计算资源增加而有效扩展。同时研究观察到了模型

性能随其参数规模增大而明确提升的趋势，这一现象在此前的同类工作中较为少见。模型学习到的推理风格与传统搜索算法存在差异，展现出弥合形式化验证与人类数学直觉之间差距的潜力。此外，报告还会提及后续工作对该框架的扩展，包括引入“测试时强化学习搜索”以增强模型的自主证明能力，以及赋予其解读错误信息并进行自我修正的“错误修复能力”。

B4-4 数据集的规模形式化

李嘉, Project Numina

摘要：Numina 系列的数据集已经累计突破 50 万的下载量，成为训练大模型的必备数据集。我们分享一下我们形式化 numina 数据集的一些工作以及后续工作的展望。

分会报告 B5 专题：AI for Optimization

B5-1 A space-decoupling framework for optimization on bounded-rank matrices with orthogonally invariant constraints

高斌, 中国科学院数学与系统科学研究院

摘要：Imposing additional constraints on low-rank optimization has garnered growing interest recently. However, the geometry of coupled constraints restricts the well-developed low-rank structure and makes the problem nonsmooth. In this paper, we propose a space-decoupling framework for optimization problems on bounded-rank matrices with orthogonally invariant constraints. The "space-decoupling" is reflected in several ways. Firstly, we show that the tangent cone of coupled constraints is the intersection of the tangent cones of each constraint. Secondly, we decouple the intertwined bounded-rank and orthogonally invariant constraints into two spaces, resulting in optimization on a smooth manifold. Thirdly, we claim that implementing Riemannian algorithms is painless as long as the geometry of additional constraint is known a prior. In the end, we unveil the equivalence between the original problem and the reformulated problem. The numerical experiments validate the effectiveness and efficiency of the proposed framework.

B5-2 基于图同构判定的大模型优化建模评测体系

丁添, 深圳市大数据研究院

摘要：。在工业界，数学优化问题的建模往往依赖运筹学专家的领域知识。近年来，大型语言模型（LLM）的发展让自动优化建模成为可能。然而，大模型在优化问题方面的数学建模能力的评估数据较少，并缺少理论可靠的系统性评测方法。为此，我们提出 Bench4Opt，一个用于评估大模型在线性规划（LP）与混合整数线性规划（MILP）建模能力的评测体系。新测评体系包括 818 个模型-数据分离的建模问题，覆盖 16 种问题类型与 40+应用领域。在评测方法上，首次提出使用改进的 Weisfeiler-Lehman 图同构检测算法对优化模型的等价性进行判定。实验评估显示，GPT-4o 与 DeepSeek-V3 在自动优化建模方面表现较为优异，分别达到 50.49%与 46.94%的整体准确率，但在不同种类问题上的表现差异显著，揭示了大模型在自动优化建模方面的潜力与现有局限性。

B5-3 人工智能驱动的大规模组合优化算法、平台与应用

孙建永, 西安交通大学

摘要：传统组合优化算法传统优化过度依赖研究者知识、不使用已有优化知识且通常在大规模问题上性能不好。针对这些问题，我们提出人工智能驱动的大规模优化算法，构建深度学习策略解决问题结构表征学习以实现不同组合优化问题间的泛化；提出轻重自编码器策略实现小规模问题到大规模问题的泛化；通过深度学习方法分治策略实现大规模问题的快速求解。另外，创新提出解决训练中强化学习的奖励稀疏以及大规模组合优化问题训练样本少甚至没有的问题。我们将所提出的算法集成到软件平台，并在无线通信以及地球物理勘探中取得了有效应用，取得了一定的经济和社会效益。

B5-4 LMask: Learn to Solve Constrained Routing Problems with Lazy Masking

李天佑, 北京大学

摘要：Routing problems are canonical combinatorial optimization tasks with wide-ranging applications in logistics, transportation, and supply chain management. However, solving these problems becomes significantly more challenging when complex constraints are involved. In this talk, we introduce LMask, a novel learning framework that utilizes dynamic masking to generate high-quality feasible solutions for constrained routing problems. The LazyMask decoding method is proposed to lazily refine feasibility masks with the backtracking mechanism. In addition, it employs the refinement intensity embedding to encode the search trace into the model, mitigating representation ambiguities induced by backtracking. We provide theoretical guarantees for the validity and probabilistic optimality of our approach. Extensive experiments on the TSPTW and TSPDL demonstrate that LMask achieves state-of-the-art feasibility rates and solution quality, outperforming existing neural methods.

分会报告 B6 专题：量子计算理论与方法

B6-1 Quantum for Science: Efficient Quantum Algorithms for Nonlinear Dynamics and Artificial Intelligence Models

刘锦鹏, 清华大学

摘要：Nonlinear dynamics play a prominent role in scientific computation and artificial intelligence. Whereas previous quantum algorithms for general nonlinear equations have been severely limited due to the linearity of quantum mechanics, we gave the first efficient quantum algorithm for nonlinear differential equations with sufficiently strong dissipation. This is an exponential improvement over the best previous quantum algorithms, whose complexity is exponential in the evolution time. Furthermore, we design the first quantum algorithm for training classical sparse neural networks with end-to-end settings. We benchmark instances of ResNet with sparse pruning applied to Cifar-100 dataset and DPM-Solver for U-ViT applied to ImageNet-100 dataset, and we find that a quantum enhancement is possible at the early stage of learning. Our work shows that fault-tolerant quantum computing can contribute to the training and inference processes of most state-of-the-art large language models and diffusion models. References: [1]Efficient quantum algorithm for dissipative nonlinear differential equations. Proceedings of the National Academy of Science 118, 35 (2021). [2]Towards provably

efficient quantum algorithms for large-scale machine learning models. Nature Communications 15, 434 (2024)
[3]Towards efficient quantum algorithms for diffusion probability models. arXiv:2502.14252

B6-2 Achieving Chemical Accuracy with Quantum Computing Enforced Language Model

李震宇, 中国科学技术大学

摘要：Finding accurate ground state energy of a many-body system has been a major challenge in quantum chemistry. The integration of machine learning and quantum computing has shed new light on resolving this problem. Here we demonstrate an integration of quantum computation with a transformer-based neural network that for the first time reaches chemical accuracy on a strongly correlated molecular system at the 40-qubit scale. Our hybrid algorithm, QiankunNet-QSCI, uses a quantum processor to efficiently sample crucial determinants of the electronic wavefunction, which are then refined by a deep transformer network. We achieve an accurate ground-state simulation of a challenging Fe_2S_2 molecular cluster while using far fewer configuration basis states than traditional approaches, which highlights the scalability of QiankunNet-QSCI and its potential applicability to even larger systems.

B6-3 面向隐优化视角的可约束神经网络

史良良, 上海数学与交叉学科研究院

摘要：随着深度学习技术的快速发展，越来越多的模型被用于科学发现、复杂系统的模拟以及 NP 难问题的求解等。然而，在实际场景中，这类问题的输出通常要求预测结果满足一定的约束条件，因此，如何使得模型的输出满足硬约束成为一个既有趣且重要的问题。我们聚焦于通过带约束的隐优化视角来解决这一问题，设计了基于张量迭代的优化求解算法作为神经网络层，进而确保输出满足（局部）双随机约束，流平衡约束等硬约束。

B6-4 From Parameterized Quantum Comb to Quantum Unitary Time-Reversal

王鑫, 香港科技大学（广州）

摘要：Quantum combs play a vital role in characterizing and transforming quantum processes, with wide-ranging applications in quantum information processing. However, obtaining the explicit quantum circuit for the desired quantum comb remains a challenging problem. We propose PQComb, a novel framework that employs parameterized quantum circuits (PQCs) or quantum neural networks to harness the full potential of quantum combs for diverse quantum process transformation tasks. This method is well-suited for near-term quantum devices and can be applied to various tasks in quantum machine learning. In particular, based on parameterized quantum combs, we obtain simpler circuits for reversing unknown qubit-unitary operations and obtain the idea of their generalization. We introduce a deterministic and exact approach to universally reverse arbitrary unknown unitary transformations and more efficient quantum algorithms for inverting unitaries with specific Hamiltonian structures.

分会报告 B7 专题：AI for Physics and Chemistry（I）

B7-1 AI 物理双驱动的化学反应路径搜索

朱通, 华东师范大学

摘要：地球生命的起源仍然是科学领域的一大未解之谜。尽管在前生物条件下进行的许多自下而上的实验为生命的自发化学起源提供了宝贵的见解，但对于其中复杂反应过程的理解仍存在显著的空白。在本研究中，我们提出了一种新颖的方法，使用以纯粹笛卡尔坐标形式化的旋转平移不变势（RTIP）来促进化学反应的自动化模拟。通过采用 RTIP 路径采样探索原始分子的反应性，我们识别出几种低能量的反应机制，如双氢转移氢化反应以及 HCOOH 催化的水合作用和氨基化作用。这些工作构建了一个全面的反应网络，展示了甘氨酸、丝氨酸和丙氨酸的合成路径。进一步的热力学分析突显了甲烯亚胺在氨基酸合成中的关键前体角色，因为与传统认知的氢氰酸相比，甲烯亚胺在偶联反应中的反应性更为有利。我们的研究表明，RTIP 方法结合分而治之的策略为复杂反应过程的模拟提供了新的视角，并为推进复杂体系化学反应机理研究提供了有价值的新工具。

B7-2 AI 和自动化加速功能分子和反应发现

朱戎, 北京大学

摘要：建设智能化“干湿闭环”的实验平台可能是一种可以大大加速合成化学科学发现的模式。一方面，合成化学领域已长期积累了大量文献、专利、实验、计算等数据，为通用科学模型与软件工具提供了可能；另一方面，高通量、高灵活度、高可靠性真实实验数据的获取和迭代，有望通过 few shot tuning，高水平地完成各种下游的任务，如性质预测、选择性预测、逆向设计等等。我们基于本实验室的研究基础，希望在一个示范性的项目中展示上述闭环逻辑，及其在探索未知的化学空间时，对新型富碳功能分子和相关“功能化”反应工具的发现的加速效应。

B7-3 通用深度学习密度泛函框架：DL-xDH

张颖, 复旦大学

摘要：密度泛函理论是目前使用最广泛的电子结构计算方法。寻找越来越精确的密度泛函近似方法是密度泛函领域核心课题。如何在不增加计算消耗的同时，开发可兼顾主族元素化学和过渡金属化学的通用泛函方法是理论计算化学领域长期以来的研究热点和难点问题。按照泛函构造变量的复杂程度，密度泛函近似方法可以归纳为包含 5 级台阶的雅各布天梯，从 Hartree 近似出发不断逼近“化学精度”这一泛函方法开发的天堂。但是随着泛函变量复杂度的增加，近似泛函的构造难度急剧增大，攀登天梯的难度逐级呈指数增长。这导致实际上越是高阶泛函近似，目前可用的泛函形式反而越单一。机器学习的时代背景下，密度泛函方法开发迎来了新的机遇。通过设计合适的神经网络模型，结合精确化学数据可以训练出复杂的泛函形式。但是，在低级别近似框架下引入机器学习技术无法替代更高级别密度泛函方法开发的必要性。另一方面，最高阶泛函近似需要引入未占轨道信息，构造难度远超前四阶近似。将海量数据驱动的机器学习手段应用于构造最高阶泛函近似既是密度泛函方法开发巨大的机遇，也是重大的挑战。近期，我们从高等级第五阶密度泛函方法出发，尝试从泛函误差和密度误差两方面着手，将多组态的概

念与机器学习手段引入泛函的开发，并取得一定进展。与此同时，为了适应智能时代下，多学科融合、多团队合作的理论方法研发趋势，我们充分计算机语言技术的最新进展，采用可兼顾安全性和效率的 RUST 语言，搭建电子结构计算平台，尝试在实现高效的主流电子结构计算功能的基础上，探索匹配新一代通用电子结构方法的低标度算法。

B7-4 神经网络赋能量子蒙特卡洛

任维络, ByteDance Seed

摘要：量子蒙特卡洛（QMC）方法长期以来一直是计算量子科学中的一种强大工具。近年来，神经网络与 QMC 的结合开辟了令人兴奋的新可能性。基于神经网络的 QMC 方法利用神经网络的表达能力，从而更好地表示量子态。在本次报告中，我将介绍变分蒙特卡洛（VMC）的基本原理，并介绍该领域在提升效率与精度等多方面的最新进展

分会报告 B8 专题：Optimization for AI

B8-1 Memory-Efficient Block Coordinate Descent and Backpropagation for LLM Training AI

李肖, 香港中文大学（深圳）

摘要：This talk concerns optimization techniques for memory-efficient training of large language models, mainly focusing on reducing the GPU memory cost raised by optimizer and Backpropagation (BP) process. We first present BAdam, an optimization method that leverages the block coordinate descent (BCD) framework with Adam's update rule. BAdam offers a memory-efficient approach to the full parameter finetuning of large language models. It finetunes the Llama 3-8B and Llama 3-70B models using a single RTX3090-24GB GPU and 4 A100-80GB GPUs, respectively. The experiment results confirm BAdam's efficiency in terms of memory usage, running time, optimization capability, and downstream performance. Second, we introduce StreamBP, which is a memory-efficient and exact BP algorithm for training LLMs on ultra long sequence (e.g., training reasoning model) or for scaling up batch sizes. StreamBP builds upon linear decomposition of the chain rule, and can be applied to SFT, PPO, GRPO, and DPO. It allows 3-5x larger sequence length / batch size compared to standard BP with gradient checkpointing, while using the same or even less BP time.

B8-2 Fine-Tuning Large Language Models with Forward-only Optimizers AI

沈力, 中山大学

摘要：Large language model (LLM) fine-tuning faces significant challenges in GPU-memory efficiency. To address this issue, we present advancing fine-tuning techniques for LLMs through forward-only optimization methods that minimize reliance on heavy backward passes or gradient computations. Firstly, for full fine-tuning tasks, we propose TeZO, a tensorized zeroth-order (ZO) optimizer that exploits low-rank structures across both model parameters and temporal gradients. By modeling ZO perturbations as a 3D tensor and applying Canonical Polyadic Decomposition (CPD), TeZO reduces memory consumption to 35% of prior ZO-Adam methods while maintaining SOTA-level performance. Its compatibility with adaptive optimizers further enhances scalability for

large-scale deployment. Secondly, we introduce MaskPro, a novel framework for semi-structured (N:M) sparsity in LLM mask fine-tuning. By learning a probabilistic categorical distribution over model weights and integrating variance-reduced policy gradients with a loss residual tracker, MaskPro achieves stable training while preserving hardware-friendly sparsity patterns. Both approaches can leverage gradient-free estimation to mitigate the computational bottlenecks of LLM fine-tuning. Through comprehensive theoretical analysis and experiments on diverse benchmarks, we demonstrate the power of forward optimizers in balancing efficiency, accuracy, and memory usage.

B8-3 Accelerating RLHF Training with Reward Variance Increase

袁雁城, 香港理工大学

摘要： Reinforcement learning from human feedback (RLHF) is an essential technique for ensuring that large language models (LLMs) are aligned with human values and preferences during the post-training phase. As an effective RLHF approach, group relative policy optimization (GRPO) has demonstrated success in many LLM-based applications. However, efficient GRPO-based RLHF training remains a challenge. Recent studies reveal that a higher reward variance of the initial policy model leads to faster RLHF training. Inspired by this finding, we propose a practical reward adjustment model to accelerate RLHF training by provably increasing the reward variance and preserving the relative preferences and reward expectation. Our reward adjustment method inherently poses a nonconvex optimization problem, which is NP-hard to solve in general. To overcome the computational challenges, we design a novel O(nlogn) algorithm to find a global solution of the nonconvex reward adjustment model by explicitly characterizing the extreme points of the feasible set. As an important application, we naturally integrate this reward adjustment model into the GRPO algorithm, leading to a more efficient GRPO with reward variance increase (GRPOVI) algorithm for RLHF training. As an interesting byproduct, we provide an indirect explanation for the empirical effectiveness of GRPO with rule-based reward for RLHF training, as demonstrated in DeepSeek-R1. Experiment results demonstrate that the GRPOVI algorithm can significantly improve the RLHF training efficiency compared to the original GRPO algorithm.

B8-4 A Memory Efficient Randomized Subspace Optimization Method for Traning

袁坤, 北京大学

摘要： The memory challenges associated with training Large Language Models (LLMs) have become a critical concern, particularly when using the Adam optimizer. To address this issue, numerous memory-efficient techniques have been proposed, with GaLore standing out as a notable example designed to reduce the memory footprint of optimizer states. However, these approaches do not alleviate the memory burden imposed by activations, rendering them unsuitable for scenarios involving long context sequences or large mini-batches. Moreover, their convergence properties are still not well-understood in the literature. In this work, we introduce a Randomized Subspace Optimization framework for pre-training and fine-tuning LLMs. Our approach decomposes the high-dimensional training problem into a series of lower-dimensional subproblems. At each iteration, a random subspace is selected, and the parameters within that subspace are optimized. This structured reduction in dimensionality allows our method to simultaneously reduce memory usage for both activations and optimizer states. We establish comprehensive convergence guarantees and derive rates for various scenarios, accommodating different optimization strategies to solve the subproblems. Extensive experiments validate the

superior memory and communication efficiency of our method, achieving performance comparable to GaLore and Adam.

分会报告 C1 专题：强化学习理论与算法（I）

C1-1 Offline learning for combinatorial optimization

陈卫, 微软亚洲研究院

摘要：Traditionally machine learning and optimization are two different branches in computer science. They need to accomplish two different types of tasks, and they are studied by two different sets of domain experts. Machine learning is the task of extracting a model from the data, while optimization is to find the optimal solutions from the learned model. In the current era of big data and AI, however, such separation may hurt the end-to-end performance from data to optimization in unexpected ways. Data-driven optimization is an effective way to tightly integrate data sampling, machine learning and optimization tasks. In this talk, I will focus on one important approach in data-driven optimization, which is on how to learn from offline sampled data with the goal of combinatorial optimization. I will briefly introduce my recent studies on optimization from structured samples and offline learning for combinatorial multi-armed bandits to effectively tackle the problem.

C1-2 uniINF: Best-of-Both-Worlds Algorithm for Parameter-Free Heavy-Tailed MABs

黄隆波, 清华大学

摘要：In this work, we present a novel algorithm, uniINF, for the Heavy-Tailed Multi-Armed Bandits (HTMAB) problem, demonstrating robustness and adaptability in both stochastic and adversarial environments. Unlike the stochastic MAB setting where loss distributions are stationary with time, our study extends to the adversarial setup, where losses are generated from heavy-tailed distributions that depend on both arms and time. Our novel algorithm `uniINF` enjoys the so-called Best-of-Both-Worlds (BoBW) property, performing optimally in both stochastic and adversarial environments without knowing the exact environment type. Moreover, our algorithm also possesses a Parameter-Free feature, i.e., it operates without the need of knowing the heavy-tail parameters (σ,α) a-priori. To be precise, uniINF ensures nearly-optimal regret in both stochastic and adversarial environments, matching the corresponding lower bounds when (σ,α) is known (up to logarithmic factors). To our knowledge, uniINF is the first parameter-free algorithm to achieve the BoBW property for the heavy-tailed MAB problem. Technically, we develop innovative techniques to achieve BoBW guarantees for Parameter-Free HTMABs, including a refined analysis for the dynamics of log-barrier, an auto-balancing learning rate scheduling scheme, an adaptive skipping-clipping loss tuning technique, and a stopping-time analysis for logarithmic regret.

C1-3 大模型背景下的强化学习

俞扬, 南京大学

摘要：2024 年图灵奖授予研究强化学习的先驱。强化学习已从早期游戏任务扩展到机器人控制等复杂

物理环境中的应用。本次报告将回顾强化学习技术发展历史，并汇报在大模型受到高度关注的背景下，强化学习技术的发展与变化。

C1-4 Settling the Sample Complexity of Online Reinforcement Learning

陈昱鑫, University of Pennsylvania

摘要： A central issue lying at the heart of online reinforcement learning (RL) is data efficiency. While a number of recent works achieved asymptotically minimal regret in online RL, the optimality of these results is only guaranteed in a ``large-sample'' regime, imposing enormous burn-in cost in order for their algorithms to operate optimally. How to achieve minimax-optimal regret without incurring any burn-in cost has been an open problem in RL theory. We settle this problem for the context of finite-horizon inhomogeneous Markov decision processes. Specifically, we prove that a modified version of Monotonic Value Propagation (MVP) achieves the minimal optimal regret for the entire range of sample size, essentially eliminating any burn-in requirement.

分会报告 C2 专题：机器学习和逼近理论（I）

C2-1 Learning performance of Off-line Q-learning algorithms

林绍波, 西安交通大学

摘要： With the help of massive data and rich computational resource, offline Q-learning has been widely used in operations research and management science and receives great success in numerous applications including, recommender system, games and robotic manipulation. Compared with avid research activities in practice, there lack solid theoretical verifications and interpretability for the success of offline Q-learning, making it be a little bit mystery. The aim of this talk is to discuss the generalization performance of two modern offline Q-learning strategies: deep Q-learning and distributed Q-learning. In the framework of learning theory, we rigorously prove that these two Q-learning approaches outperform the traditional one by showing its good generalization error bound. In particular, our results show that the main reason of the success of deep Q-learning is due to the excellent performance of deep neural networks (deep nets) in capturing special properties of rewards such as the spatially sparse and piecewise constant rather than due to their large capacities. We also show that distributed Q-learning succeeds in reducing the computational burden without sacrificing the generalization performance

C2-2 Learning Theory of Classification with Deep Neural Networks

石磊, 复旦大学

摘要： Deep neural networks have achieved remarkable success in various binary classification tasks. Despite their practical effectiveness, theoretical understanding of their generalization in binary classification remains limited. In this talk, I will present our recent progress on classification using deep neural networks.

C2-3 随机特征模型与两层神经网络分析的对偶框架

龙吉昊, 上海算法创新研究院

摘要： We consider the problem of learning functions in the Fp,π and Barron spaces, which are relevant for

understanding random feature models (RFMs), two-layer neural networks, as well as kernel methods. Through a duality analysis, we reveal an equivalence between the approximation and estimations for learning functions in the two spaces. This enables us to focus on the easier one among approximation and estimation when examining the learnability of these function spaces. To demonstrate the flexibility and versatility of our duality framework, we provide comprehensive analyses of two applications. 1) The first application is to study learning functions in $F_{p,\pi}$ with RFMs. We prove that RFM can learn functions in $F_{p,\pi}$ without the curse of dimensionality as long as $p > 1$. This result implies that RFMs can work well beyond the kernel regime as the $F_{p,\pi}$ is strictly larger than the associated reproducing kernel Hilbert space (RKHS) when $p < 2$. 2) The second application is to investigate the learnability of reproducing kernel Hilbert space (RKHS) under the L^∞ norm. By leveraging the duality principle, we relate the L^∞ learnability of a RKHS to the eigenvalue decay of the associated kernel, thereby establishing both lower and upper bounds of sample complexity. We then apply these bounds to dot-product kernels and identify conditions when the learning suffers or overcomes the curse of dimensionality. In particular, these results imply that learning with (random) ReLU features is generally intractable under the L^∞ norm. To establish the dual equivalence, we introduce an information-based complexity. We show that this complexity can effectively control minimax estimation errors in various settings, which might be of independent interest.

C2-4 Operator Learning and Neural Scaling Laws

刘皓, 香港浸会大学

摘要：Deep neural networks have demonstrated a great success in many applications. For operator learning and large language model, neural scaling laws are observed in many works. Most of the observed laws are power laws, i.e., the testing error can be written as a power of number of parameters or the number of training samples. However, theoretical explanations of the scaling laws are largely missing. In this presentation, we focus on operator learning and analyze the approximation and generalization error of some popular network architectures. We provide a theoretical explanation of neural scaling laws, and show that if the data has low-dimensional structures, one can achieve power laws.

分会报告 C3 专题：生成模型算法

C3-1 高效多模态生成：方法与应用

邓志杰, 上海交通大学

摘要：以（视觉）语言模型、扩散模型为代表的多模态生成模型是当前人工智能领域的前沿热点，然而现有模型在效率方面面临严峻挑战。本报告将系统性介绍多模态生成模型的高效建模、训推方法，并简要讨论其在图文交错生成、文生视频、长 COT 推理等场景的应用。

C3-2 基于 LLM 的合成数据有效吗？

刘勇, 中国人民大学

摘要：在大语言模型（LLMs）后训练任务中，由于高质量的特定领域数据十分稀缺，合成数据已成为

重要资源。虽然已有多种方法被用于生成合成数据，但关于合成数据的理论分析仍相对缺乏。本报告首先对当前流行的合成数据生成过程进行数学建模，然后从一个新的反信息瓶颈视角对数据合成进行了理论分析，阐明后训练模型的泛化能力关键取决于生成模型带来的信息增益。希望为合成数据生成技术的设计与后训练过程的优化提供新的理解。

C3-3 Diffusion vs. Autoregression: Which is the Key to Next-Generation LLMs

贺笛, 北京大学

摘要：（TBD）

C3-4 LLaDA：大语言模型新范式

李崇轩, 中国人民大学

摘要：本次报告聚焦一个问题：自回归是否是通向当前乃至更高水平的生成式智能的唯一范式？本次报告首先从统一概率建模的视角总结当前基础生成模型的发展，并从这个视角出发指出大语言模型的性质（如可扩展性、指令追随、情景学习、对话、无损压缩）主要来自于生成式准则，而非自回归建模独有。基于这些洞察，介绍扩散大语言模型 LLaDA 系列工作，包括基础理论、扩展定律、大规模训练、偏好对齐和多模态理解等。LLaDA 通过非自回归的方式，展示了令人惊讶的可扩展性和多轮对话能力。这些结果不仅挑战了自回归模型的统治地位，更加深了我们对生成式人工智能的理解。

分会报告 C4 专题：大模型系统

C4-1 大语言模型在异构算力环境中的部署

袁彬航, 香港科技大学

摘要：（TBD）

C4-2 从同构走向分离的大模型推理系统

章明星, 清华大学

摘要：（TBD）由于在算力和带宽两方面的明显优势，传统大模型推理架构往往以 GPU 为中心进行设计。然而，随着 GPU 利用率逐渐逼近瓶颈，进一步降低推理成本需要开拓新的优化路径。结合不同 GPU 设备乃至 CPU/DRAM 设备在带宽或容量成本上的优势，并充分利用模型本身的前序依赖性与稀疏性特征，设计适配的计算架构成为未来算法、系统与硬件协同创新的重要方向。本次报告将介绍两种具体优化思路：一是以存换算的 Mooncake 架构，它通过 以 KVCache 为中心的大模型推理架构大幅提升了 Kimi 线上业务的承载能力；在此基础上，我们进一步探讨了更多的异构分离可能性。例如，在 P/D 分离的基础上，我们发现 Decode 环节中的 MLP 和 Attention 算子具有进一步分离的潜力。为此，我们优化了相关的网络传输链路以降低延迟，并在字节跳动的环境中进行了测试。二是以存强算的 KTransformers 系统，它针对 DeepSeek V3/R1 类稀疏大模型进行 CPU/GPU 异构推理优化，显著降低了本地部署门槛。相关项目均已开源，并获得社区广泛关注。

C4-3 PAI-Llumnix：动态、弹性、可扩展的分布式推理

赵汉宇, 阿里巴巴集团

摘要：（TBD）

C4-4 复杂、动态负载下的分布式大模型训练=

符芳诚, 上海交通大学

摘要：（TBD）

分会报告 C5 专题：机器学习与材料

C5-1 Modeling Randomness Effects in High-Entropy Alloys

张露婵, 深圳大学

摘要：High-entropy alloys (HEAs), i.e., single-phase, (nearly) equiatomic multicomponent, metallic materials, have novel mechanical properties (high strength etc). We propose a stochastic Peierls-Nabarro model to understand how random site occupancy affects intrinsic strength. The stochastic Peierls-Nabarro model accounts for the randomness in the composition, characterized by both the standard deviation of the perturbation in the interplanar potential and the correlation length within the spatial compositional distribution. The model predicts the intrinsic strength of HEAs as a function of standard deviation and correlation length of the randomness. We find that compositional randomness induces an intrinsic strength. This approach provides a fundamental explanation to the origin of high strength of HEAs. We also derive stochastic continuum models for HEAs from atomistic models that incorporate the atomic level randomness and the short-range order. These stochastic continuum models theoretically validate the randomness incorporation in our stochastic Peierls-Nabarro model.

C5-2 A Stabilized Physics Informed Neural Networks Method for Wave Equations

袁成, 华中师范大学

摘要：In this work, we propose a novel Stabilized Physics Informed Neural Networks method (SPINNs) for solving wave equations. In general, this method not only demonstrates theoretical convergence but also exhibits higher efficiency compared to the original PINNs. By replacing the ξLˆ2ξ norm with ξHˆ1ξ norm in the learning of initial condition and boundary condition, we theoretically proved that the error of solution can be upper bounded by the risk in SPINNs. Based on this, we decompose the error of SPINNs into approximation error, statistical error and optimization error. Furthermore, by applying the approximating theory of ξReLUˆ3ξ networks and the learning theory on Rademacher complexity, covering number and pseudo-dimension of neural networks, we present a systematical non-asymptotic convergence analysis on our method, which shows that the error of SPINNs can be well controlled if the number of training samples, depth and width of the deep neural networks have been appropriately chosen. Two illustrative numerical examples on 1-dimensional and 2-dimensional wave equations demonstrate that SPINNs can achieve a faster and better convergence than classical PINNs method.

C5-3 Data-driven approaches for numerical PDEs: reduced order modeling & operator learning

干则成, 香港科技大学（广州）

摘要： We report some recent progress in developing data-driven approaches for numerical PDEs. In the first part, we will discuss a reduced order method (ROM) for solving the close-to-touch interaction between two nanoparticles, where tailored ROM scheme is developed to overcome the near-singular nature of such problem, significantly reducing the number of basis functions. In the second part, we will discuss an improved neural network structure for operator learning tasks based on long-range & short-range convolutions; where a new sum-of-exponentials ansatz is proposed in the long-range convolution module, significantly reducing the training cost, meanwhile improving the generalization abilities of the neural network structure.

C5-4 Frequency-adaptive Multi-scale Deep Neural Networks

黄记祖, 中国科学院数学与系统科学研究院

摘要： Multi-scale deep neural networks (MscaleDNNs) with downing-scaling mapping have demonstrated superiority over traditional DNNs in approximating target functions characterized by high frequency features. However, the performance of MscaleDNNs heavily depends on the parameters in the downing-scaling mapping, which limits their broader application. In this work, we establish a fitting error bound to explain why MscaleDNNs are advantageous for approximating high frequency functions. Building on this insight, we construct a hybrid feature embedding to enhance the accuracy and robustness of the downing-scaling mapping. To reduce the dependency of MscaleDNNs on parameters in the downing-scaling mapping, we propose frequency-adaptive MscaleDNNs, which adaptively adjust these parameters based on a posterior error estimate that captures the frequency information of the fitted functions. Numerical examples, including wave propagation and the propagation of a localized solution of the schrξ\ddot{\text{o}} ξdinger equation with a smooth potential near the semi-classical limit, are presented. These examples demonstrate that the frequency-adaptive MscaleDNNs improve accuracy by two to three orders of magnitude compared to standard MscaleDNNs.

C5-5 A Generative Model for Composition Engineering in Multi-Principal Element Alloys

项阳, 香港科技大学

摘要： Multi-principal element alloys (MPEAs), characterized by the presence of multiple primary elements in near-equiatomic or relatively high concentrations, exhibit distinctive mechanical properties arising from their complex compositional landscapes. This inherent compositional complexity leads to significant challenges in elucidating their deformation and fracture mechanisms for the efficient design. We introduce AlloyVAE, a generative machine learning framework to directly predict MPEA residual stress from composition fields (concentrations and short-range orders). This framework also enables the solution of inverse design problems for strength enhancement by composition engineering.

分会报告 C6 专题：图像处理与人工智能

C6-1 面向肿瘤疗效预测的多模态分析方法

张立, 北京大学

摘要：面向患者治疗疗效的有效预测，其疾病的表征需要融合来自异构数据源（如影像、病理和临床数据）的信息。传统方法如简单的数据拼接，往往无法捕捉这些模态间复杂的非线性相互依赖关系，因而难以形成完整的患者画像。我们提出了一个以交叉注意力机制为核心的多模态融合框架。该方法使模型能够动态地学习并权衡不同数据模态间的深层关联。实验结果表明，模型融合表征所产生的风险分层，与关键的观测终点表现出显著的相关性。这种从多源数据中创建稳健、一体化患者画像的高效策略，这为复杂生物医学系统中的定量建模展示了重要价值。

C6-2 Enhancing Full Waveform Inversion via Learned and Regularized Source Wavelet Manipulation

邱凌云, 清华大学

摘要：Full-waveform inversion (FWI) is a powerful tool for high-resolution subsurface parameter reconstruction. Due to the existence of local minimum traps, the success of the inversion process usually requires a good initial model. Our study primarily focuses on understanding the impact of source wavelets on the landscape of the corresponding optimization problem. We thus introduce a decomposition scheme that divides the inverse problem into two parts. The first step transforms the measured data into data associated with the desired source wavelet. Here, we consider inversions with known and unknown sources to mimic real scenarios. The second sub-problem is the conventional full waveform inversion, which is much less dependent on an accurate initial model since the previous step improves the misfit landscape. A regularized deconvolution method and a convolutional neural network are employed to solve the source transformation problem. Numerical experiments on the benchmark models demonstrate that our approach improves the gradient's quality in the subsequent FWI and provides a better inversion performance.

C6-3 Parametric Neural Operator for Non-Line-of-Sight Imaging

段玉萍, 北京师范大学

摘要：Non-line-of-sight (NLOS) imaging is an advanced computational imaging technology aimed at reconstructing obscured or hidden scenes using indirect light signals. These signals are typically generated through multiple reflections or scattering, resulting in weak signal strength and susceptibility to noise interference. Therefore, incorporating physical processes into the reconstruction is crucial for enhancing the quality. We propose a parametric neural operator model capable of learning complex mapping relationships. Through training, this model can simulate the propagation of light and extract useful information from indirect light signals. By leveraging the powerful fitting capabilities of neural networks, this approach can handle complex light transmission models and effectively reduce noise.

分会报告 C7 专题：算子学习

C7-1 FEALPy: A Cross-Platform Intelligent CAX Engine with Scalable Tensor Computation for Multi-Method Simulations

魏华祎, 湘潭大学

摘要： FEALPy is a cross-platform, intelligent CAX engine designed to advance multi-method simulations through scalable tensor computation. While originally rooted in finite element algorithms, FEALPy now supports a wide array of numerical methods including finite difference, finite volume, particle methods, and more. The platform's core lies in its unified mesh interface, enabling seamless transitions between different mesh types and dimensions without the need to modify the underlying code. FEALPy integrates machine learning algorithms, combining traditional CAX methods with AI to accelerate the development of next-generation intelligent CAX applications. With its multi-backend tensor computation engine, supporting libraries such as Numpy, PyTorch, and JAX, FEALPy is adaptable to modern heterogeneous hardware systems. Faithful to its mission, FEALPy aims to provide reliable, robust support for researchers and engineers, promoting innovation in CAX methods and paving the way for cutting-edge industrial applications. This presentation presents FEALPy's architecture, key technologies, and diverse application scenarios, positioning it as a steadfast companion in the field of intelligent CAX.

C7-2 Solving PDEs using deep neural networks with error control

毛志平, 宁波东方理工大学

摘要： Neural networks have shown significant potential in solving partial differential equations (PDEs). While deep networks are capable of approximating complex functions, direct one-shot training often faces limitations in both accuracy and computational efficiency. To address these challenges, we propose both Galerkin and collocation adaptive methods that uses neural networks to construct basis functions guided by the equation residual. The approximate solution is computed within the space spanned by these basis functions. As the approximation space gradually expands, the solution is iteratively refined; meanwhile, the progressive improvements serve as reliable a posteriori error indicators that guide the termination of the sequential updates. Additionally, we introduce adaptive strategies for collocation point selection and parameter initialization to enhance robustness and improve the expressiveness of the neural networks. We also derive the approximation error estimate and validate the proposed method with several numerical experiments on various challenging PDEs, demonstrating both high accuracy and robustness of the proposed methods.

C7-3 Multigrid Neural Operator and Preconditioner: Operator Learning and Fast Helmholtz Solver

刘新亮, 阿卜杜拉国王科技大学

摘要： We present Multigrid Neural Operator (MgNO) and a Multigrid Neural Preconditioner, a unified framework that integrates classical multigrid methodologies with modern deep learning to tackle complex PDEs and accelerate linear solves for Helmholtz in particular. First, we introduce the Finite Neuron Method (FNM)—a linearized ξ\mathrm{ReLU}ˆkξ network with fixed, quasi-uniform weights—and prove that it attains optimal Sobolev-space approximation rates without suffering the curse of dimensionality, thereby demonstrating the

viability of linearized networks for high-dimensional operator learning. Next, to overcome spectral bias in operator learning, we build multigrid neural operator (MgNO) on the MgNet architecture, parameterizing multigrid V-cycle components (smoothers, intergrid transfers, coarse-grid corrections) as trainable operators. Finally, we design an unsupervised, multigrid-style preconditioner for challenging systems like oscillatory Helmholtz equations, training both smoothing and coarsening maps using only PDE residuals and coefficient fields. Coupling this learned smoother, the multigrid multi-channle neural preconditioner yields substantial convergence speedups over classical alternatives.

C7-4 A deformation-based framework for learning solution mappings of PDEs defined on varying domains

金鹏展, 北京大学

摘要：In this work, we establish a deformation-based framework for learning solution mappings of PDEs defined on varying domains. The union of functions defined on varying domains can be identified as a metric space according to the deformation, then the solution mapping is regarded as a continuous metric-to-metric mapping, and subsequently can be represented by another continuous metric-to-Banach mapping using two different strategies, referred to as the D2D framework and the D2E framework, respectively. We point out that such a metric-to-Banach mapping can be learned by neural networks, hence the solution mapping is accordingly learned. With this framework, a rigorous convergence analysis is built for the problem of learning solution mappings of PDEs on varying domains. As the theoretical framework holds based on several pivotal assumptions which need to be verified for a given specific problem, we study the star domains as a typical example, and other situations could be similarly verified. We finally present several numerical experiments to validate our theoretical results.

分会报告 C8 专题：AI for Physics and Chemistry (II)

C8-1 人工智能赋能的燃烧反应动力学模型发展

杨斌, 清华大学

摘要：介绍当前燃烧反应反动学模型发展过程中的人工智能方法。包括模型分析、模型优化、实验设计等。

C8-2 复杂流动与燃烧过程的数据驱动降阶代理模型研究

王兴建, 清华大学

摘要：面向先进动力系统中的复杂流动与燃烧过程，传统实验研究方法成本高昂，数值仿真计算资源消耗大，而基于数据驱动的代理模型技术为高效预测多参数耦合条件下的物理场分布提供了新的解决方案。本报告针对流动与燃烧物理场特有强非线性和高维度特征，系统性地构建了基于流形学习的降阶模型框架，通过特征空间映射实现了复杂的低维表征。重点对比分析了本征正交分解（POD）、自编码器（AE）与张量分解等三类降阶建模方法的计算精度、泛化能力与工程适用性，为复杂燃烧系统的快速预测与优化设计提供了新的技术途径。

C8-3 结合领域结构化知识的流体数值仿真智能体方法

张天汉, 北京航空航天大学

摘要：计算流体力学（CFD）在科学与工程领域的进步中具有关键作用，但其复杂的操作流程和对广泛领域专业知识的需求却成为其发展的阻碍。本文提出了 ChatCFD，这是一种基于大型语言模型（LLM）的流程化工具，能够在 OpenFOAM 框架内实现 CFD 工作的自动化，使用户可以通过自然语言提示或发布的文献，在几乎不需要专业知识的情况下配置并执行复杂的模拟。其核心创新在于其在数据库构建、配置验证和错误反馈中的结构化思维能力，该能力将 CFD 和 OpenFOAM 领域的专业知识（如求解器、湍流模型、文件依赖关系）系统性地与通用 LLM 相结合，从而提高了准确性和适应性。具体而言，ChatCFD 采用了四阶段工作流程：（1）知识库构建，从 OpenFOAM 教程和手册中创建结构化的 JSON 数据库；（2）用户输入处理，通过多模态界面（对话、文档、网格文件）指导用户；（3）案例文件初始化，利用预处理的知识库生成案例文件；（4）模拟执行与错误修正，通过检索增强生成（RAG）运行模拟并解决错误。借助\texttt{DeepSeek-R1｝和\texttt{DeepSeek-V3｝模型、多代理架构以及专门的 OpenFOAM 知识，ChatCFD 具备交互式对话界面、分层参数提取功能以及强大的错误修正系统，能够解决维度不匹配、文件缺失、持续错误和一般性问题等。验证结果表明，ChatCFD 能够在无需人工干预的情况下自动再现已发表的 CFD 文献中的结果，这是一项涉及复杂、未见过的配置的挑战性任务，超越了基础 CFD 示例和一般 LLM 的能力。这一成就突显了其领域特定的结构化思维能力带来的显著进步。此外，验证实验表明，在 30-40%的不可压缩和可压缩 CFD 案例中，ChatCFD 实现了无错误配置，并在 60-80%的案例中成功运行，确立了自动化 CFD 的基准。通过降低专业知识的门槛，ChatCFD 提升了 CFD 的可及性，并为人工智能驱动的工程模拟提供了方法论上的启示。

C8-4 基于混合机器学习架构的复杂流场长期高保真预测方法研究

王柏森, 北京航空航天大学

摘要：准确预测高维非线性流体动力系统的长期演化行为对工业设计与科学计算具有重要意义，然而传统数值方法面临计算成本过高与误差累积的挑战，现有数据驱动模型则受限于长期预测的稳定性缺失与物理一致性不足。本研究提出一种融合状态空间建模与生成修正的混合机器学习框架。通过对流场时间序列快照进行本征正交分解，提取主导模态构建低维子空间，并采用状态空间模型建模该子空间动力学。针对本征正交截断导致的高阶模态缺失问题，提出了正交补空间生成修正策略，进一步为确保生成结果符合物理规律，引入微分方程约束嵌入机制，计算控制方程残差，通过扩散后验采样理论，使模型收敛至物理可行解域。通过后台阶流瞬态流场的验证表明，本框架可稳定预测全阶流场演化。相较于传统方法，所提出架构在长期稳定性、复杂几何细节还原及物理场一致性方面展现显著优势。

分会报告 D1 专题：强化学习理论与算法（II）

D1-1 欺骗性对齐机理与方法

杨耀东, 北京大学

摘要：（TBD）

D1-2 面向大模型智能体的强化学习

温颖, 上海交通大学

摘要：大模型的能力提升依赖于持续获取高质量的数据和反馈信号。虽然预训练阶段已利用大量优质数据，但持续增长的关键在于不断引入新的高质量数据。由于人工数据生产成本高且难以满足需求，探索大模型自我迭代生成和筛选数据的方法变得至关重要。本讲座将探讨大模型的数据再生产过程，包括生成、评估和训练三个步骤，核心挑战在于设计针对大语言模型的任务环境、奖励信号及算法，以实现数据的有效筛选和评估，通过应用不同级别的反馈信号进行强化学习，确保只有最有价值的数据用于模型的迭代训练，激发大语言模型的认知与元认知能力，以提升大语言模型智能体的泛化能力和决策任务性能。

D1-3 A Normalizing Flows-based Deep Reinforcement Learning Algorithm for Mean-Field Games

陈志平, 西安交通大学

摘要：We propose an algorithm that combines deep reinforcement learning and fictitious play to solve infinite-horizon, discounted and entropy-regularized mean-field games (MFGs) with population-dependent dynamics in continuous state-action spaces. We design a dual normalizing flows architecture to represent both current and average population distributions, enabling efficient sampling and accurate density estimation. We prove the convergence of our algorithm for entropy-regularized MFGs in continuous state and action spaces, establishing an ξO(\frac{1}{t})ξ decay rate of exploitability. Numerical experiments on three kinds of MFGs, especially high-dimensional MFGs demonstrate that our algorithm can achieve faster exploitability convergence and better performance than state-of-the-art baselines.

D1-4 多智能体强化学习与 AI Agent 研究

杨天培, 南京大学

摘要：强化学习 (RL) 在在机器人控制、游戏等序列决策任务中展现出令人瞩目的成果。与此同时，大型语言模型 (LLM) 和视觉语言模型 (VLM) 应运而生，在多模态理解和推理方面展现出令人印象深刻的能力。本报告将介绍团队在大模型、强化学习和多智能体系统交叉领域的最新工作介绍，包括多智能体通信泛化工作、Computer Use Agent，以及多 LLM Agent 在医疗领域的应用。最后，我们探讨一些开放性问题，包括多 LLM Agent 的端侧部署和交互通信等。

分会报告 D2 专题：机器学习和逼近理论（II）

D2-1 从有限元到机器学习

谢和虎, 中国科学院数学与系统科学研究院

摘要：本报告从有限元方法求解偏微分方程的角度来理解和设计基于神经网络的机器学习算法。首先从有限元方法的角度来理解机器学习算法求解偏微分方程的误差构成，得到数值积分精度也是影响机器学习精度的重要因素，同时在积分精度充分的条件下机器学习算法具有自适应的特点，即具有自动处理问题奇性的能力，非常类似于有限元中的自适应算法。在进行数值积分思想的指导下，我们也设计了可以高精度求解高维问题的张量神经网络结构，并用于设计可高精度求解高维问题的机器学习算法。

D2-2 Neural Networks, Dynamical Systems, Control Families, and Formal Languages

蔡永强, 北京师范大学

摘要：Deep learning has made significant progress in data science and natural science. Some studies have linked deep neural networks to dynamical systems, but the network structure is restricted to residual networks. It is known that residual networks can be regarded as numerical discretizations of dynamical systems. In this talk, we consider traditional network structures and prove that vanilla feedforward networks can also be used for the numerical discretization of dynamical systems, where the width of the network is equal to the input and output dimensions. The proof is based on the properties of the leaky ReLU function and the numerical technique of the splitting method for solving differential equations. The results could provide a new perspective for understanding the approximation properties of feedforward neural networks. In particular, the minimum width of neural networks and the minimal control family of dynamical systems for universal approximation can be derived. In addition, the relationship between mapping compositions and regular languages can be established.

D2-3 Norm spaces rooted in neural networks and their applications

陆帅, 复旦大学

摘要：We revisit several neural network-derived norm spaces, encompassing (extended) Barron spaces, variation spaces, Radon-BV spaces, and spectral Barron spaces. We systematically investigate the properties of these spaces and explore their applications within regularization schemes and inverse problems in partial differential equations. It is joint work with Yuanyuan Li (Fudan), Peter Mathé (WIAS)

D2-4 Learning theory of spectral algorithms under covariate shift

郭正初, 浙江大学

摘要：In machine learning, it is commonly assumed that the training and test samples are drawn from the same underlying distribution. However, this assumption may not always hold true in practice. In this talk, we delve into a scenario where the distribution of the input variables (also known as covariates), differs between the training and test phases. This situation is referred to as covariate shift. To address the challenges posed by covariate shift, various techniques have been developed, such as importance weighting, domain adaptation, and reweighting methods. In this talk, we specifically focus on the weighted spectral algorithm. Under mild conditions imposed on the weights, we demonstrate that this algorithm achieves satisfactory convergence rates. This talk is based on joint work with Prof. Jun Fan and Prof. Lei Shi.

D2-5 The Condensation Phenomenon of Deep Learning

张耀宇, 上海交通大学

摘要：Condensation (also known as quantization, clustering, or alignment) is a widely observed phenomenon where neurons in the same layer tend to align with one another during the nonlinear training of deep neural networks (DNNs). It is a key characteristic of the feature learning process of neural networks. In recent years, to advance the mathematical understanding of condensation, we uncover structures regarding the dynamical regime, loss landscape and generalization for deep neural networks, based on which a novel theoretical framework emerges. This presentation will cover these findings in detail. First, I will present results regarding the dynamical regime identification of condensation at the infinite width limit, where small initialization is crucial. Then, I will discuss the mechanism of condensation at the initial training stage and the global loss landscape structure underlying condensation in later training stages, highlighting the prevalence of condensed critical points and global minimizers. Finally, I will present results on the quantification of condensation and its generalization advantage, which includes a novel estimate of sample complexity in the best-possible scenario. These results underscore the effectiveness of the phenomenological approach to understanding DNNs, paving a way for further developing deep learning theory.

分会报告 D3 专题：机器学习与统计

D3-1 Approximation error from discretizations and its applications

赵俊龙, 北京师范大学

摘要：Converting a continuous variable into a discrete one is a commonly used technique for solving various problems in both statistics and machine learn- ing. It is well known that discretizations result in biases. However, this issue has not been studied systematically. In this paper, a general framework is proposed to understand and compare the approximation errors of different s- licing strategies. Poincare-type inequalities are first established for univariate discretizations and then generalized to the multivariate and other settings. It is shown that the bias is controlled by two factors: the distance between t- wo specific distributions that are generated with and without discretizations respectively, and the smoothness of the functions involved. Several importan- t applications are considered to illustrate the usefulness of the results. Our results help to understand the approximation error of some matrix used in the literature of dimension reduction. Furthermore, as an illustration of the usefulness of discretizations, we propose an algorithm for regression prob- lems, by combining random forest with partial discretizations of responses. Simulation results confirm the advantages of this algorithm over the classical random forest.

D3-2 Connections between context data and model weights in transformers

胡天阳, 香港中文大学（深圳）

摘要：Context data—such as few-shot examples or chain-of thoughts—has become a key ingredient in how large language models (LLMs) learn and adapt on the fly. This talk explores the connections between such context data and the model weights in transformer-based LLMs. First, we connect context data to training data

through a greedy layer-wise gradient descent algorithm. Then, we examine whether the effects of context can be directly internalized into model weights. While standard transformer architectures fall short in this regard, we find that small architectural modifications—like adding query-dependent bias terms—can bridge this gap. These insights shed light on how LLMs use context and suggest new ways to make them more adaptive and efficient.

D3-3 Uniform Inference for Kernel Gradient Flow Regression

程宇骞, 清华大学

摘要：深度生成模型在提升性能的同时，通常也伴随着巨大的计算开销，这在训练和推理阶段都构成了实际的挑战。如何提高生成模型的效率和性能，是当前领域关注的重点。本次报告将介绍两种新的技术路径。首先是 GMem，一种模块化的生成模型方法。它通过解耦记忆与泛化，将关键的语义信息存储于一个独立的外部模块中，从而降低了模型对网络自身规模的依赖。实验结果表明该方法能显著提升训练效率。 其次，我们将讨论一个统一的连续生成模型框架 (UCGM)。该框架旨在整合不同的采样方法，如多步扩散和少步一致性模型，提供统一的训练和采样流程。应用该框架，不仅可以训练出在极少步数（例如 2 步）内就达到高质量的模型，也能用于优化现有的预训练模型，在大幅减少采样步数（减少 84%）的情况下获得性能提升。

D3-4 Over-parameterization Leads to Adaptivity in High Dimensional Gaussian Sequence

丁嘉麟, 北京大学

摘要：Recently, an adaptive feature program has advocated that over-parameterized models lead to more adaptivity in regression. In this paper, we address the adaptivity of the diagonal over-parameterization in inner product kernel regression on the high dimensional sphere ξS^{d−1}ξ, where the sample size ξn \asymp d^\gammaξ for some ξ\gamma>0ξ. Motivated by the celebrated Le Cam equivalence, we first propose an alternative simplified sequence model, which captures the essential behavior of inner product kernel regression on the high dimensional sphere. We then show that the over-parameterized sequence model in high dimensional settings achieves better convergence rates than fixed kernel regression, and actually matches the minimax rate over the specified subset. Moreover, we also demonstrate that depth enhances generalization, i.e., introducing an extra D-layer parameterization improves the generalization error rate, even approaching the parametric rate as D increases in some scenarios.

分会报告 D4 专题：机器学习与优化理论

D4-1 Progress and open problems in structured optimization

张景昭, 清华大学

摘要：Oracle-complexity analysis has become a powerful tool for understanding the fundamental limits of optimization, yielding tight upper and lower bounds for a broad class of standard minimization problems, whether convex or nonconvex. Building on this mature foundation, we turn to richer formulations such as min–max and bilevel optimization, which better capture modern machine-learning objectives yet remain comparatively under explored. For these settings we present several algorithmic refinements that improve

known convergence rates, while simultaneously exposing significant gaps between the best available upper bounds and existing—or currently unknown—lower bounds, thereby charting clear directions for future research.

D4-2 Accelerated Gradient Descent by Concatenation of Stepsize Schedules

江如俊, 复旦大学

摘要：This talk considers stepsize schedules for gradient descent on smooth convex objectives. We extend the existing literature and propose a unified technique for constructing stepsizes with analytic bounds for arbitrary iterations. This technique constructs new stepsize schedules by concatenating two short stepsize schedules. Using this approach, we introduce two new families of stepsize schedules, achieving a convergence rate of ξO(nˆ{-1.2716...}〉)ξ with a state-of-the-art constants for the objective value and gradient norm of the last iterate, respectively. Furthermore, our analytically derived stepsize schedules either match or surpass the existing best numerically computed stepsize schedules.

D4-3 随机梯度下降算法在高维回归问题中正则效应与泛化性能分析

方聪, 北京大学

摘要：随机梯度下降算法是求解机器学习问题中的常见算法。在高维学习问题中，随机梯度下降算法的迭代次数往往低于模型参数量，算法对于模型的产生隐式正则效应是模型具有良好泛化的主要原因。本次讲座，我们将研究随机梯度下降算法在不同学习情境下求解线性与简单非线性模型的泛化性能，并进行定量比较。在线性模型中，我们将分别讨论算法在不同学习尺度（即样本数与问题维度不同依赖关系）与协变量偏移条件下的学习效率，尝试理解算法对于学习问题的适应性与涌现发生的条件。在非线性模型，我们将阐明算法能够自适应问题结构，突破一阶算法在离线情形下面临的统计-计算鸿沟（statistical to computational gap）诅咒。

D4-4 FZOO: Fast Zeroth-Order Optimizer for Fine-Tuning Large Language Models towards Adam-Scale Speed

叶海山, 西安交通大学

摘要：Fine-tuning large language models (LLMs) often faces GPU memory bottlenecks: the backward pass of first-order optimizers like Adam increases memory usage to more than 10 times the inference level (e.g., 633 GB for OPT-30B). Zeroth-order (ZO) optimizers avoid this cost by estimating gradients only from forward passes, yet existing methods like MeZO usually need tens of times more steps to converge. Can this trade-off between speed and memory in ZO be fundamentally improved? Normalized-SGD, for instance, demonstrates strong empirical performance with greater memory efficiency than Adam. In light of this, we introduce FZOO, a Fast Zeroth-Order Optimizer towards Adam-Scale Speed. On the one hand, FZOO reduces the total forward passes needed for convergence by employing batched one-sided estimates that adapt step-sizes based on the standard deviation of batch losses. On the other hand, it accelerates per-batch computation through the use of Rademacher random vector (±1) perturbations coupled with CUDA's paral lel processing capabilities. Extensive experiments on diverse models (including RoBERTa-large, the OPT family (350M-66B), Phi-2, and Llama3) across 11 varied downstream tasks validate FZOO's effectiveness. On average, FZOO outperforms

MeZOby+3%inaccuracy while requiring 3× fewer forward passes. Notably, for the RoBERTa-large model, FZOO achieves average improvements of +5.6% in accuracy and 18× reduction in forward passes compared to MeZO, achieving con vergence speeds comparable to Adam. We also provide theoretical analysis proving FZOO's formal equivalence to a normalized-SGD update rule and establishing its convergence guarantees. Beyond full-parameter tuning, FZOO plugs smoothly into PEFT techniques, unlocking even larger memory savings. Taken together, our results make single-GPU, high-speed, full-parameter fine-tuning realistic today and point toward future work on memory-efficient pre-training.

分会报告 D5 专题：深度学习与科学计算

D5-1 OmniFluids: Unified Physics Pre-trained Modeling of Fluid Dynamics

张瑞, 中国人民大学

摘要：Computational fluid dynamics (CFD) underpins progress in numerous scientific and engineering fields, yet high-fidelity simulations remain computationally prohibitive. While machine learning approaches promise acceleration, they typically specialize in single physical systems or demand extensive training data, hindering their practical deployment. We introduce OmniFluids, a unified, physics-only pre-trained model that captures fundamental fluid dynamics laws and adapts efficiently to diverse tasks with minimal data. To achieve this, we develop a training framework combining physics-only pre-training, coarse-grid operator distillation, and few-shot fine-tuning. This enables OmniFluids to retain broad physics knowledge while delivering fast and accurate predictions. Architecturally, OmniFluids integrates a mixture of operators, a multi-frame decoder, and factorized Fourier layers, which enable efficient and scalable modeling of diverse physical tasks while maintaining seamless integration with physics-based supervision. Across a broad range of two- and three-dimensional benchmarks, OmniFluids outperforms state-of-the-art AI-driven methods in flow field prediction and turbulence statistics, delivering 10-100x speedups compared to classical solvers, and accurately recovers unknown physical parameters from sparse, noisy data. This work demonstrates the potential of training a unified CFD solver solely from physics knowledge, establishing a new paradigm for efficient and generalizable surrogate modeling across complex fluid systems.

D5-2 复杂流场环境的智能感知与控制

蔡声泽, 浙江大学

摘要：流场实验测量与计算模拟在航空航天、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。在实验测量方面，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，本报告介绍基于数据驱动深度学习的粒子图像与粒子跟踪测速方法；随后，采用物理启发式神经网络（PINNs），实现从稀疏测量到全局流场的反问题计算。为构建流体力学模拟的替代模型，本报告简要介绍基于神经网络算子的

流体动力学计算方法。同时，借助流体预测模型与强化学习算法，实现复杂流场环境下的智能体导航控制。

D5-3 基于神经 PDE 求解器的生物医学反散射成像

孙赫, 北京大学

摘要：深层生物组织的高分辨率成像长期受到强散射与非线性传播的计算瓶颈制约。例如，肌骨和颅脑等含骨器官一直被视为超声成像的“禁区”。神经 PDE 求解器凭借其高效、稳定地逼近复杂 PDE，有望成为反散射物理建模的有力工具。然而，受限于生物医学数据匮乏及现有神经算子网络泛化能力不足，其成像效果仍不令人满意。本报告将介绍一种融合生成式 AI 与新型强散射神经算子架构的反散射求解框架。该方法仅需数十张跨模态 CT 图像，即可构建高精度超声 PDE 模拟器，并高效重建三维超声影像。基于此，我们首次实现了媲美磁共振（约 1 mm）分辨率的人体肌骨组织三维超声成像，为运动损伤和肿瘤等疾病的临床诊断提供了全新定量评估方案。该框架还可推广至类器官与胚胎等厚样本的光学衍射层析成像，为细胞生物学研究开辟新路径。

D5-4 基于人工智能的科学知识自动发现

陈云天, 宁波东方理工大学

摘要：本报告聚焦于人工智能（AI）在科学知识发现领域的最新进展，探讨了 AI 如何助力科学探索与物理化学规律揭示。通过符号数学与 AI 算法的结合，我们成功验证了方法对带交互项 Burgers 方程、具有高阶导数的 KdV 方程和含指数项 Chafee-Infante 方程的提取能力，充分证实了该方法的准确性与稳健性。进一步地，我们应用该方法揭示了此前未知的全新方程，用于粘性重力流、复杂地形降水和化学极性等复杂现象的建模。在最新研究中，我们基于从数学手册中提取的 200 余类方程进行训练，构建出当前效果最佳的方程发现模型。报告将重点介绍 SGA、DISCOVER 及应用于科学探索的大语言模型 LLM4ED 等前沿模型，为科学家跨越知识与数据壁垒、深入理解自然提供了全新方法。

分会报告 D6 专题：机器学习与复杂系统

D6-1 地球系统复杂性及 AI

樊京芳, 北京师范大学

摘要：2021 诺贝尔物理学奖颁给了研究复杂系统的真锅淑郎、克劳斯·哈塞尔曼以及乔治·帕里西以表彰他们对因“对我们理解复杂系统的开创性贡献”。而前两位对地球气候的物理建模、可变性量化和全球变暖的可靠预测领域做出了突出成就。作为复杂自适应的地球系统，可能存在多个潜在的临界要素。而各个临界要素之间的相互作用可能对其他子系统产生稳定或不稳定的影响，从而可能导致突然地级联失效,使得气候变化的突变和不可逆转的威胁越来越大，人们必须切实的采取有效地行动来缓和气候变化带来的负面影响。然而，由于地球系统本身的复杂结构以及存在着众多非线性相互作用，使得人们对于上述灾难性事件的理解，尤其是预测方面变得困难重重。这也是科学界和公共政策的决策者极为关注的话题之一。本次报告我将会对有关复杂性理论以及 AI 如何应用于地球复杂系统进行阐述。

D6-2 Restoring Network Evolution with Transferable Graph-Based Machine Learning

胡延庆, 南方科技大学

摘要：The structural evolution of complex networks, such as biological and social systems, poses significant challenges for analysis due to their intricate dynamics and diverse domains. Here, we introduce a transferable machine-learning framework that integrates graph neural networks and graph transformers to reconstruct the evolutionary trajectories of networked systems. Validated across multiple network domains, our approach achieves up to 20% higher prediction accuracy, reducing model complexity by over 60% and computation time by more than two orders of magnitude compared to state-of-the-art methods. By leveraging transfer learning, it reliably infers any network's evolution without prior temporal data. Applying this framework, we infer the formation times of over 2.6 million neural connections in the Drosophila brain for the first time, revealing a strong correlation between connection formation time and functional essentiality. Our work paves the way for decoding the evolution of complex networks and harnessing cross-domain transfer learning, unlocking new frontiers in network science and beyond.

D6-3 凸优化框架下的物理信息机器学习

赖志路, 香港科技大学（广州）

摘要：物理信息机器学习（Physics-Informed Machine Learning, PIML）提供了一种深度融合数据与物理规律以求解重要科学问题（如方程求解、参数估计、推测隐含物理量、方程挖掘和状态预测等）的新范式。然而，物理信息机器学习在实际应用中仍面临诸多优化方面的挑战，严重限制了其推广与应用。为克服优化困难，本工作提出了一种凸优化框架，将训练物理信息机器学习模型转化为凸优化问题，该框架被称为 Convex-PIML。首先引入 B-样条基函数（B-spline basis functions）的线性组合来拟合数据以增强损失函数的凸性。同时，使用适当的凸函数近似损失函数中的非凸部分（一般为物理损失项），将原始非凸优化问题转化为一系列逐步细化的凸优化子问题。该转化允许成熟的凸优化算法被用来高效且稳定地求解凸优化子问题，最终提升了原始优化问题的求解精度。此外，本工作还引入了一种基于误差估计的自适应基函数节点优化方法，以缓解基函数的谱偏差问题，进一步提升了优化精度。我们在包含不同类型物理先验的多个典型场景验证了该框架的有效性，结果表明所提框架可以高效且高精度地求解物理信息机器学习的优化问题，具有广阔的应用潜力。

D6-4 基于时空信息转换的高维复杂系统预测与表征算法研究

彭昊, 华南理工大学

摘要：在自然科学和工程技术领域，高维非线性动力系统被广泛应用于描述各类复杂现象（例如气象变化、物理过程以及生物演化等）。对这些系统状态进行精准预测和表征，不仅是分析复杂系统的重要工具，也为揭示其内在演化规律提供了关键支持。然而，高维复杂系统具有高维度、多时空依赖性及强非线性耦合的特征，传统基于数学模型的方法难以全面刻画其动态演化机制。复杂系统的高维状态在相空间中不仅揭示了变量间的复杂交互作用（空间信息），还蕴含了系统在一段时间内的动态演化特征（时间信息）。基于 Takens 延迟嵌入定理和广义嵌入定理，我们建立了时空信息转换（Spatial-Temporal Information transformation, STI）方程。通过数据驱动的方法，我们构建了拓扑共轭的嵌入映射，将高维

系统的空间信息转换为低维显变量或隐变量的时间信息，发展了系列的高维复杂系统预测和表征方法，以探索时空信息转换在复杂系统分析、建模及相关应用中的潜在价值。

D6-5 Latent Iterative Refinement Flow: A Geometric-Constrained Approach for Few-Shot Generation

高婷, 华中科技大学

摘要：Few-shot generation, which synthesizes diverse, high-quality samples from limited training examples, is still a core chal- lenge in generative modeling. Current methods face two critical issues: models trained from scratch often overfit and memorize sparse data, while fine-tuning large pre-trained models can inherit domain biases but struggle to capture the latent space’s geometric structure. To address these limitations, we introduce Latent Iterative Refinement Flow (LIRF), a novel framework for few-shot generation as the progressive densification of latent manifold. We provide theoretical guarantees and convergence theorem in Hausdorff distance between generated and true data manifolds. Our method achieves substantial performanceimprovements, evidenced by a FID of 30.29 on CIFAR-10, significantly outperforming baselines such as Lottery Ticket Hypothesis (41.47) and AdvAug (41.25). We also demonstrate the framework’s scalability by generating coherent, high-resolution images on AFHQ-Cat. Comprehensive ablation studies confirm the critical necessity of both our manifold-preserving latent space and the contractive correction mechanism. In summary, LIRF offers a theoretically grounded and practically effective, domain-agnostic solution for data-scarce generative modeling.

分会报告 D7 专题：大模型数据准备

D7-1 Data-centric AI 基础设施

张文涛，北京大学

摘要：人工智能正从模型为中心（Model-centric AI）转向以数据为中心（Data-centric AI, DCAI），本次报告将探讨面向 DCAI 的数据基础设施体系，包括支持多模态数据统一管理的 AI 数据库，DataFlow 数据准备与动态训练工具。该体系突破了传统数据湖和数据处理工具的局限，实现了数据与模型的高效协同。通过大模型预训练、企业知识库构建等创新应用验证，展示了 DCAI 基础设施在提升模型性能、降低开发门槛方面的突破性价值，为人工智能向智能化计算新范式演进提供了系统解决方案。

D7-2 具身智能 VLA 大模型的数据研究

陆鸣，英特尔中国研究院

摘要：视觉-语言-动作 (VLA)大模型是具身智能领域的核心技术，然而相比于视觉-语言大模型 VLM，VLA 的基座大模型进展缓慢，主要原因在于 VLA 大模型的海量训练数据采集困难。本次报告会结合 VLM 的发展过程，从数据的角度深入分析 VLA 大模型的未来，同时也会介绍我们正在做的研究工作。

D7-3 视频生成背后的多模态理解技术

张远行, 快手

摘要：大语言模型中的经验启发着视频生成技术，在数据链路、模型架构、训练范式等方面带来了新的视角，推动着更加真实、可控、有创意的 AIGC 视频生成工具的发布。在视频生成技术的背后，多模态理解技术是最重要的支撑性技术之一。通用的多模态理解模型虽然能够在多种榜单上取得不错的成绩，但在视频生成视角下的精细化描述、理解等任务上表现的远远达不到及格线。在这个重要的落地场景中，专用的多模态理解模型将被用于多模态内容 Caption、多模态表征、视觉元素编辑等真实业务，实打实地影响到端到端的视频生成表现。本次报告中将着重介绍视频生成背后的多模态理解技术，围绕任务优化目标、数据组织、模型技术、评测技术几个方面展开，更进一步也将展示由多模态理解驱动的新 AIGC 应用玩法。

D7-4 多源多模态的下一代 ra g

陈冲, 华为

摘要：（TBD）

分会报告 D8 专题：生成式 AI 交叉研究

D8-1 智能医学成像和处理

陈阳，北京大学

摘要：人工智能正从模型为中心（Model-centric AI）转向以数据为中心（Data-centric AI, DCAI），本次报告将探讨面向 DCAI 的数据基础设施体系，包括支持多模态数据统一管理的 AI 数据库，DataFlow 数据准备与动态训练工具。该体系突破了传统数据湖和数据处理工具的局限，实现了数据与模型的高效协同。通过大模型预训练、企业知识库构建等创新应用验证，展示了 DCAI 基础设施在提升模型性能、降低开发门槛方面的突破性价值，为人工智能向智能化计算新范式演进提供了系统解决方案。

D8-2 深度学习赋能的药物发现与开发

符天凡, 南京大学

摘要：药物设计和开发是一个既漫长又昂贵的过程，涉及从分子发现到临床试验的多个复杂步骤。人工智能（AI）技术展示了巨大的潜力，可以显著加速这一过程并降低成本。在药物发现的初期阶段，目标是识别具备理想药理特性的分子。本报告将深入探讨最新的药物设计方法，包括连续空间深度生成模型和离散空间药物设计路径搜索算法。这些先进的 AI 工具能够高效地探索化学空间，预测新化合物的活性和安全性，并优化候选药物的设计，以满足特定的治疗需求。进一步讲，在药物开发的后期阶段，重点转向了临床试验，这是评估药物对人体安全性和有效性的重要环节。为了提高临床试验的成功率和效率，本报告将介绍一系列最新的可信赖的方法，包括可解释性、不确定性感知的临床试验设计与预测技术。这些方法不仅能够模拟真实的临床试验过程，还能帮助科学家更好地理解潜在的风险和收益，从而做出更加明智的决策。

D8-3 Towards AI for Genomics: GENERator & GENERanno

李秋熠, 阿里云

摘要：The rapid advancement of DNA sequencing technologies has significantly expanded our capacity to decode genomic information. However, the accurate interpretation and functional understanding of complex biological sequences remain major challenges in genomics. To address these challenges, we introduce two innovative tools—GENERator and GENERanno—that leverage large language model techniques for genomic sequence modeling and analysis. GENERator is a generative genomic foundation model built upon a Transformer decoder architecture. It achieves state-of-the-art performance across multiple benchmarks while maintaining remarkable efficiency: it matches the performance of Evo2 with only 1% of its training cost, making it one of the most resource-efficient DNA language models to date. GENERator excels at generating biologically plausible protein-coding sequences consistent with the central dogma of molecular biology, as well as optimizing regulatory elements such as enhancers with programmable activity profiles. These capabilities position it as a powerful tool for synthetic biology, functional genomics, and therapeutic design. In contrast, GENERanno is a compact yet powerful encoder-based model specifically optimized for metagenomic annotation. It consistently outperforms traditional HMM-based methods (e.g., GLIMMER3, GeneMarkS2, Prodigal) and recent LLM-based approaches (e.g., GeneLM), demonstrating robust generalization on archaeal genomes and novel species. Leveraging advanced contextual understanding, GENERanno pioneers key annotation capabilities—including pseudogene identification, taxonomic classification, gene fitness prediction, and antibiotic resistance profiling—directly from raw DNA sequences, eliminating reliance on reference databases or comparative genomics workflows. Together, GENERator and GENERanno form a synergistic framework that bridges the gap between generative modeling and precise functional annotation in genomics. This dual-model paradigm not only enhances our ability to interpret and engineer biological systems but also lays the foundation for next-generation AI applications in functional genomics, metagenomics, and beyond.

D8-4 语言与知识驱动的科学智能体

张强, 浙江大学

摘要：科学研究日益依赖于各类专业计算工具，但高效使用这些工具通常需要深厚的领域专业知识。尽管大型语言模型（LLMs）在工具自动化方面展现出潜力，但它们在整合和协同多个工具以完成复杂科研流程方面仍存在挑战。在此，我们提出 SciToolAgent，一个由 LLM 驱动的智能体，能够自动化地调用生物学、化学和材料科学领域的数百种科研工具。SciToolAgent 的核心是一个科学工具知识图谱，借助基于图的检索增强生成（retrieval-augmented generation）机制，实现智能的工具选择与执行。该系统还集成了完善的安全审查模块，以确保工具使用的规范性与伦理性。在精心构建的基准测试中，SciToolAgent 显著优于现有方法。其在蛋白质工程、化学反应性预测、化学合成以及金属有机框架筛选等任务中的案例研究进一步展示了其自动化复杂科研流程的能力，使先进的科研工具对专业人士与非专业人士均更加可及。

学生分会 E1 专题：AI 理论（I）

E1-1 (De)-regularized Maximum Mean Discrepancy Gradient Flow

陈宗昊，伦敦大学学院

摘要： We introduce a (de)-regularization of the Maximum Mean Discrepancy (DrMMD) and its Wasserstein gradient flow. Existing gradient flows that transport samples from source distribution to target distribution with only target samples, either lack tractable numerical implementation (f-divergence flows) or require strong assumptions, and modifications such as noise injection, to ensure convergence (Maximum Mean Discrepancy flows). In contrast, DrMMD flow can simultaneously (i) guarantee near-global convergence for a broad class of targets in both continuous and discrete time, and (ii) be implemented in closed form using only samples. The former is achieved by leveraging the connection between the DrMMD and the ξ\chi^2ξ-divergence, while the latter comes by treating DrMMD as MMD with a de-regularized kernel. Our numerical scheme uses an adaptive de-regularization schedule throughout the flow to optimally trade off between discretization errors and deviations from the ξ\chi^2ξ regime. The potential application of the DrMMD flow is demonstrated across several numerical experiments, including a large-scale setting of training student/teacher networks.

E1-2 Context-Size Scaling for Operator and In-Context Learning

刘雨濠，清华大学

摘要： Meta-learning, seeking to develop models capable of rapidly adapting to new tasks from limited data, has emerged as a prominent paradigm in modern machine learning. Notable examples include in-context learning and operator learning. Such tasks often exhibit a mismatch between training and test-time context sizes, for instance, in prompt length variation for in-context learning and resolution shifts for operator learning. In this talk, we investigate the transferability of model performance across varying context size and analyze how performance scales with both training and test-time context size.

E1-3 SGD Achieves Optimality for Least Squares via Power-Decay Learning Rates

王梓麟，北京大学

摘要： In this work, we study the problem of solving the least-squares regression task via one-pass stochastic gradient descent (SGD) in infinite-dimensional Hilbert spaces. Under standard source and capacity conditions, we investigate the role of learning rate schedules (LRS) in accelerating convergence. We propose a novel LRS, termed power decay, and provide a sharp theoretical analysis showing that SGD equipped with this schedule achieves the minimax optimal convergence rate. Remarkably, our method eliminates the logarithmic factors present in prior analyses, yielding the first provable instance where SGD attains optimality with a carefully designed LRS. Leveraging a continuous-time approximation framework, we further derive intuitive insights into the dynamics of SGD and establish a sufficient condition under which general LRSs attain optimal rates. Our results offer both theoretical advancements and practical guidelines for designing effective learning rate schedules in stochastic optimization.

E1-4 Architecture induces invariant manifolds of neural network training dynamics

赵佳杰，上海交通大学

摘要： The architecture of deep neural networks is widely recognized as the primary factor behind their exceptional performance. However, how architectural design influences training dynamics has remained unclear

for decades, largely due to the difficulty of isolating architectural effects from other variables. To address this challenge, we introduce a novel analytical framework based on geometric control theory for studying neural network dynamics. This framework involves: (i) relaxing the gradient flow dynamics into a geometric control problem to isolate the influence of architecture on training dynamics; and (ii) analyzing the resulting control orbits to identify architecture-induced invariant manifolds within the gradient flow. Our analysis reveals that symmetry—particularly permutation symmetry—is a key mechanism that gives rise to a hierarchy of invariant manifolds, ranging from low to high dimensions. The invariant manifolds demonstrate neuron condensation and equivalence to reduced-width networks, with dynamics that yield low-complexity fittings of the training data. Overall, our framework establishes a strong connection between deep learning and dynamical systems theory, opening new avenues for theoretical advances through the rich concepts and tools of dynamical systems.

学生分会 E2 专题：AI 理论（II）

E2-1 DICE: Data Influence Cascade in Decentralized Learning

朱同天，浙江大学

摘要：Decentralized learning offers a promising approach to crowdsource data consumptions and computational workloads across geographically distributed compute interconnected through peer-to-peer networks, accommodating the exponentially increasing demands. However, proper incentives are still in absence, considerably discouraging participation. Our vision is that a fair incentive mechanism relies on fair attribution of contributions to participating nodes, which faces non-trivial challenges arising from the localized connections making influence ``cascade'' in a decentralized network. To overcome this, we design the first method to estimate Data Influence CascadE (DICE) in a decentralized environment. Theoretically, the framework derives tractable approximations of influence cascade over arbitrary neighbor hops, suggesting the influence cascade is determined by an interplay of data, communication topology, and the curvature of loss landscape.DICE also lays the foundations for applications including selecting suitable collaborators and identifying malicious behaviors.

E2-2 The Underlying Mechanism behind Deep Learning: From Empirical Discoveries to Theoretical Attempts

周展鹏，上海交通大学

摘要：Despite the successes of modern deep neural networks, theoretical understanding of them still lags behind. Just like in many other scientific disciplines, a crucial step toward formulating a comprehensive theory of deep learning lies in empirical investigations of the learning pipeline, intending to uncover nontrivial phenomena that shed light on the underlying mechanisms. In the first part, I will present a study on the Sharpness-Aware Minimization (SAM). We find that SAM selects flatter minima over Stochastic Gradient Desecent (SGD) even when applied only during the last few epochs of training. We theoretically build a two-phase picture of the training dynamics of SAM in the late phase. This study advances our understanding of the

surprising generalization ability of neural networks. In the second part, I will introduce an intriguing phenomenon, Layerwise Linear Feature Connectivity (LLFC), which greatly strengthens the Linear Mode Connectivity (LMC) phenomenon that has been widely studied in the community. By adopting a feature-centric perspective, the study of LLFC transcends and advances our understanding of LMC.

E2-3 Why Rectified Flow is Better? Elucidating VP, VE and RF-based diffusion models

杨若峰，上海交通大学

摘要：Recently, rectified flow (RF)-based models have achieved a great performance in 2D, 3D, and video generation compared with previous variance preserving (VP)-based models (SD XL) and variance exploding (VE)-based models (EDM, Consistency Models). However, the theoretical explanation for the great performance of RF-based modes is lacking. This work starts from the sample complexity perspective and explicitly explains why RF-based models enjoy a better complexity than VP and VE-based models.

E2-4 Diffusion for Discriminative Modeling and Certification

陈焕然，清华大学

摘要：扩散模型已经在生成领域大红大紫。在本系列工作中，我们发现，扩散生成模型无需任何训练，即为一个鲁棒的判别模型，展示了生成模型与判别模型的对偶性。我们证明了扩散模型的 ELBO 具有 O(1) Lipschitzness，由此给出了扩散模型与扩散分类器的鲁棒性下界。最后，我们将这套理论分析拓展到了扩散模型以外的模型上。对于任意平滑分布与（有界的）神经网络，平滑后的函数的鲁棒性下界都可转化为分数背包问题或 0-1 背包问题进行求解。有趣的结论包括：RGB 空间的扩散模型，不可能被 Lp 的对抗样本攻击、不可能被基于优化的方法数据投毒、不可能做基于优化的数据蒸馏。当原始模型是有界函数时，平滑后的模型的鲁棒性下界可以被分数背包的贪心算法求解。当原始模型是二值函数时，平滑后的模型的鲁棒性下界可以通过 0-1 背包的动态规划算法得到更紧的解。当扩散模型、MaskGen、自回归模型达到相同的干净准确率时，扩散模型的鲁棒性下界严格大于 MaskGen 大于自回归模型。

学生分会 E3 专题：AI 算法

E3-1 Affine Equivariant Networks Based on Differential Invariants

李艺康，北京大学

摘要：In the field of geometric deep learning, equivariant networks enhance model efficiency and generalization by embedding symmetry prior knowledge into model design. However, most existing methods require discretization or sampling of groups, leading to increased model sizes for larger groups, with the affine group being a representative challenge. In this work, we build affine equivariant networks based on differential invariants from the viewpoint of symmetric PDEs, without discretizing or sampling the group. In the model construction, we innovatively normalize polynomial relative differential invariants under a special norm to create a new affine invariant, which effectively improves numerical stability when replacing classical differential invariants. For further flexibility, we design an equivariant layer, which can be directly integrated

into various standard network architectures. Moreover, the proposed framework for constructing equivariant networks is highly general and widely applicable, suitable for designing corresponding equivariant networks for the affine group and its continuous subgroups.

E3-2 LLaDA-V：对扩散语言模型进行视觉指令微调

游泽彬，中国人民大学

摘要：本报告介绍了 LLaDA-V，它是一个完全基于离散扩散模型构建的多模态大语言模型。该工作旨在探索替代主流自回归范式的新型架构，通过对掩码序列进行迭代式去噪来生成完整的答案。实验结果验证了该方法的可行性与竞争力，证明了扩散模型在多模态理解领域具备潜力，并为该领域提供了新的架构选择。

E3-3 The Sharpness Disparity Principle in Transformers for Accelerating Language Model Pre-Training

王锦波，北京大学

摘要：Transformers consist of diverse building blocks, such as embedding layers, normalization layers, self-attention mechanisms, and point-wise feedforward networks. Thus, understanding the differences and interactions among these blocks is important. In this paper, we uncover a clear sharpness disparity across these blocks, which emerges early in training and intriguingly persists throughout the training process. Motivated by this finding, we propose Blockwise Learning Rate (LR), a strategy that tailors the LR to each block's sharpness, accelerating large language model (LLM) pre-training. By integrating Blockwise LR into AdamW, we consistently achieve lower terminal loss and nearly 2× speedup compared to vanilla AdamW. We demonstrate this acceleration across GPT-2 and LLaMA, with model sizes ranging from 0.12B to 2B and datasets of OpenWebText, MiniPile, and C4. Finally, we incorporate Blockwise LR into other optimizers such as Adam-mini (Zhang et al., 2024c), a recently proposed memory-efficient variant of Adam, achieving a combined 2× speedup and 2× memory saving. These results underscore the potential of exploiting the sharpness disparity to improve LLM training.

学生分会 E4 专题：优化

E4-1 Bilevel Reinforcement Learning via the Development of Hyper-gradient without Lower-Level Convexity

杨俨，中国科学院数学与系统科学研究院

摘要：Bilevel reinforcement learning (RL), which features intertwined two-level problems, has attracted growing interest recently. The inherent non-convexity of the lower-level RL problem is, however, to be an impediment to developing bilevel optimization methods. In this talk, 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. This, remarkably, distinguishes our development of hyper-gradient from the general AID-based bilevel frameworks since we take advantage of the specific structure of RL problems. Moreover, we design both model-based and model-free bilevel reinforcement learning

algorithms, facilitated by access to the fully first-order hyper-gradient. Both algorithms enjoy the optimal convergence rate. To extend the applicability, a stochastic version of the model-free algorithm is proposed, along with results on its iteration and sample complexity.

E4-2 SPARKLE: A Unified Single-Loop Primal-Dual Framework for Decentralized Bilevel Optimization

孔博傲，北京大学

摘要：In this talk, we focus on decentralized bilevel optimization, in which multiple agents collaborate to solve problems involving nested optimization structures with neighborhood communications. Most existing literature primarily utilizes gradient tracking to mitigate the influence of data heterogeneity, without exploring other well-known heterogeneity-correction techniques such as EXTRA or Exact Diffusion. Additionally, these studies often employ identical decentralized strategies for both upper- and lowerlevel problems, neglecting to leverage distinct mechanisms across different levels. To address these limitations, we propose SPARKLE, a unified Single-loop Primal-dual AlgoRithm frameworK for decentraLized bilEvel optimization. SPARKLE offers the flexibility to incorporate various heterogeneitycorrection strategies into the algorithm. Moreover, SPARKLE allows for different strategies to solve upperand lower-level problems. We present a unified convergence analysis for SPARKLE, applicable to all its variants, with state-of-the-art convergence rates compared to existing decentralized bilevel algorithms. Our results further reveal that EXTRA and Exact Diffusion are more suitable for decentralized bilevel optimization, and using mixed strategies in bilevel algorithms brings more benefits than relying solely on gradient tracking.

E4-3 Stochastic optimization over expectation-formulated generalized Stiefel manifold

姜林硕，中国科学院数学与系统科学研究院

摘要：In this talk, we consider a class of stochastic optimization problems over the expectation-formulated generalized Stiefel manifold \eqref{sogse}, where the objective function f is continuously differentiable. We propose a novel constraint dissolving penalty function with a customized penalty term \eqref{cdfcp}, which maintains the same order of differentiability as f. Our theoretical analysis establishes the global equivalence between \ref{cdfcp} and \ref{sogse}, in the sense that they share the same first-order and second-order stationary points under mild conditions. These results on equivalence enable the direct implementation of various stochastic optimization approaches to solve \ref{sogse}. In particular, we develop a stochastic gradient algorithm and its accelerated variant by incorporating an adaptive step size strategy. Furthermore, we prove their $\mathcal{O}(\varepsilon^{-4})$ sample complexity for finding an ε-stationary point of \ref{cdfcp}. Comprehensive numerical experiments show the efficiency and robustness of our proposed algorithms.

E4-4 Subspace Optimization for Large Language Models with Convergence Guarantees

何雨桐，北京大学

摘要：Subspace optimization algorithms, such as GaLore (Zhao et al., 2024), have gained attention for pre-training and fine-tuning large language models (LLMs) due to their memory efficiency. However, their convergence guarantees remain unclear, particularly in stochastic settings. In this paper, we reveal that GaLore does not always converge to the optimal solution and provide an explicit counterexample to support this finding.

We further explore the conditions under which GaLore achieves convergence, showing that it does so when either (i) a sufficiently large mini-batch size is used or (ii) the gradient noise is isotropic. More significantly, we introduce GoLore (Gradient random Low-rank projection), a novel variant of GaLore that provably converges in typical stochastic settings, even with standard batch sizes. Our convergence analysis extends naturally to other subspace optimization algorithms. Finally, we empirically validate our theoretical results and thoroughly test the proposed mechanisms. Codes are available at https://github.com/pkumelon/GoLore.

学生分会 E5 专题：AI and PDE（I）

E5-1 In vivo 3D ultrasound computed tomography of musculoskeletal tissues with generative neural PDE solvers

曾祉竣，清华大学

摘要：Ultrasound computed tomography (USCT) holds great promise as a radiation-free, high-resolution modality for clinical imaging. However, its translation to bone-containing tissues —such as musculoskeletal systems—remains hampered by conventional ray-based beamforming reconstruction that neglects strong wave scattering physics. Here, we present an innovative Real2Sim2Real framework that fuses generative neural networks with physics-informed partial differential equation (PDE) solvers to achieve fast, high-fidelity 3D USCT. By learning a compact surrogate of the complete physics for ultrasonic wave propagation from only dozens of cross-modality images, our approach combines the accuracy of wave PDE modeling with the computational efficiency and stability of deep neural networks. This enables, for the first time, accurate and efficient quantitative wave-based imaging of in vivo human musculoskeletal tissues, providing spatial maps of acoustic properties rather than conventional reflection-mode images. On both synthetic benchmarks and in vivo human data (breast, arm, and leg), we reconstruct 3D maps of quantitative tissue parameters in under ten minutes, achieving unprecedented sensitivity to biomechanical properties (e.g., sound speed) in muscle and bone regions and delivering imaging resolution comparable to magnetic resonance imaging. By overcoming the computational bottleneck in strongly scattering regimes, our method paves the way for routine clinical USCT assessment of musculoskeletal diseases such as sarcopenia. This transformative deep learning framework also extends to other biomedical imaging challenges.

E5-2 Redefining Neural Operators in ξd + 1ξ Dimensions

宋昊泽，香港科技大学（广州）

摘要：Neural Operators have emerged as powerful tools for learning mappings between function spaces. Among them, the kernel integral operator has been widely validated on universally approximating various operators. Although recent advancements following this definition have developed effective modules to better approximate the kernel function defined on the original domain (with ξdξ dimensions, ξd=1, 2, 3...ξ), the

unclarified evolving mechanism in the embedding spaces blocks our view to design neural operators that can fully capture the target system evolution. Drawing on recent breakthroughs in quantum simulation of partial differential equations (PDEs), we elucidate the linear evolution process in neural operators. Based on that, we redefine neural operators on a new ξd+1ξ dimensional domain. Within this framework, we implement our proposed Schrödingerised Kernel Neural Operator (SKNO) aligning better with the ξd+1ξ dimensional evolution. In experiments, our ξd+1ξ dimensional evolving linear block performs far better than others. Also, we test SKNO's SOTA performance on various benchmark tests and also the zero-shot super-resolution task. In addition, we analyse the impact of different lifting and recovering operators on the prediction within the redefined NO framework, reflecting the alignment between our model and the underlying ξd+1ξ dimensional evolution.

E5-3 Harnessing Scale and Physics: A Multi-Graph Neural Operator Framework for PDEs on Arbitrary Geometries

李志豪，香港科技大学（广州）

摘要：Partial Differential Equations (PDEs) underpin many scientific phenomena, yet traditional computational approaches often struggle with complex, nonlinear systems and irregular geometries. This paper introduces the \textbf{AMG} method, a \textbf{M}ulti-\textbf{G}raph neural operator approach designed for efficiently solving PDEs on \textbf{A}rbitrary geometries. AMG leverages advanced graph-based techniques and dynamic attention mechanisms within a novel GraphFormer architecture, enabling precise management of diverse spatial domains and complex data interdependencies. By constructing multi-scale graphs to handle variable feature frequencies and a physics graph to encapsulate inherent physical properties, AMG significantly outperforms previous methods, which are typically limited to uniform grids. We present a comprehensive evaluation of AMG across six benchmarks, demonstrating its consistent superiority over existing state-of-the-art models. Our findings highlight the transformative potential of tailored graph neural operators in surmounting the challenges faced by conventional PDE solvers. Our code and datasets are available on \url{https://github.com/lizhihao2022/AMG}.

E5-4 Point Cloud Neural Operator for Parametric PDEs on Complex and Variable Geometries

曾晨宇，北京大学

摘要：Surrogate models are critical for accelerating computationally expensive simulations in science and engineering, particularly for solving parametric partial differential equations (PDEs). Developing practical surrogate models poses significant challenges, particularly in handling geometrically complex and variable domains, which are often discretized as point clouds. In this work, we systematically investigate the formulation of neural operators—maps between infinite-dimensional function spaces—on point clouds to better handle complex and variable geometries while mitigating discretization effects. We introduce the Point Cloud Neural Operator (PCNO), designed to efficiently approximate solution maps of parametric PDEs on such domains. We evaluate the performance of PCNO on a range of pedagogical PDE problems, focusing on aspects such as boundary layers, adaptively meshed point clouds, and variable domains with topological variations. Its

practicality is further demonstrated through three-dimensional applications, such as predicting pressure loads on various vehicle types and simulating the inflation process of intricate parachute structures.

学生分会 E6 专题：AI and PDE（II）

E6-1 Weak Generative Sampler to Sample Invariant Distribution of Stochastic Differential Equation

蔡志强，香港中文大学

摘要：The solution of many typical high-dimensional PDEs (such as the Fokker-Planck, and McKean-Vlasov equations) is associated with a probability distribution. To solve such PDEs by deep learning techniques is usually to simply find a neural network for the density function itself, subject to certain positivity and normalization conditions. The further utilization of the solution requires random sampling again. We introduce a framework of Weak Generative Sampler (WGS) to both solve the PDE and generate samples more efficiently than the PINN and the Ritz method. Our proposed loss function is based on the weak form and the generic probability interpretation of the loss function. The details of this talk will explain why the efficiency and adaptivity are so easy to achieve in this WGS for high-dimensional PDEs.

E6-2 Data-driven optimized high-order WENO schemes with low-dissipation and low-dispersion

周金蕊，电子科技大学

摘要：Classical high-order weighted essentially non-oscillatory (WENO) schemes are designed to achieve optimal convergence order for smooth solutions and to maintain non-oscillatory behaviors for discontinuities. However, their spectral properties are not optimal, which limits the ability to capture high-frequency waves and small-scale features. In this paper, we propose a data-driven optimized method to improve the spectral properties of the WENO schemes. By analyzing the approximate dispersion relation (ADR), the spectral error of the schemes can be bounded by the reconstructed errors of a series of trigonometric functions with different wavenumbers. Therefore, we propose the new schemes WENO5-JS/Z-NN that introduce a compensation term parameterized by a neural network to the weight function of the WENO5-JS/Z schemes. The neural network is trained such that the generated weights can minimize the reconstructed errors over a large number of stencils, and furthermore, improve the spectrum accuracy. Meanwhile, the Total Variation Diminishing (TVD) constraint and anti-dissipation penalization are incorporated into the loss function to enhance the shock-capturing capability and preserve stability in simulating high-frequency waves. Compared to WENO5-JS/Z, our schemes maintain the ability to capture discontinuities while providing higher resolution for problems with fine-scale flow features. The ADR indicates that the new schemes can match the exact spectrum more accurately over a broader range of wavenumbers.

E6-3 A physics-informed deep learning method for solving hydrate dissociation problems in sediment

卜凡，中国地质大学（北京）

摘要：Natural gas hydrates are considered a promising energy resource due to their vast reserves, yet the dissociation process during thermal stimulation remains complex and difficult to model accurately. This study presents a nested neural network framework for solving the hydrate dissociation problem in sediment. The

proposed approach integrates governing conservation laws and thermodynamic assumptions into a deep learning model comprising two nested neural networks—an outer network for predicting the temperature field and an inner network for estimating the location of the moving phase-change boundary. The model controls the residuals, initial and boundary conditions, and interface constraints at moving boundaries of partial differential equations through a customized adaptive loss function. Compared with fractional-order models, this framework demonstrates better performance in handling thermal effects and strong thermal coupling.

E6-4 High Order Integrated Reconstruction for Finite Volume Scheme

张晨晔，北京大学

摘要：In [L. Chen and R. Li, An Integrated Quadratic Reconstruction for Finite Volume Schemes to Scalar Conservation Laws in Multiple Dimensions, (2017)], an integrated quadratic reconstruction was proposed for finite volume methods on flexible unstructured grids, which satisfy a local maximum principle and has 3-th order for smooth solutions. However, the optimization process consumes an excessive amount of time overhead. To address the need for efficient solving of large-scale small-scale optimization problems, this study employs neural networks to learn the solution mapping of optimization problems, significantly improving computational efficiency.

E6-5 SPIKE: stable physics-informed kernel evolution method for solving hyperbolic conservation laws

苏华，北京大学

摘要：We present a Stable Physics-Informed Kernel Evolution (SPIKE) method for solving inviscid one-dimensional hyperbolic conservation laws under periodic boundary conditions—a purely physics-driven approach requiring no training data. By evolving adaptive kernels through direct minimization of the governing equation loss, SPIKE intrinsically preserves the conservation invariants of the system while dynamically aligning kernel propagation with local characteristic speeds. Numerical validation on prototypical equations demonstrates oscillation-free shock formation/propagation, with theoretical analysis proving strict adherence to Rankine-Hugoniot conditions. This framework outperforms physics-informed ML methods (e.g., PINN, EDNN) in robustness and efficiency, establishing a new paradigm for structure-preserving computation of conservation laws.

学生分会 E7 专题：AI for Science（I）

E7-1 AeroGTO: An Efficient Graph-Transformer Operator for Learning Large-Scale Aerodynamics of 3D Vehicle Geometries

刘鹏伟，浙江大学

摘要：Obtaining high-precision aerodynamics in the automotive industry relies on large-scale simulations with computational fluid dynamics, which are generally time-consuming and computationally expensive. Recent advances in operator learning for partial differential equations offer promising improvements in terms of efficiency. However, capturing intricate physical correlations from extensive and varying geometries while balancing large-scale discretization and computational costs remains a significant challenge. To address these

issues, we propose AeroGTO, an efficient graph-transformer operator designed specifically for learning large-scale aerodynamics in engineering applications. AeroGTO combines local feature extraction through message passing and global correlation capturing via projection-inspired attention, employing a frequency-enhanced graph neural network augmented with k-nearest neighbors to handle three-dimensional (3D) irregular geometries. Moreover, the transformer architecture adeptly manages multi-level dependencies with only linear complexity concerning the number of mesh points, enabling fast inference of the model. Given a car's 3D mesh, AeroGTO accurately predicts surface pressure and estimates drag. In comparisons with five advanced models, AeroGTO is extensively tested on two industry-standard benchmarks, Ahmed-Body and DrivAerNet, achieving a 7.36% improvement in surface pressure prediction and a 10.71% boost in drag coefficient estimation, with fewer FLOPs and only 1% of the parameters used by the prior leading method.

E7-2 Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport
张振毅，北京大学

摘要：Reconstructing dynamics using samples from sparsely time-resolved snapshots is an important problem in both natural sciences and machine learning. Here, we introduce a new deep learning approach for solving regularized unbalanced optimal transport (RUOT) and inferring continuous unbalanced stochastic dynamics from observed snapshots. Based on the RUOT form, our method models these dynamics without requiring prior knowledge of growth and death processes or additional information, allowing them to be learnt directly from data. Theoretically, we explore the connections between the RUOT and Schrödinger bridge problem and discuss the key challenges and potential solutions. The effectiveness of our method is demonstrated with a synthetic gene regulatory network. Compared with other methods, our approach accurately identifies growth and transition patterns, eliminates false transitions, and constructs the Waddington developmental landscape. This paper is accepted in ICLR 2025 as an oral presentation.

E7-3 Predicting Dynamical Systems across Environments via Diffusive Model Weight Generation
李瑞堃，清华大学

摘要：Data-driven methods offer an effective equation-free solution for predicting physical dynamics. However, the same physical system can exhibit significantly different dynamic behaviors in various environments. This causes prediction functions trained for specific environments to fail when transferred to unseen environments. Therefore, cross-environment prediction requires modeling the dynamic functions of different environments. In this work, we propose a model weight generation method, \texttt{EnvAd-Diff}. \texttt{EnvAd-Diff} operates in the weight space of the dynamic function, generating suitable weights from scratch based on environmental condition for zero-shot prediction. Specifically, we first train expert prediction functions on dynamic trajectories from a limited set of visible environments to create a model zoo, thereby constructing sample pairs of prediction function weights and their corresponding environments. Subsequently, we train a latent space diffusion model conditioned on the environment to model the joint distribution of weights and environments. Considering the lack of environmental prior knowledge in real-world scenarios, we propose a physics-informed surrogate label to distinguish different environments. Generalization experiments across

multiple systems demonstrate that a 1M parameter prediction function generated by \texttt{EnvAd-Diff} outperforms a pre-trained 500M parameter foundation model.

E7-4 基于混合机器学习架构的复杂流场长期高保真预测方法研究

李沛函，北京航空航天大学

摘要：准确预测高维非线性流体动力系统的长期演化行为对工业设计与科学计算具有重要意义，然而传统数值方法面临计算成本过高与误差累积的挑战，现有数据驱动模型则受限于长期预测的稳定性缺失与物理一致性不足。本研究提出一种融合状态空间建模与生成修正的混合机器学习框架。该方法的核心架构包含三个紧密耦合的技术模块：首先通过对计算流体力学（CFD）时间序列快照进行本征正交分解（POD），提取包含 99%能量的主导模态构建低维子空间，并采用 Mamba 状态空间模型建模该子空间动力学。针对 POD 截断导致的高阶模态缺失问题，第二模块提出基于隐空间扩散模型（LDM）的正交补空间生成修正策略，通过图核算子神经网络实现非结构化 CFD 网格到规则拓扑结构的几何自适应映射。为确保生成结果符合物理规律，第三模块引入微分方程约束嵌入机制，利用图核算子编码后的网格不变特征解析推导拉普拉斯算子与梯度算子的离散估计值，进而计算 Navier-Stokes 方程残差，通过扩散后验采样理论，在 DDPM 的逆扩散采样阶段迭代注入物理残差约束，使生成样本收敛至物理可行解域。通过后台阶流瞬态流场的验证表明，本框架可稳定预测全阶流场演化。相较于传统方法，所提出架构在长期稳定性、复杂几何细节还原及物理场一致性方面展现显著优势。

学生分会 E8 专题：AI for Science（II）

E8-1 面向机器学习的材料数据质量评价体系与指标构建方法

左维，上海大学

摘要：随着机器学习在材料科学领域的广泛应用，高质量数据已成为保证模型可靠性与预测精度的前提。然而，现有研究多停留于数据清洗或单一指标评估，缺乏面向机器学习需求的系统化质量评价框架和评价指标。因此，本文提出了一种面向机器学习的材料数据质量评价体系与指标构建方法，从材料数据全生命周期和评价学五大公理出发，系统定义了材料数据的静态质量与动态质量两类属性，并构建了可覆盖不同材料数据质量类型的抽象指标体系。在此基础上，该方法创新性地基于“继承—发展”思想，将抽象指标向多模态、不同粒度的数据延伸，能够系统性地形成适用于具体应用场景和任务需求的量化指标。最后，以材料结构化数据和文本数据为例，对所提方法进行了实例化应用，给出了统一的量化计算流程与分层级阈值标准，为后续的质量诊断提供了可量化的依据，保证了材料数据质量评价的客观性和可比性。所构建的评价体系与指标方法具有较强的系统性与可扩展性，可为材料领域的机器学习应用提供规范化技术支持。

E8-2 统一匹配框架：少样本场景下的分子性质预测任务新解

李瑞凤，浙江大学

摘要：药物发现在识别治疗各种疾病的候选药物中起着关键作用。然而，其成功率较低，常常导致标注

数据稀缺，从而形成小样本学习问题。现有方法主要关注单尺度特征，忽略了决定不同分子性质的分子层级结构。为了解决这些问题，我们提出了通用匹配网络（UniMatch），一种双重匹配框架，通过元学习将显式的分子层级匹配与隐式的任务级匹配相结合，连接多层次分子表示与任务级泛化能力。具体而言，我们的方法通过层级池化与匹配显式捕捉多个层次（如原子、子结构和分子）的结构特征，从而实现精确的分子表示与比较。此外，我们采用一种元学习策略进行隐式任务级匹配，使模型能够捕捉任务间的共享模式，并快速适应新任务。该统一匹配框架既实现了有效的分子对齐，又利用了共享的元知识进行快速迁移。 我们的实验结果表明，UniMatch 在 MoleculeNet 和 FS-Mol 基准测试上优于当前最先进的方法，AUROC 提升了 2.87%，ΔAUPRC 提升了 6.52%。此外，UniMatch 在 Meta-MolNet 基准测试中也展现出优秀的泛化能力。

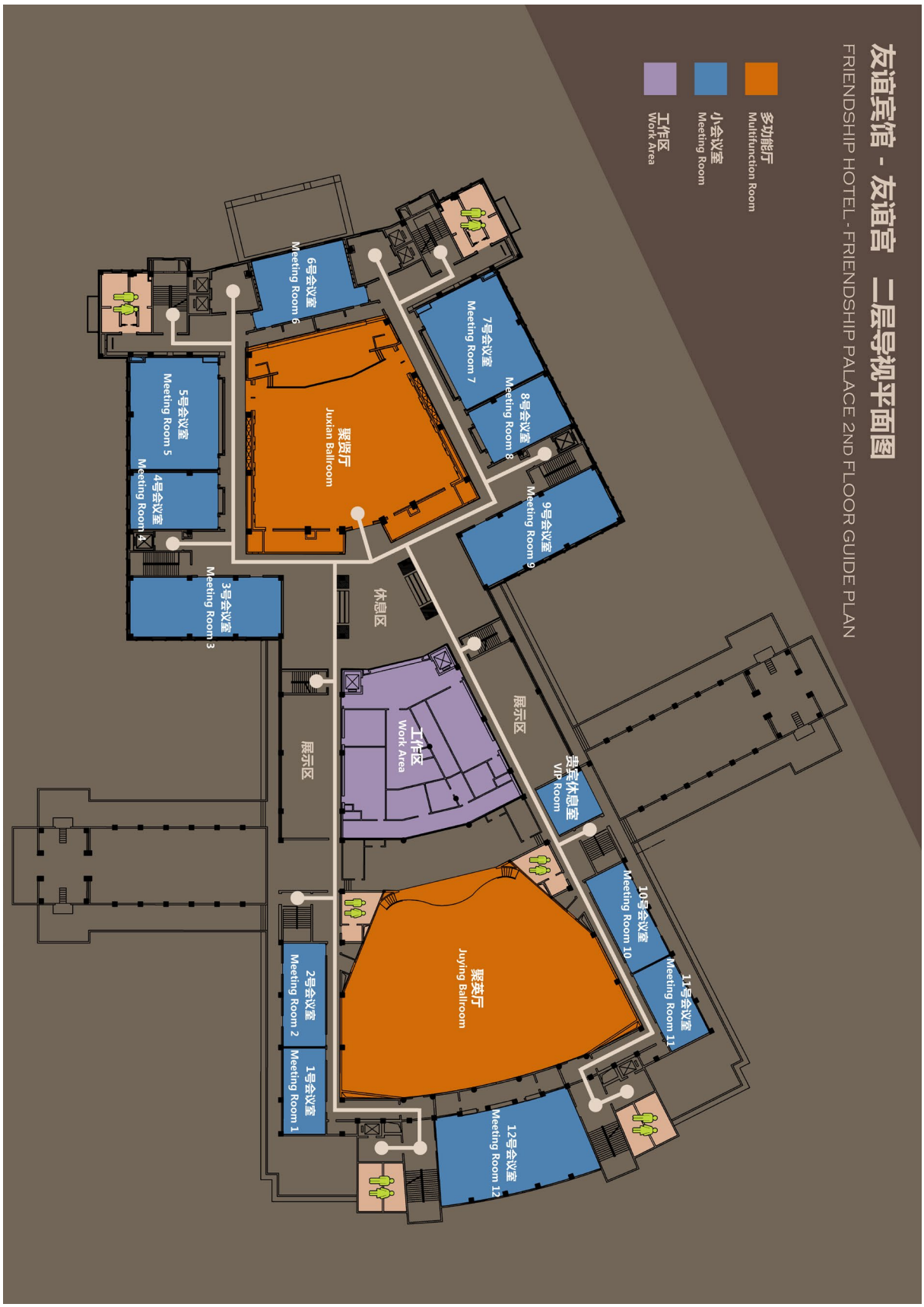
E8-3 面向高比能固态锂电池的聚合物电解质——从分子设计到智能预测

周倩，中国科学院青岛生物能源与过程研究所

摘要：聚合物电解质对固态电池的性能具有重要影响。报告人前期通过分子设计提出了多种新型聚合物电解质，分别改善了聚合物电解质的室温离子传输性能、电化学稳定窗口、电池安全等性能。然而，传统基于实验试错法和经验指导的聚合物电解质设计方法，存在研发周期长、效率低、难以突破认知局限等问题。最近，随着 AI for Sicence 热潮的持续升温，AI 在材料科学中的成功应用为加速新型功能材料的设计与筛选提供了全新路径。 申请人通过系统的控制变量法与多参数相关性分析发现，聚合物基体的第一电离能与其氧化稳定性之间存在显著正相关性，可作为聚合物电解质抗氧化性能评估的有效描述符。然而，现有数据库中聚合物的电离能等关键物理参数极为稀缺，且第一性原理计算成本高昂，严重制约了该方向的进一步发展。 为此，申请人构建了一种融合因果解耦建模与迁移学习策略的深度学习预测框架，成功实现了对超过 13,000 种聚合物分子结构的电离能等电子性质的高效、精准预测。基于该模型的预测结果，筛选出多种在氧化稳定性方面显著优于传统聚环氧乙烷（PEO）和聚碳酸丙烯酯（PPC）的新型聚合物电解质体系。 本工作首次提出并验证了聚合物电解质氧化稳定性相关的有效描述符，并结合数据驱动的机器学习方法，构建了从分子结构到电化学性能的映射关系，实现了对聚合物电解质材料的大范围快速筛选。该研究不仅为高电压固态锂电池的发展提供了新的材料解决方案，也为聚合物功能材料的智能化设计建立了可推广的研究范式。 鉴于该成果在材料智能设计与能源电化学领域的前沿性和广泛适用性，特此申请在 CSML2025 会议中做报告，以期与广大同仁深入交流，共同推动 AI+材料的融合发展。

六、会场分布示意图


北京友谊宾馆示意图
Plan of Beijing Friendship Hotel
北 N
通往理工大学
Lead to BIT
北环西路
Beisanhuan West Rd.
中关村南大街
Zhongguancun South St.
四通桥
Sitong Bridge
11号楼 Building 11
西北门 Northwest Gate
颐园写字楼 Yiyuan Office Building
颐园公寓 Yiyuan Apartment
瑞宾楼 Building 7
康体中心 Fitness Center
嘉宾楼 Building 8
怡宾楼 Building 4
邮局 Post Office
Building 2
敬宾楼
友谊宫 Friendship Palace
悦宾楼 Building 5
Building 3
迎宾楼
乡园公寓 Xiangyuan Apartment
北京科学会堂 Beijing Science Hall
北门 North Gate
中心庭院 Central Garden
环岛 Round About
环岛 Round About
贵宾楼 Grand Building
东门 East Gate
12号楼 Building 12
苏园写字楼 Suyuan Office Building
苏园公寓 Suyuan Apartment
雅园写字楼 Yayuan Office Building
雅园公寓 Yayuan Apartment
结账中心
首都机场班车站 Airport Shuttle Bus Station
地铁4号线人民大学站 Renmin University Station Subway line 4


友谊宾馆 - 友谊宫 二层导视平面图
FRIENDSHIP HOTEL - FRIENDSHIP PALACE 2ND FLOOR GUIDE PLAN
多功能厅 Multifunction Room
小会议室 Meeting Room
工作区 Work Area
聚贤厅 Juxian Ballroom
聚英厅 Juying Ballroom
工作区 Work Area
休息区
展示区
展示区
贵宾休息室 VIP Room
1号会议室 Meeting Room 1
2号会议室 Meeting Room 2
3号会议室 Meeting Room 3
4号会议室 Meeting Room 4
5号会议室 Meeting Room 5
6号会议室 Meeting Room 6
7号会议室 Meeting Room 7
8号会议室 Meeting Room 8
9号会议室 Meeting Room 9
10号会议室 Meeting Room 10
11号会议室 Meeting Room 11
12号会议室 Meeting Room 12