

顶会观察

ICLR 2023

上海人工智能实验室青年科学家 李弘扬

国际学习表征会议 (International Conference on Learning Representations, ICLR) 是深度学习的顶级会议之一, 受到学术研究者广泛认可, 在谷歌发布的人工智能领域 “top publication” 排名中位列前十。该会议由深度学习三大巨头之二的 Yoshua Bengio 和 Yann LeCun 牵头创办。Yoshua Bengio 是蒙特利尔大学教授, 深度学习三巨头之一, 他领导蒙特利尔大学的人工智能实验室 MILA 进行 AI 技术的学术研究。Yann LeCun 同为深度学习三巨头之一的他现任 Facebook 人工智能研究院 FAIR 院长、纽约大学教授。ICLR 的创办旨在提供一个场所, 能够让学者们交流分享表征学习领域所关心的话题, 为深度学习提供一个专业化的交流平台。今年的 ICLR 于 2023 年 5 月 1-5 日在卢旺达举办第十一届会议, 这是机器学习领域顶会首次在非洲国家举办, 这将会极大促进非洲国家科技发展及民众对前沿科技的认知, 进而带来广泛而深远的国际影响力。

一、会议亮点

开放的评审机制: ICLR 采用了开放评审的评审制度, 根据规定, 所有提交的论文都会公开姓名等信息, 并且接受所有同行的评价及提问, 任何学者都可或匿名或实名地评价论文。而在公开评审结束后, 论文作者也能够对论文进行调整和修改。ICLR 是在开放评审方面做得最公开、影响范围最大的一个会议。目前 ICLR 的历届所有论文及评讨论论的内容, 都完整地保存在 OpenReview 上, 它也是 ICLR 的官方投稿入口。OpenReview 是马萨诸塞大学阿默斯特学院 Andrew

McCallum 为 ICLR 2013 牵头创办的一个公开评审系统, 秉承公开同行评审、公开发表、公开来源、公开讨论、公开引导、公开推荐、公开 API 及开源等八大原则, 得到了 Facebook、Google、NSF 和马萨诸塞大学阿默斯特中心等机构的支持。

开源代码: 随着 ICLR 影响力的进一步提升, 越来越多的研究工作发布了代码或数据, 开源已经成为趋势。据 GitHub 项目的不完全统计, 今年已经超过 400 篇论文公开了源代码。代码开源的优势明显, 免费透明, 不仅可以增加研究者之间的协作机会, 而且还能提升研究工作的影响力。当然, 也存在一定的未知风险, 这需要大家共同努力完善开源代码。

二、录用情况

ICLR2023 收到的有效投稿和录用数量都有显著提高, 大会共收到了 4922 篇有效投稿, 相比上一年增加了 32.2%。最终接收了近 1573 篇论文, 接收率约为 31.8%, 录用率与去年基本持平。此外, 今年还有一个变化是接收论文的标签会有两个, 一个是论文类型 (oral, spotlight, poster), 另一个是 presentation 的方式。ICLR2023 会议涵盖的方向主要有强化学习、深度学习、表征学习、图神经网络等方向, 并且今年 Transformer 论文接收率下降, 语言模型论文增多。位于 top5% 共有 90 篇论文, 内容涉及 Transformer、in-context learning、扩散模型等内容。今年 ICLR2023, 共有 4 篇论文获得杰出论文奖, 5 篇论文获得杰出论文奖提名。其中, 来自北京大学的张博航、罗胜杰、王立威、贺笛共同获得一篇杰出论文奖, 清华大学孔祥哲、中国人民

大学高瓴人工智能学院黄文炳、清华大学刘洋共同获得一篇杰出论文奖提名。据统计，清华大学朱军教授和谷歌大脑的 Dale Schuurmans 两位学者发文数量并列。

三、 主题报告

本次 ICLR2023 会议邀请了六位 Keynote 演讲者，主题报告内容涵盖了一系列不同的研究主题，从生成模型，医疗保健和健康公平中的机器学习，统计机器学习/理论，到对话式人工智能和人工智能的创造力。每位演讲者分别进行了 60 分钟的演讲，主题报告激发了研究者们深入思考深度学习的技术基础及其对社会日益增长的影响。

Entanglements, Exploring Artificial Biodiversity. 随着生成系统的快速发展，生成系统已经在艺术创作过程中得到广泛的应用。Sofia Crespo 分享了她的艺术实践和使用生成系统的旅程，特别是神经网络，作为探索推测性生命体的一种手段，以及技术如何使我们更接近自然世界。

Understanding Systematic Deviations in Data for Trustworthy AI. 微软 AI for Good Research Lab 实验室的首席研究科学家 Girmaw Abebe Tadesse 讨论了在可信人工智能研究中的数据偏差问题。随着采用机器学习 (ML) 模型来协助决策的趋势越来越明显，为了实现值得信赖的 ML 应用，检查模型和其相应的数据的潜在系统偏差是至关重要的。这种检查过的数据可能被用于训练、测试或由模型本身产生。对系统性偏差的理解在资源有限和/或错误敏感的领域尤其关键，例如医疗保健。在这次演讲中，报告者将反思他们最近的工作，这些工作利用系统偏差的自动识别和表征来完成医疗领域的各种任务，包括：数据质量理解；时间漂移；异质干预效果分析；以及新类检测。此外，人工智能驱动的科学发现正越来越多地使用生成模型来促进。报告者分享他们以数据为中心的多层次评估框架如何帮助量化生成式模型的能力，并以材料科学作为一个使用案例，以领域诊断和可解释的方式。除了经常用来训练 ML 模型的策划数据集的分析，类似的以数据为中心的分析也应该被考虑在传统的数据源上，比如教科书。最后，报告者介绍最近在皮肤病学学术材

料中进行的自动表示分析的合作工作。

Importance-Weighting Approach to Distribution Shift Adaptation. RIKEN 中心主任、东京大学教授 Masashi Sugiyama 介绍了分布转移自适应问题中的重要性加权方法。对于可靠的机器学习，克服分布偏移是最重要的挑战之一。在这次演讲中，报告者首先概述经典的重要性加权方法来适应分布偏移，它包括一个重要性估计步骤和一个重要性加权的训练步骤。然后，报告者介绍一种较新的方法，即同时估计重要性权重和训练一个预测器。最后，报告者讨论一个更具挑战性的连续分布转移的场景，即数据分布随时间不断变化。

AI, History and Equity. 波士顿大学副教授 Elaine Nsoesie 介绍了人工智能在社会问题中的应用。大型数据集越来越多地被用来训练人工智能模型，以解决社会问题，包括健康方面的问题。有偏见的人工智能模型的社会影响已被广泛讨论。然而，在对话中有时缺少的是历史上的政策和不公正在塑造现有数据和结果方面的作用。通过历史视角评估数据和算法可能对社会变革至关重要。

Dialogue Research in the Era of LLMs. 来自 stealth AI startup 公司的 Dilek Hakkani-Tur 对大语言模型时代的对话研究进行了介绍。最近的大型语言模型 (LLMs) 由于能够对任何用户的要求产生连贯的自然语言反应，使得开放域对话系统取得了重大进展。它们记忆和进行组合推理的能力使对话相关的任务得以准确执行，如语言理解和响应生成。然而，这些模型存在局限性，例如，幻觉、不希望捕捉到的偏见、对特定政策的概括性困难，以及缺乏可解释性。为了解决这些问题，自然语言处理界提出了一些方法，例如，在训练或推理过程中向语言模型注入知识，使用多步骤推理和 API/工具检索相关知识，等等。在这次演讲中，报告者对相关工作进行概述。

Learned optimizers: why they're the future, why they're hard, and what they can do now. 来自 Google Brain 的研究员 Jascha Sohl-Dickstein 报告了在深度学习中广泛应用的优化器所存在的问题和改进方案。深度学习的成功有赖于所学函数在许多任

务中的表现大大超过手工设计的函数。然而，研究者们目前仍然使用手工设计的优化器来训练模型，这些优化器作用于手工设计的损失函数。报告者论证了这些手工设计的组件通常与期望的行为不匹配，研究者可以期望元学习的优化器表现得更好。报告者讨论使元学习优化器难以训练的挑战和原因。其中包括：混乱和高方差的元损失景观；元训练的极端计算成本；缺乏全面的元训练数据集；设计具有正确归纳偏差的学习型优化器的挑战；解释学习型优化器的行动方法的挑战。报告者分享了其中一些挑战的解决方案。报告者展示了实验结果，在许多情况下，学习型优化器的性能优于手工设计的优化器。此外，报告者讨论了元训练学习型优化器所带来的新功能。

四、 热点论文

在今年的 ICLR2023 的获奖论文中，共有 4 篇论文获得杰出论文奖，5 篇论文获得杰出论文奖提名。其中，来自北京大学的张博航、罗胜杰、王立威、贺笛共同获得一篇杰出论文奖，清华大学孔祥哲、中国人民大学高瓴人工智能学院黄文炳、清华大学刘洋共同获得一篇杰出论文奖提名。

杰出论文奖

论文 1: Universal Few-shot Learning of Dense Prediction Tasks with Visual Token Matching, 来自 KAIST 和微软亚洲研究院。该论文提出了一种用于密集预测任务的少样本学习 pipeline，密集预测任务包括语义分割、深度估计、边缘检测和关键点检测等。该研究提出了一个简单的统一模型，可以处理所有密集预测任务，并包含多项关键创新。该研究将激发密集预测的进一步发展，所提方法——例如视觉 token 匹配、情景 (episodic) 元学习——可以用于相关的多任务学习问题。

论文 2: Rethinking the Expressive Power of GNNs via Graph Biconnectivity, 来自北京大学，该论文基于双连通性 (biconnectivity) 提出一种 GNN 表达性度量新指标。具体来说，该研究提出了一种利用节点间距离的新算法，并在合成数据和真实数据中进行了演示。该研究表明：双连通性问题在理论和实践中都

有着广泛的潜在应用。

论文 3: DreamFusion: Text-to-3D using 2D Diffusion, 来自谷歌研究院和加州大学伯克利分校，该论文提出了一种基于文本生成 3D 模型的有效方法，而无需 3D 模型作为训练数据。该论文的关键思想是利用本生成图像的扩散模型，并通过将误差信号反向传播到 3D 模型的神经辐射场来生成 3D 模型。该方法是 SOTA 图像生成和 3D 建模的巧妙组合，在实践中效果极好，并将启发各种后续工作，包括基于文本的 3D 视频生成。

论文 4: Emergence of Maps in the Memories of Blind Navigation Agents, 来自佐治亚理工学院、Meta AI 等，该论文基于认知科学和机器学习的跨学科方法，让仅具备自我运动 (egomotion) (不具备任何其他感知) 的导航智能体学得有效表征，并实现有效导航。该研究对表征学习具有重要意义。

杰出论文奖提名

论文 1: Towards Understanding Ensemble, Knowledge Distillation and Self-Distillation in Deep Learning, 来自 Meta 和 MBZUAI，该论文试图从一个新的理论视角来理解知识蒸馏。作者认为对于自然的多视图结构，没有蒸馏的情况下神经网络只能训练为仅依赖于部分特征，而蒸馏可以缓解这个问题。这篇论文提供了证明这一点的简化示例，有助于人们更好地理解知识蒸馏的有效性。

论文 2: Mastering the Game of No-Press Diplomacy via Human-Regularized Reinforcement Learning and Planning, 来自 Meta AI 和 MIT，该论文的主题是多回合、多阶段、多人游戏的算法开发，提出使用一种类似于自我对弈 (self-play) 的策略来找到游戏均衡 (equilibrium) 状态，并在一个受人类玩家欢迎的复杂多人棋盘游戏上测试了该算法。其中，将寻求平衡的策略与行为克隆相结合。

论文 3: On the duality between contrastive and non-contrastive self-supervised learning, 来自 Meta AI 等，在自监督学习领域，各种方法似乎没有任何共同点，但在实践中却表现相似。该论文对各种自监督学习方法进行了分析探究，发现了它们的共同点。

该论文展开研究了一些流行的自监督学习方法，证明其提出的理论能用于实际方法。这篇论文对自监督学习领域具有重要意义。

论文 4: Conditional Antibody Design as 3D Equivariant Graph Translation, 来自清华大学计算机系, 清华大学智能产业研究院, 中国人民大学高瓴人工智能学院以及北京智源人工智能研究院, 抗体设计是药物研发的一个重要问题, 具有重要的应用前景。本文提出一种基于等变图神经网络的抗体设计方法 MEAN, 在给定抗原、抗体重链和轻链的条件下, 实现了抗体 CDRs 的设计和优化。与以往方法不同, MEAN 不但考虑了更全的「上下文信息」, 而且能直接生成抗体 CDRs 的 1D 氨基酸序列及其 3D 构象, 具有更高效率。在多个数据集的完整实验上, MEAN 显著优于已有方法。论文有望为后续湿实验研究提供一种高效的算法工具。

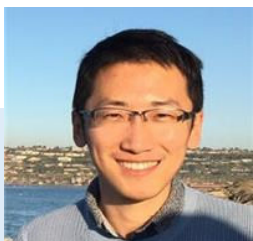
论文 5: Disentanglement with Biological Constraints: A Theory of Functional Cell Types, 来自斯坦福大学、牛津大学和 UCL, 该研究受生物学启

发, 揭示了机器学习和神经科学之间有趣的联系, 并从数学上证明机器学习中的约束会导致线性网络解缠结 (disentanglement)。该研究还通过实验表明, 相同的约束对于非线性情况也是有效的。总的来说, 这项研究从数学的角度对单个神经元和大脑结构给出了更深层的理解。

五、 总结展望

本年度 ICLR 大会中强化学习、深度学习、表征学习、图神经网络等领域依旧保持高热度。深度学习领域还面临许多挑战应该会包括: 从物理世界到传感器与机器认知过程中, 现有方法的能力上界是什么、人类已经发现的规律如何辅助机器认知物理世界、机器如何探知人类未曾发现的物理世界规律、深度学习模型的可解释性和可靠性等。回答好上述问题将会使得深度学习更好地迈向更高层级的智能。

责任编辑 魏秀参



李弘扬

李弘扬, 上海人工智能实验室青年科学家、高级工程师。2019 年香港中文大学博士毕业, 在读期间荣获香港政府奖学金。以第一作者身份, 在相关国际会议与期刊如 CVPR、NeurIPS、T-PAMI 等发表文章 10 余篇, 累计谷歌学术引用 1600 余次, 提交专利 20 余项。担任 CVPR、NeurIPS 2023 等会议领域主席。主持国家自然科学基金青年项目、上海市启明星项目。连续三年担任清华大学《计算机视觉》课程主讲人、上海交通大学行业博士生导师。2022 年获 Waymo 自动驾驶国际知名挑战赛第一名; 团队提出的工作 (BEVFormer、UniAD 等) 在国际上取得领先地位, 为多家自动驾驶公司提供了实际量产落地方案。提出的环视 3D 检测工作获 2022 年全球最有影响力的 AI 论文 Top 100。UniAD 工作获 CVPR 2023 最佳论文提名奖。
Email: lihongyang@pjlab.org.cn