

7码5期必中的倍投法

EMCm7DuGMf9IBRLV

7码5期必中的倍投法大模型何以擅长小样本学习？ICLR 2025这项研究给出详细分析

近年来，大语言模型（LLM）在人工智能领域取得了突破性进展，成为推动自然语言处理技术发展与应用人工智能实现的核心力量。上下文学习能力（In-Context Learning, ICL）是 LLM 最显著且重要的能力之一，它允许 LLM 在给定包含输入输出示例的提示（prompt）后，直接生成新输入的输出，这一过程仅通过前向传播而无需调整模型权重。这种能力使得 LLM 能够基于上下文中的示例快速理解并适应新任务，展现出强大的小样本学习和泛化能力。理解 LLM 是如何实现 ICL 的，对于提高模型性能与效率、提升模型可解释性与 AI 安全、推广大模型应用与改进小样本学习算法具有重要意义，也是近来机器学习研究热点之一。有以下关键问题需要回答：

1. LLM 能够学到哪些学习算法，例如梯度下降、比较近邻等？
2. 在具体问题的 ICL 过程中在执行哪一种学习算法？
3. 如何进一步提升 LLM 的 ICL 能力？

ICL 通常建模为将多个已知样例与预测目标输入一起，拼接成序列输入 LLM 中的 transformer 模型，输出对目标的预测（图 1 左）。现有工作已证明 ICL 在不同模型和数据分布条件下，能够分别实现如线性回归和梯度下降等具体的学习算法，从已知样例中学习对应输入输出映射，并作用于目标输入上产生预测输出。而这种学习算法是 transformer 模型通过预训练过程得到的，现实中 LLM 的预训练涉及海量的文本数据，含有复杂的语义信息，难以用单一的数学分布建模。现有工作对 ICL 实现小样本学习算法的解释难以泛化到真实世界场景或实际 LLM。为了对 ICL 的小样本学习能力有更直观的认识，在近期发表于 ICLR2025 的工作“Why In-Context Learning Models are Good Few-Shot Learners?”中我们对 ICL 模型作为元学习器的本质进行了建模与研究，以对上面三个问题进行了回答。

1. 将 LLM 建模为元学习器覆盖学习算法空间

ICL 模型可以学到所有传统元学习器学到的算法。元学习（Meta-Learning）是一种“学习如何学习”的方法，可通过设计模型使其能够快速适应新任务应用于小样本学习。它通过在多个相关任务上进行训练，学习到一种通用的学习策略或算法，从而在面对新任务时能够快速调整自身参数或结构，实现快速优化和泛化。借助元学习领域成熟的理论基础与方法经验，理论证明了作为实现学习算法的模型，基于 transformer 的 ICL 模型与传统的元学习器相比具有更强的表达能力（图 1 右）。

图 1 大语言模型的上下文学习示例，以及上下文学习模型在学习算法空间中与传统元学习模型的关系。

2. ICL 模型学到并执行在预训练分布上最优的算法

ICL 算法的学习是通过对预训练数据分布的拟合。在预训练充足的情况下，ICL 模型能够学习到在预训练任务集上最优（在与训练数据分布上最小化损失）的学习算法，从而在仅有少量样本的情况下实现快速适应。我们构建三类已知最优算法（Pair-wise metric-based/Class-prototype metric-based/Amortization-based 三种任务的最优算法分别可由元学习器 MatchNet/ProtoNet/CNPs 学习得到，图 3a）的任务。首先在单一种类任务集上训练，测试表明 ICL 性能与该预训练数据下能学到的最优算法表现相当（图 2 上）。然后再混合三种任务集上训练，三种传统元学习器的性能都有所下降，而 ICL 的性能依然与单一种类任务训练得到的最优性能一致（图 2 下）。以上结果说明 ICL 模型能够学习到预训练任务集上最优

的学习算法，并且与传统的元学习器相比 ICL 模型具有更强的表达能力，因为它们不仅能够学习到已知的最优学习算法，还能够根据数据的分布特性表达出传统视野之外的学习算法，这使得 ICL 模型在处理多样化任务时具有显著优势。

图 2 (上) 分别在三种任务集上训练并对应测试的测试表现；(下) 在混合任务集上训练并分别测试三种任务的性能表现。

我们还对 ICL 模型学习到的算法的泛化性进行了实验研究。展示出了其作为神经网络受数据分布影响的特性：其预训练过程本质上是在拟合以特定结构输入的训练任务集的数据分布，而无法保证学习到显式的基于规则的学习算法，这一发现纠正了现有工作将 ICL 解释为算法选择 (Algorithm Selection) 的过程。这将导致 ICL 模型在预训练数据受限或测试数据分布有偏移时性能表现不及预期 (图 3)。

图 3 ICL 与“算法选择”行为的比较 (a) 两种模型在三类已知最优算法的任务上训练，在未知最优算法任务上测试；(b) 对于测试任务 ICL 可以处理而“算法选择”无法处理；(c) ICL 对测试数据分布敏感而“算法选择”不敏感。

3. 将传统深度网络的相关方法迁移到元学习层面以提升 ICL 性能

基于上述对 ICL 模型作为学习算法强表达、难泛化的认识，可以将 ICL 模型对特性与传统深度神经网络的特性进行类比。我们提出通过“样本 - 任务”的概念映射将传统深度学习技巧迁移到元学习层面以优化 ICL 模型。例如实现了基于任务难度的元课程学习提升 ICL 模型预训练过程的收敛速度：图 4 展示了对于线性回归任务以递增非零维度数量作为课程的效果，元 - 课程学习能有效加速 ICL 模型的收敛，但不一定提升其最终性能。

图 4 元 - 课程学习 (左) 训练过程 loss 变化；(中) 200000 episodes 时的测试结果；(右) 500000 episodes 时的测试结果。

又例如实现了基于领域划分的元 - 元学习，即将训练数据划分为多个领域，每个领域含有一个训练任务集和验证任务集，即可将以单个任务为输入的 ICL 模型作为待适应网络，构建元 - 元学习器在每个领域上利用训练任务集进行适应。实验效果如图 5 所示，提升了 ICL 模型的有限垂域数据高效适应能力。

图 5 采用元 - 元学习的 ICL 模型分别在给定每领域 64/256/1024 个任务时的适应表现。

4. 总结

本文通过将 ICL 模型建模为元学习器，证明了 ICL 模型具有超过已有元学习器的表达学习算法的能力；ICL 执行在预训练数据分布上最优的算法，而不一定具有可泛化的规则；可以将传统神经网络有关技术迁移到元学习层面用以提升 ICL，如元 - 课程学习加速预训练收敛，元 - 元学习提升少数据领域微调快速适应能力。

作者介绍

吴世光，清华大学电子工程系博士研究生，本科毕业于清华大学电子工程系。当前主要研究方向包括元学习与大语言模型。

王雅晴，现任北京雁栖湖应用数学研究院 (BIMSA) 副研究员，长期从事机器学习、人工智能和科学智能的研究，致力于构建高效、低成本智能算法，以精准匹配海量数据的科学解释并解决现实问题。她在 NeurIPS、ICML、ICLR、KDD、WWW、SIGIR、TPAMI、JMLR、TIP 等国际顶级会议和期刊上发表 27 篇论文，总被引用 4500 次。2024 年，她入选全球前 2% 顶尖科学家榜单。

姚权铭，现任清华大学电子工程系助理教授，研究方向为机器学习和深度学习。共发表文章 100 + 篇，涵盖机器学习领域顶级期刊 JMLR、IEEE TPAMI 和顶级会议 ICML、NeurIPS、ICLR 等，累计引用超 1.2 万余次。担任 ICML、NeurIPS、ICLR 等会议领域主席，NN、TMLR、MLJ 等期刊（资深）编委。获首届蚂蚁 In Tech 科技奖、国际人工智能学会（AAAI）学术新星、国际神经网络学会（INNS）青年研究员奖、吴文俊人工智能学会优秀青年奖，同时入选全球 Top 50 华人 AI 青年学者榜和福布斯 30under30 精英榜。

[飞艇5码三期计划技巧](#)

[168澳洲10官网历史开奖记录](#)

[飞艇精准冠军计划全天免费](#)

[澳洲幸运10九码不死技巧](#)

[中国体育彩票内部资料网](#)

[快三数据分析器软件](#)

[10元小投资平台](#)

[澳洲190打分表](#)

[单双长期最稳公式](#)

[168澳洲10开奖结果查询](#)

[单双稳赢公式100%](#)

[大小单双计划导师赚钱](#)

[众赢国际版zygjbcom](#)

[澳洲10一期计划在线版](#)

[澳洲10的计划](#)

[pk10澳洲](#)

[最新澳洲幸运5开奖结果](#)

[如何分析澳洲幸运10](#)

[澳洲幸运10正规吗百度知道](#)