复旦大學

数据科学与深度学习博士生与青年学者论坛

从模仿到超越：Transformer如何学会比教师更高效的推理策略



曹原
（香港大学）

时间： 2025年12月1日 14:00-15:00
地点： 光华楼东主楼 2001

摘要： 现代Transformer常通过生成和改进中间推理步骤来解决复杂问题，而这些步骤通常来自人类示范。这引出了一个核心问题：Transformer是否只能复现教师的推理过程，还是能够学会不同且更高效的推理方式？本报告从“Transformer可被视为算法执行器”的理论视角出发，在上下文学习问题中研究这一现象。我们首先分析上下文逻辑回归，表明在梯度下降监督下训练的Transformer实际上会学会执行归一化梯度下降，并获得比教师算法更快的收敛速度。随后，我们研究上下文线性回归，发现进一步使用真实解进行监督后，模型会学会重球法（heavy-ball method），从而相较于梯度下降教师实现加速。这些结果表明，Transformer的推理并非简单模仿教师，而是可以在任务结构和监督信号的作用下发展出更高效的推理方式。

个人简介： 曹原是香港大学计算与数据科学学院助理教授。加入港大之前，他曾在加州大学洛杉矶分校从事博士后研究。他本科毕业于复旦大学，并在普林斯顿大学获得博士学位。曹原的研究兴趣包括深度学习理论、非凸优化和高维统计。