



郑庆华，教授，中国工程院院士，同济大学党委书记，教育部科技委学部委员，教育部大学计算机教学指导委员会主任，国家自然科学基金创新群体负责人。研究领域为大数据知识工程、机器记忆智能。获得国家科技进步二等奖3项，国家教学成果一等奖2项、二等奖3项，省部级科学技术一等奖5项、中国自动化学会科技进步特等奖以及何梁何利基金“科学与技术进步奖”、中国科协“求是”杰出青年奖、中国青年科技奖、国务院政府特殊津贴专家等奖项和荣誉。

人脑认知启发的机器记忆智能

郑庆华

以 ChatGPT、DeepSeek 为代表的大模型采取“大数据、大算力、强算法”的技术路线，极大地推动了语言理解、意图识别、内容生成、问题求解等领域的进展，成为当前人工智能的主流和巅峰。然而，在令人瞩目的能力跃迁背后，其固有缺陷与结构性瓶颈也日益凸显。

当前大模型主要面临四个方面的局限。其一，过度消耗数据与算力。参数规模动辄千亿乃至万亿级，需要海量数据拟合并消耗巨大算力与能源；推理模型因生成大量“思考”Token，能耗进一步攀升。其二，灾难性遗忘。在面向新任务微调时，模型通过全局参数更新适配新数据，却以牺牲既有知识为代价，难以实现真正的持续学习。其三，抽象能力弱。模型更擅长从大数据中提取统计模式，而非形成对概念本质的深层理解，推理鲁棒性差，出现各类幻觉。其四，黑盒难解释。复杂网络结构与高度非线性映射，使推理过程难以解释与溯因，更难以“知错就改”。

究其根源，在于当前人工神经网络架构、训练与推理机制的固有局限。首先，人工神经网络架构缺乏可塑性，缺少与问题背后机理的联系，难以刻画因果关系。其次，训练时反向传播机制以全局误差为优化目标，迭代更新所有参数，消耗大量训练数据与算力，并导致遗忘。再次，推理时所有参数都参与正向传播运算，消耗大量算力。这些根源性问题进一步导致大模型存在幻觉、可控性差、鲁棒性不足等缺陷。更为严峻的是，大模型的原创概念、核心技术与算力生态是由美国打造和主导，所做的是“延长线”式的工作。

回顾人工智能发展历史，脑科学一直推动着人工智能的发展，今天的大模型正是借鉴了大脑在网络拓扑、信息表征以及编码机制方面的理论。因此，深度借鉴脑科学的研究成果，有望找到一条更加有效的机器智能发展路径。

受此启发，我们提出了人脑记忆智能启发的机器智能新思路。脑科学的研究表明：记忆是智能的基础，深刻影响人脑学习、抽象、联想、推理等高级认知活动。第一，记忆的稀疏激活、预测编码等机制使人脑高度节能。人脑针对特定任务一般只激活约 1%~4% 的神经元，预测编码则减少了冗余信息处理，使得人脑高效提取记忆，快速学习新信息。第二，突触可塑性与吸引子机制是人类持续学习的基础。大脑采取局部学习策略，仅更新少量特定脑区的突触连接；同时人脑吸引子具有良好的抗噪声能力，能回忆出完整信息，避免了灾难性遗忘。第三，抽象-具象协同机制使得人脑能快速适应新任务、新场景。大脑能从具象信息中提取抽象概念，并将二者关联，形成完整、灵活且高效的记忆表征。第四，抽象与联想协同使得人脑推理更具逻辑性与可解释性。抽象有助于人脑在不同场景中进行类比和推理，联想因其连贯思维使得推理可解释。这些机制共同铸就了人脑低功耗、小样本、强泛化、可解释的智能特征，为设计机器记忆智能提供了天然蓝图。

“机器记忆智能”的核心思想是：将人脑记忆机理引入机器智能的表征、学习与推理过程。以吸引子为机器记忆基本单元，构建“层间抽象-层内联想”的表征结构；借鉴突触可塑性与稀疏激活机制，形成低功耗、小样本、强泛化的持续学习机制；将联想-混沌的动力学引入记忆激活与推理过程，实现过程可解释、结果能溯因并在不符合常识的环境下具有探索性推理的能力。机器记忆智能有望从根本上摆脱传统数据驱动的路径依赖，推动人工智能与脑科学深度交叉融合，走出一条不同于大模型的人工智能自主创新之路。