BrainMoE: A Brain-Inspired Modular Mixture-of-Experts Model for Multimodal Reasoning

Siyao Song
State Key Laboratory of Multimodal Artificial intelligence Systems, Institute of Automation, CAS, Beijing, China

Yunhao Zhang
State Key Laboratory of Multimodal Artificial intelligence Systems, Institute of Automation, CAS, Beijing, China

Yaping Zhang
State Key Laboratory of Multimodal Artificial intelligence Systems, Institute of Automation, CAS, Beijing, China

Chengqing Zong
National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, Beijing, China

Abstract

Inspired by the coordinated interaction of distributed brain networks that support distinct cognitive functions, recent brain-inspired language models have increasingly adopted modular architectures to improve efficiency and interpretability. However, most existing brain-inspired MoE models confined to text-only reasoning and lack effective mechanisms for multimodal integration. To address this gap, we propose BrainMoE, a brain-inspired multimodal mixture-of-experts architecture with six specialized experts for language processing, logical reasoning, theory of mind, world knowledge, visual perception, and auditory perception, coordinated by a token-level router. BrainMoE outperforms text-only baselines by an average of 4.13 accuracy points on text reasoning benchmarks and achieves competitive performance on multimodal evaluations. Further analyses show that BrainMoE routes tokens to task-relevant experts, while expert ablations cause significant task-specific performance drops, demonstrating interpretable specialization and causal importance. Overall, BrainMoE highlights the potential of brain-inspired modular architectures for interpretable multimodal reasoning.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*