Chen Institute Retreat 2023

Presenter: Pantelis Vafeidis

Poster title: Learning continuous attractors that allow zero-shot generalization in recurrent neural networks by task demands

Abstract: Humans and animals display a remarkable capacity to generalize to novel contexts, while artificial agents often fail to predict beyond their training distribution. Recent experimental and theoretical studies have pointed towards abstract, or disentangled, representations as an explanation of how such generalization to unseen parts of the state space is possible. Here we extend these findings to dynamical, real-life settings by training recurrent neural networks to integrate evidence streams over time towards a decision. We find that the networks learn abstract representations in the form of continuous attractors which store a short-term memory of the Cartesian product of accumulated evidence, but only when task demands span the latent space. We demonstrate the flexibility of our approach using more than one task type, continual learning and allowing for free reaction time decisions. Our findings directly map to decision-making experiments in humans and primates where similar representations have been discovered, and to the problem of path-integration where grid cells form continuous attractors that store the current location of the agent.