Relating Hopfield Networks to Episodic Control

Hugo Chateau-Laurent
Inria centre at the university of Bordeaux, Bordeaux, France

Frederic Alexandre
Inria centre at the university of Bordeaux, Bordeaux, France

Abstract

Neural Episodic Control is a powerful reinforcement learning framework that employs a differentiable dictionary to store non-parametric memories. It was inspired by episodic memory on the functional level, but lacks a direct theoretical connection to the associative memory models generally used to implement such a memory. We first show that the dictionary is an instance of the recently proposed Universal Hopfield Network framework. We then introduce a continuous approximation of the dictionary readout operation in order to derive two energy functions that are Lyapunov functions of the dynamics. Finally, we empirically show that the dictionary outperforms the Max separation function, which had previously been argued to be optimal, and that performance can further be improved by replacing the Euclidean distance kernel by a Manhattan distance kernel. These results are enabled by the generalization capabilities of the dictionary, so a novel criterion is introduced to disentangle memorization from generalization when evaluating associative memory models.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*