DYNAMICAL SIGNATURES OF LEARNING IN RECURRENT NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Recurrent neural networks are powerful computational tools that operate best near the edge of chaos, where small perturbations in neuronal firing are transmitted between neurons with minimal amplification or loss. In this article, we depart from the observation that both stimulus and noise can be seen as perturbations to the intrinsic dynamics of a recurrent network, however stimulus information needs to be reliably preserved, while noise needs to be discarded. Here, we show that self-organizing recurrent networks (SORNs) that learn the spatio-temporal structure of their inputs, increase their performance on memory and prediction tasks by preferentially propagating the relevant stimulus-specific structured signals and by increasing their robustness to random perturbation. The computational advantages gained through learning are accompanied by a shift from critical to ordered dynamics that is more pronounced for longer sequences and is stronger towards the end of a predictable sequence. We show that SORNs with subcritical dynamics outperform their random RNNs counterparts with critical dynamics on a range of tasks, including a temporal MNIST and a sequential shape-rotation task. We propose that this shift in criticality is a signature of specialization and we expect it to be found in all cases in which general-purpose recurrent networks acquire self-correcting properties by internalizing the statistical structure of their inputs.