Parallel Processing Capability Versus Efficiency of Representation in Neural Networks

Sebastian Musslick[1] **(musslick@princeton.edu), Biswadip Dey**[2] **(biswadip@princeton.edu)**
Kayhan Özcimder[2] **(ozcimder@princeton.edu), Md. Mostofa Ali Patwary**[3] **(mostofa.ali.patwary@intel.com),**
Theodore L. Willke[3] **(ted.willke@intel.com), and Jonathan D. Cohen**[1] **(jdc@princeton.edu)**
[1]Princeton Neuroscience Institute, Princeton University
[2]Department of Mechanical and Aerospace Engineering, Princeton University
[3]Parallel Computing Lab, Intel Corporation

Introduction

A key feature of neural networks is their ability to support the simultaneous interaction among large numbers of processes in the learning and processing of representations. However, how the richness of such interactions trades off against the ability of a network to simultaneously carry out multiple independent processes – a salient limitation in many domains of human cognition – remains largely unexplored. In this work we address this question using graph analytic tools in combination with neural network simulations. We describe initial findings that we hope will help lay the groundwork for a rigorous exploration of the factors that affect the tension between efficiency of learning, generality of representation, and parallel processing capability (PPC) in neural networks.

Methods

For the purpose of our graph analytic explorations, we consider a bipartite graph of task representations as a single-layer, feed-forward network in which edges between input and output nodes constitute a process (referred to as task). Each task corresponds to a unique mapping from an input representation to an output representation (simplified as input and output nodes respectively). Based on the assumption that cross-talk between processes arises due to shared representations, we derive an interference graph in which nodes correspond to tasks and edges represent interference between pairs of tasks. Finding the maximum independent set of this graph corresponds to finding the PPC of the network, i.e. the largest set of tasks that can be performed in parallel without interference. We investigate how network size (number of input/output nodes) and the degree of process overlap (out-degree of input nodes) affect the PPC across different networks. We also describe how the interference graph can be extracted from task representations encoded in neural networks or from neuroimaging data, and use this to compare predictions about multitasking performance for specific task combinations with actual multitasking performance in trained 2-layer neural networks. Finally, we analyze how multitasking capability trades off with the degree to which shared representations develop as a function of feature-overlap across tasks. To further investigate the tradeoff between efficiency of representation and multitasking capability, we assessed learning speed, generalization performance, and multitasking performance for networks in which weight vectors for similar tasks are either initially correlated or de-correlated (inducing a bias towards shared or separate task representations).

Results and Discussion

In line with earlier numerical work (*1*), we found that the average PPC of single-layer networks drops precipitously as a function of process overlap, and scales highly sublinearly with network size (Fig. 1A). Our simulations show that the extracted interference graph of a trained neural network can predict how well the network performs a given set of tasks simultaneously (Fig. 1B). Under the assumption of task-set inertia, parallel-processing limitations extend to networks in which multitasking is implemented sequentially, by switching between tasks as rapidly as possible: tasks predicted to be interfering were subject to greater switch costs that constrained serial multitasking performance (Fig. 1C). In accordance with observations that shared task representations are likely to develop in environments with high correlations between task-relevant features, we observe that multitasking capability of trained networks drops in such environments. We finally observe that weight priors on task similarity improve learning speed and generalization performance but lead to strong constraints on PPC.

Our simulation results identify a tradeoff between learning efficiency and multitasking capability, suggesting that multitasking limitations of a neural architecture may arise as the result of a bias towards shared representations and the value of learning efficiency.

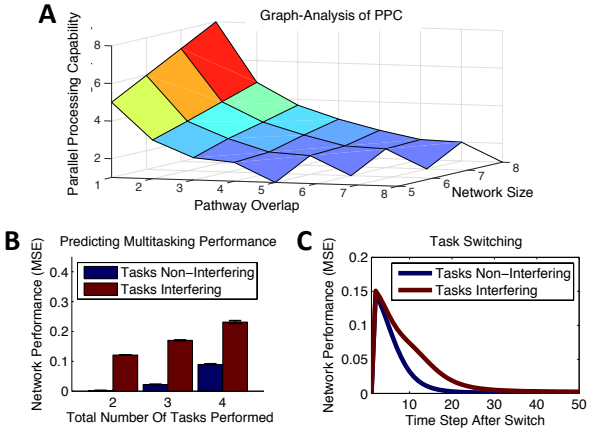

Figure 1: Graph analytical and neural network simulation results.

References

1. S. F. Feng, M. Schwemmer, S. J. Gershman, J. D. Cohen, *Cognitive, Affective, & Behavioral Neuroscience* **14**, 129 (2014).