Teaching a neural network to count: reinforcement learning with "social scaffolding"

Qihong Lu (qihong.lu@wisc.edu)
Department of Psychology, University of Wisconsin-Madison

James L. McClelland (mcclelland@stanford.edu)
Department of Psychology, Stanford University

Introduction

Counting skill is a foundation for more sophisticated math concepts, such as addition. Though it seems to be a simple task, it takes children several years to learn to do it well. Here, we consider the role of the teacher in helping a simple system learn to count.

Model training detail

We built a simple feed-forward neural network (Figure 1) that learns a counting-related task: given an array of identical objects randomly initialized on a one-dimensional line, the task is to touch every object exactly once and terminate the task at the end. The model has an eye and a hand that always move together, and its possible actions consist of movements within the range [-7, +7] from left to right. This will update what the model can see at the next time step, since objects in the periphery are fuzzy: as in the retina, the "acuity" of the input layer deteriorates with distance from the center. When the model's position overlays an object, we consider the model to be "touching" that object.

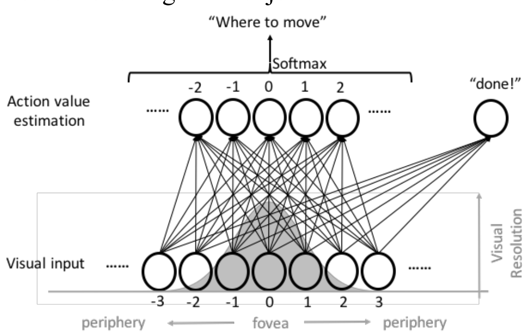

Figure 1: The architecture of the model

We compare several teaching regimes, including a standard reinforcement learning (RL) regime. Here the network learns to estimate action values with Q-learning (Minh et al. 2015; Sutton & Barto, 1998), and the network only receives direct environmental feedback when it actively terminates the task. Additionally, we designed two "teaching strategies" inspired from the social scaffoldings received by the children: intermediate feedback and demonstration. With intermediate feedback, the model receives a small positive reward each time it touches the next object from left to right. In the demonstration condition, we alternated demonstration trials, in which we forced the model to touch every object exactly once from left to right, with standard RL trials (reward comes only at the end in both kinds of trials).

Results

Compared to standard RL, both intermediate feedback and teacher demonstration improved learning significantly. In a fourth condition, we combined the two strategies, and this produced the best learning outcome (Figure 2), because the structured feedback and instructions relaxed the difficulties of the temporal credit assignment in the reinforcement learning setting. For example, providing intermediate feedback makes the task more supervised and teacher demonstration forces exposure to the Q-value of the optimal solution. These results provide insights about how social scaffoldings support learning from a computational perspective. Further research will extend these explorations to multi-layer recurrent architectures and more complex task settings.

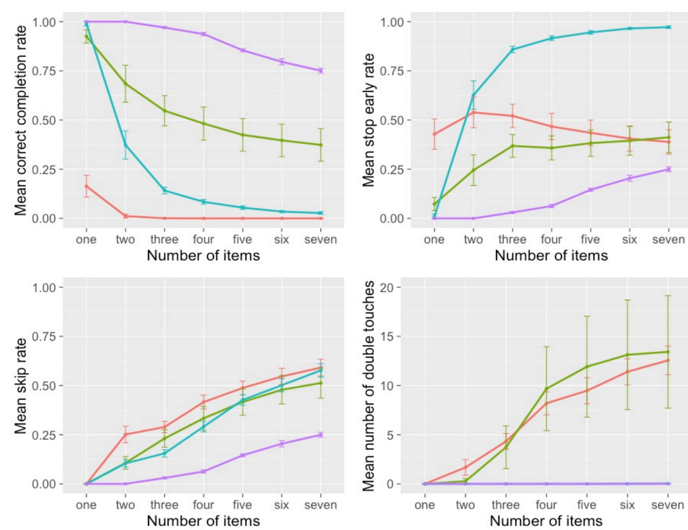

Figure 2. Compare different ways of teaching in terms of model's performance on the counting-related task.

References

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*: An introduction. MIT press.