Using pseudo-recurrent connectionist networks to solve the problem of sequential learning

Robert M. French (`rfrench@ulg.ac.be`)
Psychology Department, B32
University of Liège, Liège, Belgium

Introduction

One of the most important problems facing connectionist models of memory — in fact, facing *any* model of memory — is how to make them simultaneously sensitive to, but not disrupted by, new input. This problem is often referred to as the “sensitivity-stability” dilemma. It is particularly relevant for connectionist networks, especially since they can be afflicted by a particularly severe manifestation of the sensitivity-stability problem known as catastrophic interference. Catastrophic interference is the tendency of neural networks to abruptly and completely forget previously learned information in the presence of new input. The “pseudo-recurrent” architecture proposed in this paper relies on physically separating previously learned representations from those currently being learned (analogous to the hippocampal/neocortical separation in humans. See McClelland, McNaughton, O'Reilly, 1995). *Pseudopatterns* (Robins, 1995), i.e., approximations of the previously learned data, are extracted from the “neocortical” part of the network and are added to the set of new patterns to be learned in the “hippocampal” part.

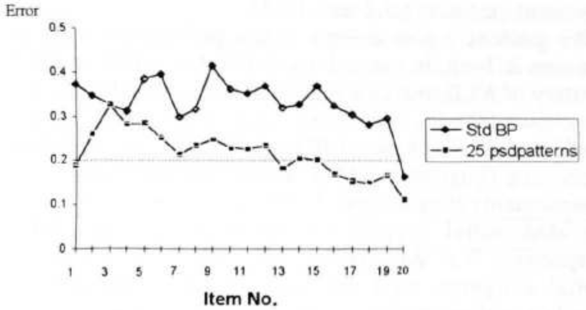
The “pseudo-recurrent” architecture

This architecture consists of a feedforward backpropagation network that is divided into two parts (“early-processing” and “final-storage”), each interactively training the other. When a new pattern from the environment is presented to the early-processing system, this will cause a number of “pseudopatterns” to be generated by the final-storage system. These pseudopatterns from final-storage reflect the network's prior learning. They are learned along with the new pattern. In short, pseudopatterns stand in for the patterns originally learned by the network. In this way, the crucial interleaving of new information with (an approximation of) old information is achieved. Once the early-processing system has learned the new pattern along with the set of pseudopatterns, this information is transferred to final-storage either by copying the early-storage weights to final-storage (assuming identical network topologies) or by means of pseudopatterns generated this time by the early-processing area and learned by final storage.

Experiments

In tests of the network, using both Ebbinghaus-like savings measures and tests of exact recognition of previously learned information, this type of network was shown to effectively overcome the problem of catastrophic interference, thus allowing effective sequential learning with gradual (not catastrophic) forgetting.

In one experiment, 20 patterns were presented *sequentially* to the network. A new pattern was presented only after the previous one had been learned to criterion. After the 20th item had been learned, all 20 were tested. In the pseudo-recurrent network the final 8 remained at or below criterion, 9 more within 0.05 of criterion. For standard backpropagation, however, *none* of the previously learned items are within 0.1 of criterion. Clearly, forgetting occurs more gradually in the pseudo-recurrent network than in standard BP (see figure below).



In addition, the mechanism of pseudopattern transfer automatically produces sparse internal representations in final storage, the more pseudopatterns, the sparser the representations. Compact representations are more efficient to process and less likely to cause catastrophic interference, but are more susceptible to damage. This could provide an underlying reason for selective category memory losses in elderly amnesiacs (French, 1996).

References

French, R. M. (1996) Pseudo-recurrent connectionist networks: An approach to the “sensitivity–stability” dilemma. ULG Psych. Dept. TR (under review).

McClelland, J., McNaughton, B., & O'Reilly, R. (1995), Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*. 102, 419–457.

Robins, A. (1995) Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, 7(2), 123–146.