Noise, Pseudopatterns, and Information Transfer in the Brain

Robert M. French

Quantitative Psychology & Cognitive Science
Psychology Department
University of Liège,
4000 Liège, Belgium
rfrench@ulg.ac.be

Nick Chater

Institute for Applied Cognitive Science
Department of Psychology
University of Warwick
Coventry, CV4 7AL, UK
nick.chater@warwick.ac.uk

1. Overview

Connectionist learning methods for distributed feedforward networks typically result in "catastrophic forgetting." That is, as new information is learned, old information is rapidly obliterated. This contrasts with the much more gradual forgetting observed in human memory, thus casting doubt on the cognitive plausibility of connectionist learning methods. It has been shown previously that paired neural networks that are "cross-trained" in a rather intricate manner can address this problem (French, 1997; Ans & Rousset, 1997). Here, we show that a simpler single-network approach that makes use only of noise passing through the network can also significantly reduce catastrophic interference. We speculate that this kind of method might be involved in human learning.

2. The Hessian pseudopattern backpropagation algorithm

When a neural network learns (perfectly) a set of patterns, $\{P_i : I_i \to O_i\}_{i=1}^{N}$, this defines a unique error surface, $E(w)$, with respect to the weights of the network. Learning the set of patterns means that the network has found a local minimum w_0 in weight-space for which $E(w_0) = 0$ and $E'(w_0) = 0$ where $E'(w)$ represents the first derivative of the error function. The problem is that when new patterns are subsequently learned by the network, a new error surface is created. If the previously learned patterns have not been interleaved with the new patterns, the point in weight-space corresponding to a minimum of the error surface associated with the new patterns may not correspond at all to an error minimum for the previously learned patterns. Hence, catastrophic forgetting of the old patterns.

Now assume that the original patterns $\{P_i\}_{i=1}^{N}$ are no longer available for interleaving with the new patterns, but that we would nonetheless like to approximate the original error surface $E(w)$. If the function *f* underlying the original set of patterns is relatively "nice" (i.e., continuous, reasonably smooth, etc.), then by generating a set of pseudopatterns $\{\psi_i : \hat{I}_i \to \hat{O}_i\}_{i=1}^{M}$ whose input values are drawn from a random distribution and whose associated output is simply the result of passing the random input through the network, we can produce a reasonable approximation of *f* (Robins, 1995). Just as the original set of patterns $\{P_i\}_{i=1}^{N}$ had a unique error surface associated with it, the same is true for the set of pseudopatterns $\{\psi_i\}_{i=1}^{M}$. We will call this latter error surface $\hat{E}(w)$. It follows from the definition of pseudopatterns that $\hat{E}(w_0) = 0$ and $\hat{E}'(w_0) = 0$. The question is how to develop an approximation of this error surface in the vicinity of w_0 without having recourse to the original patterns. We develop a Taylor series expansion of $\hat{E}(w)$, which requires the second derivative of $\hat{E}(w)$ (i.e., the Hessian), but, unlike the first derivative, *the Hessian does not disappear when evaluated at* w_0. Once we have the second derivative, we immediately obtain $\hat{E}(w)$, the desired approximation of the original error surface.

Somewhat counter-intuitively, approximating the original error surface using noise (i.e., randomly chosen inputs fed through the network) *does not require any explicit error information*; noise moving through the system is sufficient to approximate this error surface. When combined with the error surface of the new patterns to be learned, an overall error surface is produced on which gradient descent will be performed. Using this combined error surface, instead of the error surface for the new patterns alone, significantly reduces catastrophic interference.

We show by means of a neural network simulation that this technique is, indeed, effective in reducing catastrophic interference.

References

Ans, B. and Rousset, S. (1997) Avoiding catastrophic forgetting by coupling two reverberating neural networks. *Academie des Sciences, Sciences de la vie*, 320, 989 - 997

French, R. M. (1997) Pseudo-recurrent connectionist networks: An approach to the “sensitivity–stability” dilemma. *Connection Science, 9*, 353-379.

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