Mitigation of Catastrophic Forgetting in Recurrent Neural Networks using a Fixed Expansion Layer
Robert Coop and Itamar Arel

Catastrophic Interference
Example – Binary Pattern Learning
- Standard feed-forward neural network used to learn auto-associative 32-bit binary patterns (using 16 hidden neurons)
- 16 patterns learned to accuracy of at least 85%
- New interfering pattern learned, then accuracy re-tested

How much information is lost when a new pattern is learned?

FEL Operation – Feedforward
Operation
- Weight Initialization
- Feedforward, FEL neuron triggering
- Backpropagation

Feature-sign search algorithm
- Efficiently finds sparse coefficients given an input signal and a set of fixed basis vectors
- Solves $\frac{1}{2}\|y - Ax\|^2 + \gamma \|x\|_1$ where $x$ is hidden layer activation, $A$ is FEL weight matrix, $\gamma$ is a penalty constant

Auto-associative binary pattern reconstruction
32-bit binary patterns, target is input to input 20 patterns learned until 90% accuracy

Non-stationary MNIST sequence classification
Networks have 64 hidden neurons

Example of Forgetting Effect
- Original accuracy for batch: 99.6%
- Loss
- After learning 1 pattern: 67.9%, 32.7%
- After learning 6 patterns: 56.9%, 43.1%

FEL in Elman Recurrent Networks
Recurrent FEL Network
- Elman RNN retains context
- Values of hidden neurons at time $t$ fed into network as input at time $t+1$
- Expansion layer ‘inserted’ between context and input

Auto-associative binary sequence reconstruction
20 sequences consisting of 25 32-bit binary patterns Goal is to reconstruct final part of sequence after observing complete sequence sequences learned with 90% accuracy

FEL Neuron Affinity
FEL neurons expected to ‘ latch’ onto input
- Measured affinity of neurons
- During MNIST sequence
- Recorded input for neurons
- Input weighted by neuron value
- Neurons exhibit latching
- Show positive response to digit 1 at same time as negative response to digit 2

Fixed Expansion Layer Network
The Fixed Expansion Layer (FEL) Feedforward Neural Network
- Motivation
- Hidden layer of network contains a dense signal – transforms that signal into a sparse space in the Fixed Expansion Layer
- Sparsity of FEL prevents network weight convergence and mitigates catastrophic forgetting

Non-stationary Gaussian distribution classification
2 dimensional real input (4 classes as output)
Networks have 20 hidden neurons
50,000 training iterations
1,000 testing iterations

Non-stationary MNIST classification
128 dimension input
4 classes as output
FEL network: 80 hidden, 512 FEL
Standard network: 784 hidden

Publications
R. Coop and I. Arel. Mitigation of Catastrophic Forgetting in Recurrent Neural Networks using a Fixed Expansion Layer. In Neural Networks (IJCNN), The 2013 International Joint Conference on, accepted for publication, August 2013.