Outline

Theoretical Background
- Overview
- Overview II
- Basic Algorithm

Comparisons
- Two Problems with Backpropagation
- Pros & Cons

Applications

- Feed-forward, supervised learning
- Uses backpropagation algorithms
- coNN - constructive neural network
  - 2 main categories:
    - evolutionary based (what is done at Brock)
    - generally constructive. CCNN is main exemplar of this group
- creates it's own topology starting with minimal network
  - input and output layers only, as usual connected by weights.
- input of nodes code the problem being presented to the network
- output of nodes code the network's response to the input problem
- typically quickprop is used as the learning rule

Step 1 Training input weights

A new hidden neuron is added one at a time Instead of a minimum, we find a maximum: the maximum correlation between the candidate output and the residual error of the networks output.

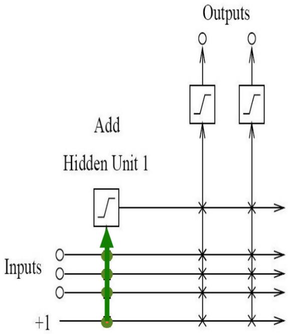
1. generate a population of candidate nodes with randomized input weights
2. inputs are connected, but not outputs
3. repeat training steps until no correlation improvement
   3.1 one epoch of the training data is run through
   3.2 update input weights using any learning rule, such that correlation is increased

Train input weights - correlation

S is sum over all output units o of correlation with error

$$S = \sum_{o} \left| \sum_{p} (V_p - V)(E_{p,o} - E_o) \right|$$

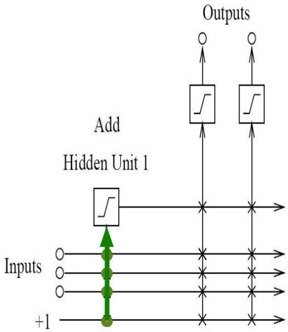
o	network output at which the error is measured
p	the training pattern
σ	network output
V_p	candidate output for input pattern p
$E_{p,o}$	network output error for output o, pattern p
$\overline{V}$	average of candidate output over all patterns
$\overline{E_o}$	average of output errors over all patterns



Train input weights - correlation

$$\frac{\delta S}{\delta w_i} = \sum_{p,o} \sigma_o (E_{p,o} - \overline{E_o}) f_p{}' I_{i,p}$$

w_i	weight to be updated
σ_o	sign of correlation
$f_p{}'$	derivative for pattern p of the nodes fn wrt of inputs
$\mathrm{I}_{\mathrm{i,p}}$	input node gets from unit i for input p
$\mathrm{E}_{\mathrm{p,o}}$	network output error for output o, pattern p
$\overline{E_o}$	average of output errors over all patterns

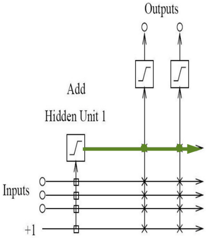


Step 1 Training input weights

- candidate with highest correlation is put in the network, others are discarded.
- correlation sign doesn't matter, only magnitude
- the bias unit effectively implements a learnable resting activation level for each hidden and output unit.
- subsequent hidden neurons are attached to previous hidden neurons - this is where the cascade term comes from.

Step 2 Training output weights

1. connect output of new node to output layer nodes use randomized weights that have adjusted sign to reduce error
2. input weights are fixed
3. only output weights are trained using any learning rule, until no improvement in error reduction
4. once done the new node is fixed permanently.
5. any new training involves adding another new node on it's own downstream layer



The Step Size Problem

- vanilla backpropagation requires small steps for convergence - slow
- we do not have the information to pick an optimal learning rate, manually selected

The Moving Target Problem

- complication when many factors changing at the same time
- error signal defines problem unit trying to solve, but this keeps changing
- dramatic slowdown of training with increasing number of hidden layers
- herd effect:
  - 2 tasks A, B. If A has bigger effect, all nodes redundantly train for A, ignoring B
  - But when all nodes move toward B at once, problem A response becomes worse.
  - eventually nodes split to train for separate problems A and B, but it takes a long time
  - a randomly initialized network prevents nodes from behaving identically, but this tends to dissipate as the network is trained
- One way to combat: allow only a few weights to change while keeping the others constant

Pros

- At elast 10 times faster than standard backpropagation
  - [performanceChart2.png]
- The network determines its own size and topologies
- Incremental/life long learning: new training, new information can be added, with an already trained network
- effectively deals with the step size and moving target problems

Cons

- Very susceptible to overfitting

Applications

In spite of the many CoNN algorithms surveyed in (Kwok & Yeung, 1997a), the most popular for regression problems is no doubt the Cascade Correlation algorithm and maybe the second most popular is the DNC... The popularity of Cascade Correlation can be attested by the various ways this algorithm has inspired new variations and also has been used in the combined approaches between learning methods. [Franco et al, 2009]

References I

- S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture."
- K. Khatter and J. Kaur, "Global Journal of Engineering Science and Research Management."
- S. K. Sharma and P. Chandra, "CONSTRUCTIVE NEURAL NETWORKS: A REVIEW," International Journal of Engineering Science and Technology, vol. 1, no. 2, pp. 7847–7855.
- T.-Y. Kwok and D.-Y. Yeung, "Constructive algorithms for structure learning in feedforward neural networks for regression problems," IEEE Transactions on Neural Networks, vol. 8, no. 3, pp. 630–645, 1997.
- G. Balázs, "Cascade-Correlation Neural Networks: A Survey."
- Y. Guo, L. Bai, S. Lao, S. Wu, and M. S. Lew, "A Comparison between Artificial Neural Network and Cascade-Correlation Neural Network in Concept Classification," in Advances in Multimedia Information Processing – PCM 2014, 2014, pp. 248–253.
- B. K. Wong, T. A. Bodnovich, and V. S.-K. Lai, "The Use of Cascade-Correlation Neural Networks in University Fund Raising," The Journal of the Operational Research Society, vol. 51, no. 8, pp. 913–920, 2000.
- A. B. Nassif, L. F. Capretz, and D. Ho, "Software Effort Estimation in the Early Stages of the Software Life Cycle Using a Cascade Correlation Neural Network Model," in 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2012, pp. 589–594.
- S. Saha, A. Konar, A. Saha, A. K. Sadhu, B. Banerjee, and A. K. Nagar, "EEG based gesture mimicking by an artificial limb using cascade-correlation learning architecture," in 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 4680–4687.

References II

- P. Blonda, G. Pasquariello, and J. Smid, "Comparison of backpropagation, cascade-correlation and Kokonen algorithms for cloud retrieval," in Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan), 1993, vol. 2, pp. 1231–1234 vol.2.
- N. Sharma and H. Om, "Cascade correlation neural network model for classification of oral cancer," WSEAS Transactions on Biology and Biomedicine, vol. 11, pp. 45–51, 2014.
- B. Chandra and P. P. Varghese, "Applications of Cascade Correlation Neural Networks for Cipher System Identification," World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 1, no. 2, pp. 369–372, 2007.
- N. A. Singh and T. Naryanan, "Application of Cascaded Correlation Neural Network for Financial Performance Prediction and Analysis of BSNL," in Swarm, Evolutionary, and Memetic Computing, 2014, pp. 502–513.