# Controlling the Structure of Neural Networks that Grow and Shrink

Norbert Jankowski

Department of Computer Methods
Nicholas Copernicus University
Torun, Poland, norbert@phys.uni.torun.pl

**Abstract:** The main subject of my research is the automatic control of the neural network structure. It is crucial for the learning process that the structure (number of neurons and weights) is adequate to the complexity of a given problem, especially if the structure of the network can change during the adaptive process while new data is presented. Networks that can change their structure learn faster and have better generalization ability. In this paper we are going to concentrate on neural networks estimating the probability density using localized functions (for example RBF networks).

Many pruning methods were described in the last decade, but pruning leads to the removal of the network connections and unnecessary neurons, but frequently many neurons contribute to decision borders that could be represented by smaller network without decreasing of accuracy. Therefore one should merge two (or even more – it may be more complicated computationally) neurons keeping the current shape of the decision surface unchanged as much as possible. Two neurons can be replaced by another one with if the ratio:

$$\frac{\int_{\mathbf{d}\subseteq\mathcal{D}^n} |\phi_i(\mathbf{x}) + \phi_j(\mathbf{x}) - \phi_{new}(\mathbf{x})| \, d\mathbf{x}}{\int_{\mathbf{d}\subseteq\mathcal{D}^n} |\phi_i(\mathbf{x}) + \phi_j(\mathbf{x})| \, d\mathbf{x}} < \alpha \tag{1}$$

is smaller than some confidence parameter $\alpha$. Here $d$ is the subspace in which localized neuron transfer functions (scaled by the networks weights) $\phi_i(\mathbf{x})$ and $\phi_j(\mathbf{x})$ have values greater than a small threshold and $\phi_{new}(\mathbf{x})$ is the new neuron whose transfer function replaces the combination of neurons $i$ and $j$.

Assuming that $i$ and $j$, two candidate neurons for merging, were found, and that it is possible to pre-initialize density function parameters of the neuron $new$, the above criterion can be simplified through sampling the space around neurons $i$ and $j$ (using adequate distribution function for the density of neurons $i$ and $j$) and computing *weighted* mean squared error for a given number of points.

$$\frac{\sum_{d\in\mathbf{d}} \left(\phi_i(\mathbf{x}) + \phi_j(\mathbf{x}) - \phi_{new}(\mathbf{x})\right)^2}{\sum_{d\in\mathbf{d}} \left(\phi_i(\mathbf{x}) + \phi_j(\mathbf{x})\right)^2} < \alpha \tag{2}$$

To speed up calculations *incremental checking* is introduced: the inequality Eq. 2 is checked in each iteration using $\alpha$ that has initially large value, reduced in each iteration to the final, small value. The checking-phase should be continued as long as the inequality Eq. 2 is satisfied, otherwise there is no chance for merging and the learning process can start again.

The *Incremental Network* uses statistical information obtained from the Extended Kalman Filter learning algorithm to grow and prune its structure. Two criteria are defined to answer the questions whether a neuron should be added and/or an unnecessary neuron exists. Methods that can use pruning and growing in simultaneously the learning process are still very rare. Incremental network has been applied to classification of medical data and to approximation of functions, obtaining highly accurate results.

Another research goal was to find transfer functions which are dimensionally separable, with independent biases and slopes in each dimension, and are able to rotate in multidimentional space without using full covariance matrix. One solution to this problem is based on the product form of the combination of sigmoids:

$$C_P(\mathbf{x};\mathbf{t},\mathbf{t}',\mathbf{R}) = \prod_i \left(\sigma(\mathbf{R}_i\mathbf{x} + t_i) - \sigma(\mathbf{R}_i\mathbf{x} + t'_i)\right) \tag{3}$$

where $\mathbf{R}_i$ matrix has slopes on the diagonal ($N$ parameters) and only one non-zero off-diagonal value ($N-1$ rotation parameters). Such function provides very flexible decision borders using relatively small number of parameters.

# Controlling the Structure of Neural Networks that Grow and Shrink


Norbert Jankowski

Presenting author name: Norbert Jankowski

Department of Computer Methods
Nicholas Copernicus University
in Torun, Poland
ul. Grudziadzka 5, 87 – 100 Torun
phone: +48 56 21065, fax: +48 56 21543
norbert@phys.uni.torun.pl


Type of presentation: poster

Classification: T – supervised learning, neural system models