A Two-Layer Learning Method for Radial Basis Function Networks Using Combined Genetic and Regularised OLS Algorithms

S Chen, Y Wu and K Alkadhimi

University of Portsmouth, UK

email: schen@ee.port.ac.uk

Abstract

The paper presents a novel two-layer learning method for radial basis function (RBF) networks. At the lower layer, a regularised orthogonal least squares (ROLS) algorithm is employed to construct RBF networks while the two key learning parameters, the regularisation parameter and hidden node's width, needed by the ROLS algorithm are optimized using the genetic algorithm at the higher layer. Networks constructed by this learning method have superior generalisation properties, and the computational complexity of the method is reasonable. Nonlinear time series modelling and prediction is used as an example to demonstrate the effectiveness of this hierarchical learning approach.

1 Introduction

The genetic algorithm (GA) [1] is a powerful nonlinear optimisation technique and has attracted considerable attention of the neural network community. A key advantage of using the GA as a neural network learning method is that it is capable of achieving optimal or near optimal network parameter settings under given network structure and training conditions. This is however obtained at the cost of extensive computational requirements.

Simpler learning can often be achieved if a neural network has a linear-in-the-parameters structure. When the width parameter is fixed, a RBF network has such a structure and an orthogonal least squares (OLS) algorithm [2] has been developed for constructing parsimonious RBF networks. A well constructed small neural network often has better generalisation properties compared with a large full-size neural network model.

If training data are highly noisy, the parsimonious principle alone may not be sufficient to guarantee good generalisation performance. By combining the parsimonious principle with the regularisation method [3], a ROLS algorithm has been derived [4], which has superior generalisation properties under severely noisy conditions. A good regularisation parameter required by the ROLS algorithm can be obtained by iterations using a Bayesian formula [5]. The regularisation parameter so generated however may not necessarily be the best one as will be demonstrated later.

We propose a two-level learning hierarchy for constructing RBF networks based on the combined GA and ROLS algorithms. Because the generalisation performance is a complex multimodal function on the space of the width and regularisation parameter, these two key parameters are optimised using the GA at the upper level. Given these two parameters, the ROLS

Genetic Algorithms in Engineering Systems: Innovations and Applications
12-14 September 1995, Conference Publication No. 414, © IEE, 1995

algorithm is employed to construct parsimonious RBF networks at the lower level. Since the GA only optimises two parameters and the lower layer only involves linear learning problems, the computational complexity of this combined approach is much less than that of using the GA to learn all the network parameters. RBF networks produced by this learning hierarchy have significantly better generalisation performance as is demonstrated by the included examples of nonlinear time series modelling and prediction.

The RBF network considered in this paper has a Gaussian nonlinearity with a uniform width ρ, and the network output is defined by

$$F_r(\mathbf{x}(k)) = \sum_{i=1}^{n} w_i \exp(-\|\mathbf{x}(k)-\mathbf{c}_i\|^2/\rho) \quad (1)$$

where w_i are the weights and $\mathbf{c}_i$ are the centre vectors. The approach developed however is not restricted to this particular Gaussian RBF (GRBF) network.

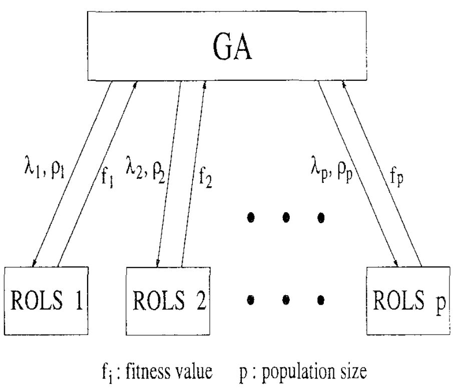

Figure 1: Schematic of two-layer learning hierarchy for RBF networks.

2 The learning scheme

Fig.1 illustrates the proposed two-layer learning scheme. At the upper layer, the GA, with a population size of p, learns the width ρ and the regularisation parameter λ based on the fitness function f values provided by the lower layer. The lower layer consists of the p parallel ROLS algorithms. The data set is divided into a training set and a testing set. Given ρ_i and λ_i, the ith ROLS algorithm constructs a RBF network. The generalisation performance, the mean square error (MSE) over the testing set, of the resulting RBF network is computed. The inverse of this generalisation performance is the fitness function value f_i for the given ρ_i and λ_i.

The ROLS algorithm [4] uses a forward regression procedure to construct parsimonious RBF networks based on the following zero-order regularised error criterion

$$J_R = \mathbf{e}^T\mathbf{e} + \lambda \mathbf{g}^T\mathbf{g} \quad (2)$$

where $\mathbf{e}$ is the error vector between the desired outputs and the network outputs, and $\mathbf{g}$ is the orthogonal weight vector. The implementation of this algorithm is similar to that of the OLS algorithm [2]. The simple zero-order regularisation employed significantly improves the generalisation properties of the constructed network model.

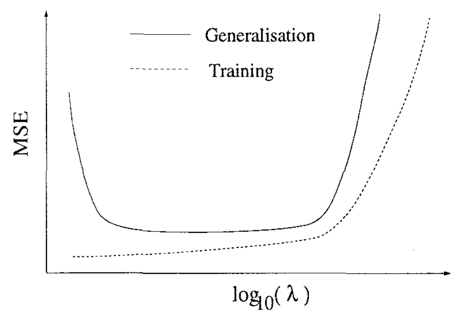

Figure 2: Mean square error as a function of regularisation parameter.

It was often observed that regularised learning exhibits the characteristics of Fig.2 [6,7]. This appears to suggest that generalisation performance curve may have a flat region. A good λ value can be obtained using the following iterative procedure: Given an initial guess of λ, the learning algorithm constructs a RBF network. This in turn allows an updating of λ using the evidence formula

$$\lambda = \frac{\gamma}{N-\gamma} \frac{\mathbf{e}^T\mathbf{e}}{\mathbf{g}^T\mathbf{g}} \quad (3)$$

where γ is the number of good parameter measurements [5], and N is the number of training data.

It should be emphasized that the above Bayesian procedure in general can only obtain a local optimal value of λ and Fig.2 does not provide a complete picture. In fact, the generalisation performance or fitness function f is a highly complicated multimodal function on the space of ρ and λ. The characteristics of Fig.2 may only be obtained under a particular value of ρ. We use a simple example to demonstrate these points.

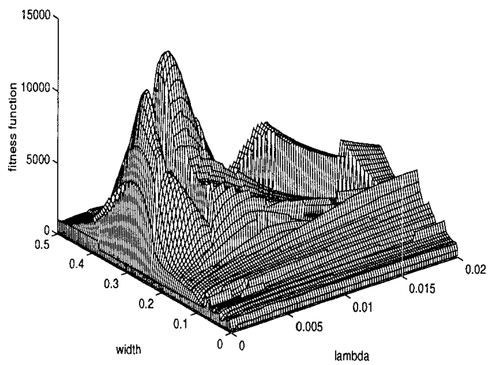

Figure 3: Generalisation performance surface on space of ρ and λ.

Consider the modelling of the scalar function

$$F_s(x) = \sin(2\pi x), \quad 0 \le x \le 1 \tag{4}$$

by a GRBF network. The training data was generated from $F_s(x) + e$, where the noise had a Gaussian distribution with zero mean and variance 0.02 and x was taken from the uniform distribution in (0, 1). The training data had a signal-to-noise ratio (SNR) of 14 dB. Given values of ρ and λ, the ROLS algorithm constructed GRBF networks. The learning procedure was terminated when the regularised error reduction ratio [7] was smaller than a preset threshold. GRBF networks constructed had 5 to 7 nodes depending on values of ρ and λ. The MSE between the noise-free system output $F_s(x)$ and the network response $F_r(x)$ was computed. The inverse of this generalisation performance as the function of ρ and λ is plotted in Fig.3.

Even for such a simple example, the complexity of the generalisation performance surface is apparent. A local optimal method cannot in general learn the global optimal values of ρ and λ. Furthermore, performance improvement by achieving the global optimum is very significant. The choice of the GA as the upper level learning method is therefore well justified. The lower level of the learning hierarchy consists of p parallel linear learning problems. A population size of $p = 100$ is usually adequate. The overall computational complexity of this learning approach is within the computing power of a standard PC or workstation.

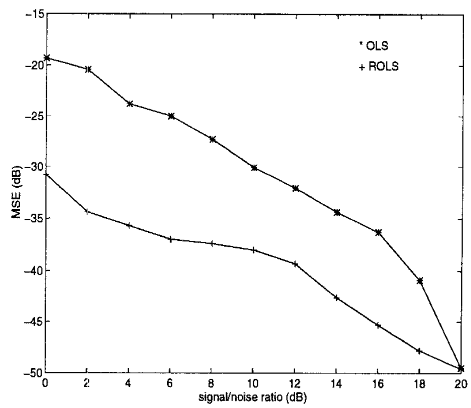

Figure 4: Generalisation performance with and without regularisation for example 1.

3 Examples

Example 1. This is the simple example of modelling the scalar function used to generate Fig.3. To demonstrate superior generalisation properties of regularised learning under severely noisy condition, we used the two-layer learning scheme to construct GRBF networks for different SNR conditions with and without regularisation. For the case of no regularisation, the lower level employed the OLS algorithm and the upper level only learnt ρ. Fig.4 depicts the generalisation performance, the MSE between $F_s(x)$ and $F_r(x)$, for these two cases. It can be seen that the simple regularisation technique employed has superior generalisation performance under highly noisy training conditions. For the training data of SNR=14 dB, it was observed that the optimal $\rho_{opt} = 0.27$ and $\lambda_{opt} = 2.5 \times 10^{-3}$ (corresponding to the highest peak $f_{opt} = 1.5 \times 10^4$ in Fig.3) was achieved by the combined GA and ROLS learning.

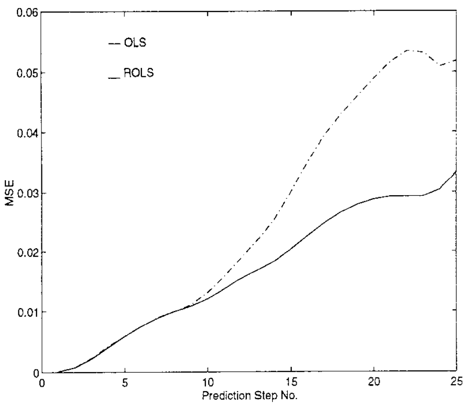

Figure 5: Multi-step prediction performance for Mackey-Glass time series.

Example 2. The second example was Mackey-Glass time series prediction. To make the task more realistic, small noise was added to the time series samples, giving rise to a SNR of 40 dB. A data set of 1000 samples were generated with the first 500 samples used as the training set and the last 500 samples as the testing set. The GRBF predictors were constructed with and without regularisation. Again in the case of no regularisation, the upper level only learnt ρ. In the both cases, the constructed GRBF networks had 25 centres. The multi-step-ahead prediction accuracies over the testing set were then computed and the results were plotted in Fig.5. From Fig.5, it can be seen that better generalisation performance was achieved with regularisation.

Example 3. This example was sunspot time series prediction based on the 280 sunspot observations over the years 1700-1979. The observations of 1761–1979 were used as the training set and the observations of 1700-1760 were used as the testing set. Two GRBF models of 25 centres were constructed using the learning hierarchy with and without regularisation. The normalized multi-step-ahead prediction accuracies of the two resulting models over the testing set are plotted in Fig.6. Figs.7 and 8 compare the 5-step-ahead predictions obtained by the two models while Figs.9 and 10 show the 11-step-step-ahead predictions of the two models.

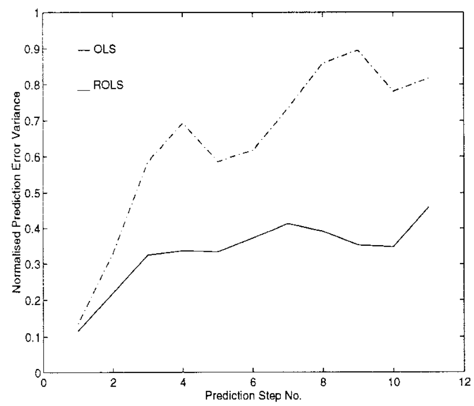

Figure 6: Normalized prediction performance for sunspot time series.

4 Conclusions

A novel two-level learning hierarchy has been developed for RBF networks by combining the GA with the ROLS learning. The GA at the upper layer finds the global optimum of the width and regularisation parameters while the ROLS algorithm at the lower layer automatically constructs RBF networks. The method is computationally more efficient compared with using the GA to directly learn all the network parameters. Time series modelling and prediction has been used to illustrate superior generalisation properties of the proposed learning method.

5 References

[1] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA.: Addison-Wesley, 1989.

[2] S. Chen, C.F.N. Cowan and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE*

Trans. Neural Networks, Vol.2, No.2, pp.302–309, 1991.

[3] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, Vol.78, No.9, pp.1481–1497, 1990.

[4] S. Chen, "Regularised OLS algorithm with fast implementation for training multi-output radial basis function networks," to be presented at *IEE 4th Int. Conf. Artificial Neural Networks*, Cambridge, 26–28 June, 1995.

[5] D.J.C. MacKay, "Bayesian interpolation," *Neural Computation*, Vol.4, No.3, pp.415–447, 1992.

[6] C. Bishop, "Improving the generalization properties of radial basis function neural networks," *Neural Computation*, Vol.3, No.4, pp.579–588, 1991.

[7] S. Chen, E.S. Chng and K. Alkadhimi, "Regularised orthogonal least squares algorithm for constructing radial basis function networks," submitted to *Int. J. Control*, 1994.

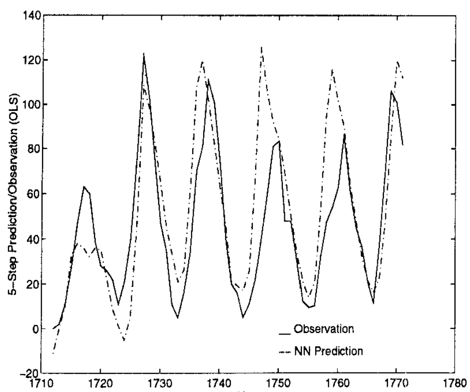

Figure 7: 5-step-ahead prediction using model obtained without regularisation.

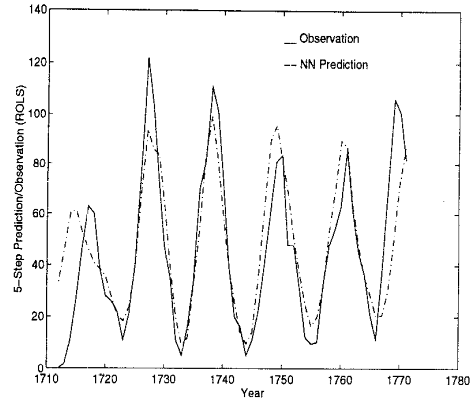

Figure 8: 5-step-ahead prediction using model obtained with regularisation.

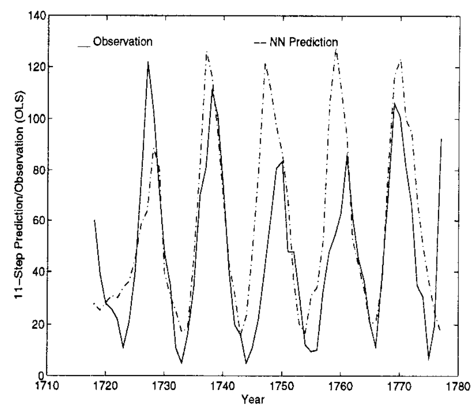

Figure 9: 11-step-ahead prediction using model obtained without regularisation.

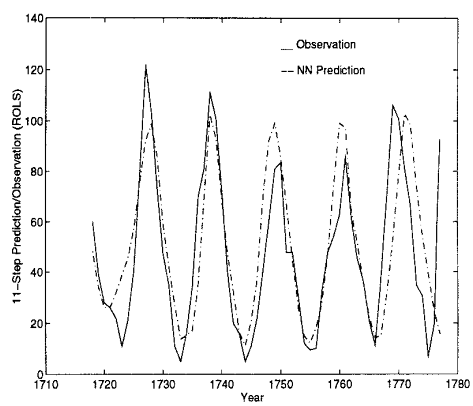

Figure 10: 11-step-ahead prediction using model obtained with regularisation.