ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 1-6

What are the main factors involved in the design of a Radial Basis Function Network ?

I.Rojas[1], M.Anguita[1], E.Ros[1], H.Pomares[1], O.Valenzuela[2], A.Prieto[1]
1 Departamento de Electrónica y Tecnología de Computadores
2 Departamento de Estadística e Investigación Operativa
Universidad de Granada. Spain.

Abstract

Many papers have been published concerning the obtention and analysis of new algorithms for neural network learning, and in particular the paramenter optimization of a Radial Basis Function network system (RBF). However, there has been no analysis in the influence in RBF design for the selection of the following factors: which of the many avalible functions to select for the non-linear functions within the hidden neurons; selection of distances; number of neurons in the hidden layer; number of bits assigned to weight resolution, etc. The present statistical study was motivated by the great variety of alternatives that a designer has to take into account when developing an RBF. Thus, complementing the existing intuitive knowledge, it is necessary to have a more precise understanding of the significance of the different alternatives. In the present contribution, the relevance and relative importance of the parameters involved in such the design are investigated by using a statistical tool, the ANalysis Of the VAriance (ANOVA)[2]. In order to analyze the behaviour of the RBF, three different groups of examples were used: time series forecasting, approximation of functions from samples and classification problems.

1. Introduction

As there are many possibilities for the set of basic functions, parameters and operators used in the design of an RBF and in general in Artificial Neural Networks, the search for the most suitable operators and functional blocks, together with their characterization and evaluation, is an important topic in the field of Neural Network design. As shown in papers dealing with real applications, the designer has to select the operator to be used in each phase in the design of an RBF, and this decision is usually taken in terms of the most common operations performed. Nevertheless, it is very important to know which stages have the greatest influence on the behaviour and performance of the neural system. Therefore, the designer should pay more attention to the phase in which the selection of the operator is most statistically significant. In this way, it would be possible to obviate a detailed analysis of different configurations that lead to systems with very similar performance.

Much research has been carried out on different applications and learning algorithms [1],[6],[7],[12]. However, to the best of our knowledge, the consequences of selecting different functional blocks have not been examined in detail. Consequently, the goal of this paper is to get a better insight into which are the most relevant blocks in an RBF design, in order to establish the elemental operations whose alternatives should be carefully studied when a real application is developed. The main factors considered are: the type of RBF used (the output is the weighted sum or is the weighted average of the radial basis function), the distance used, the number of neurons in the hidden layer, the different alternatives for the nonlinear function in the hidden layer, the number of bits used for weight storage and the noise amplitude used for corrupting the input pattern. To do this, an appropriate statistical tool has been used: the multifactorial analysis of the variance [2], which consists of a set of statistical techniques that allow the analysis and comparison of experiments, by describing the interactions and interrelations between either the quantitative or qualitative variables (called factors in this context) of the system.

2. Main factors considered in the design of an RBF

An unknown multidimensional function F(X): $X^n \rightarrow \mathcal{L}$, where $X^n \in \mathbb{R}^n$ and $\mathcal{L} \in \mathbb{R}$, can be approximated by the weights (w_i) and the radial function $\Phi_i(X)$ in two similar ways. In the first one, the output is the weighted sum of the radial basis function ($\widetilde{F}_{RBF}$), while the alternative is to calculate the weighted average ($\widetilde{F}^*_{RBF}$), with the addition of lateral connections between the radial neurons. The use of the second method has been presented in different studies [6] as an approach which, due to its normalization properties, is very convenient and provides better performance than the weighted sum method for function approximation problems. The weighted average has the disadvantage of a higher degree of computational complexity compared to the weighted sum, due to the need for output normalization. The non-linear function of the hidden neuron is expressed as: $\Phi_i(x)=\Phi(\|x-c_i\|/d_i)$, where $c_i \in R^n$ is the centre of basis function Φ_i, $d_i \in R$ is a dilation or scaling factor for the radius $\|x-c_i\|$, and $\|\|$ is typically the Euclidean norm on R^n, but we will use the Minkowski distance, expressed as $D(u,v) =(\sum|u_i - v_i|^{\lambda})^{1/\lambda}$. Choices of Φ considered in theoretical research and practical applications can be found [1],[11],[12]. Also we will consider different numbers of neurons in the hidden layer, the number of bits used for weight storage and the noise amplitude used for corrupting the input pattern.

3. Application of ANOVA in the design of an RBFN

The analysis of variance (commonly referred to as ANOVA) is one of the most widely used statistical techniques [2]. ANOVA examines the effects of one, two or more quantitative or qualitative variables (termed factors) on one quantitative response. ANOVA is useful in a range of disciplines when it is suspected that one or more factors affect a response. In its simplest form, ANOVA is a method of estimating the means of several populations, assumed to be normally distributed and independent, all having the same variance and as an initial approximation, with averages that can be expressed as a linear combination of certain unknown parameters. ANOVA allows us to determine whether a change in the measure of a given variable is caused by a change in the level of a factor, or is originated by a random effect.

To decide whether the levels of some factors affect the response of the system in a different way, it is necessary to define the **F-ratio** test statistic. If the F-ratio is greater than the F-Snedecor distribution (for defined degrees of freedom) with a sufficiently high confidence level (usually 95%), means that there are levels of some factors that affect the response of the system in a different way. The comparison between

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 1-6

the F-ratio and the F-Snedecor distribution is expressed through the significance level (**Sig. Level**). If this significance level is lower than 0.05 then the corresponding levels of the factor are statistically significant with a confidence level of 95%. Thus, this is the statistical parameter that we will mainly consider in Section 5 to derive our conclusions about the different factors within the fuzzy inference process. Table 1 presents the levels of the factors used.

Table 1 Levels of each factor considered in the statistical analysis

	Structure of the RBF	Non-linear function	Distance	Neurons	Bits	Noise
Level 1	Weighted sum	Gaussian	Manhattan	18	6	0 %
Level 2	Weighted average	Epanechnikov	Supremum	19	7	2%
Level 3	---	Triangular	Euclidean	20	8	4 %
Level 4	—	Cubic	Minkowski (λ=3)	22	10	6 %
Level 5	—	Inverse multiquadratic	—	24	12	8 %
Level 6	—	Thin plate spline	—	—	—	—

4. Illustrative examples

4.1 Time series forecasting

We will use time series generated from a differential of difference equation governed by determinism (in which, once the initial value is given, the subsequent states are all determined) . This is the deterministic chaos of a dynamic system. The Mackey-Glass time serie, the Lorenz system, the logistic map, and the quadratic map (originally proposed for biological system modeling) are considered. Prediction of these time series is recognized as a benchmark for testing various neural-network architectures [3],[5],[6],[9],[10]. Other real time series from natural and social phenomena, for example, sunspot, seasonal change in atmospheric temperature, electric power demand, traffic density, etc, obtained from [8] are also considered. For example, Fig.1 illustrates the sunspot time series [9].

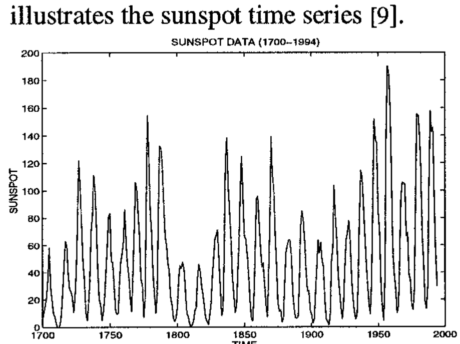

Fig. 1 Sunspot data (yearly from 1700-1994)

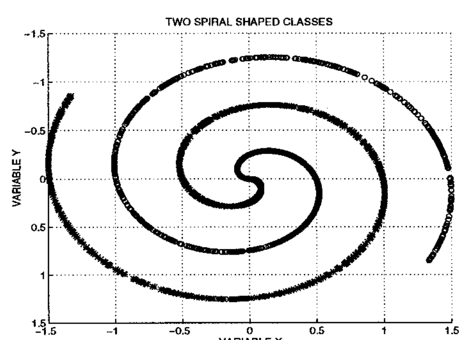

Fig. 2 Two spiral classes used for learning an RBF

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 1-6

4.2 Classification problems

We will analyze two different problems: two-spiral classification and character recognition. In the two-spiral benchmark the task of the tested data classifier is to separate a rectangular fragment of a two dimensional space into disjoint areas, each containing a different spiral curve (Fig.2). For character recognition, a network must be designed and trained to recognize 26 different letters. Each letter is represented as a 5 by 7 grid of integer values. Therefore, the neural network designed receives a 35-element input vector and it will then be required to identify a letter by responding with a 26-element output vector. Each of the 26 neurons of the output represents a letter. To operate correctly, the network should respond with a 1 in the position of the letter being represented to the network and all other values should be equal to zero.

4.3 Function approximation

The problem of function approximation from a finite number of data points has been and is still a fundamental issue in a variety of scientific and engineering fields [4],[6]. The principal goal is to learn an unknown functional mapping between an input vector and the output vector, using a set of known training samples. Once the mapping is generated, it can be used to obtain the output values given new input vectors. We will use functions with different numbers of inputs proposed in the bibliography [4]. Fig.3 shows an example of a two-input function.

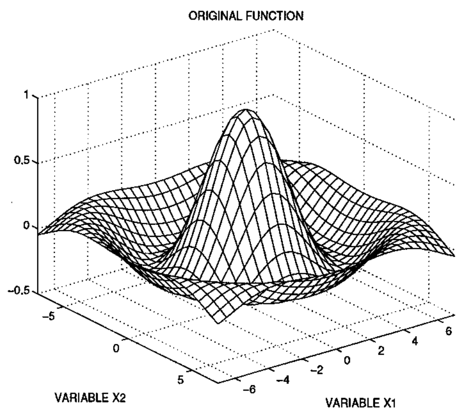

Fig. 3 Example of a function to be learned

5. Results of the statistical study

For the statistical study, all the possible configurations of factors used (presented in Table 1) are evaluated for each of the different problems presented, in order to obtain wide-ranging results. Table 2 gives the six-way variance analysis for whole set of examples of RBFN systems studied. As can be seen, the type of RBF structure used (weight average ($\widetilde{F}^*_{RBF}$) or weighted sum of the radial basis function ($\widetilde{F}_{RBF}$)), the type of radial function in the hidden layer and the distance used, present the greatest statistical relevance, in this order (because, the higher the F-Ratio or the smaller the Significance level, the greater the relevance of the corresponding factor).

Analyzing the different levels of each of these factors it is possible to understand their influence on the characteristics of the RBFN output. Figs. 4-6 show the mean value for each of the factors considered. The levels of each of the factors grouped by ellipses represent a homogeneous group. This means that the values corresponding to the levels of the given factor cannot be considered statistically different; therefore the designer can equally select any one of these, because the selection between the various levels belonging to a homogeneous group has no significant repercussion on the output. From Fig.4, it is clear that there are two

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 1-6

homogeneous groups of RBFN structures that are disjoint. The different distances analyzed are grouped in 3 sets (Fig.5). The first group is composed of the Manhattan and the Euclidean distance; the second comprises the supremum and finally, the third group includes the Minkowski distance with λ=3. Fig 6 shows the results for the type of function in the hidden layer, giving three homogeneous groups. One group includes the gaussian and the triangular functions, another contains the Epanechnikov and the inverse multiquadratic functions, and the last one groups the thin plate spline and the cubic functions. As can be observed, there exists an overlap between the groups (some non-linear functions belong simultaneously to more than one group). The biggest difference in the mean appears between the gaussian and the cubic.

Table 2 ANOVA table of the main factors

Main Factors	F-Ratio	Sig. Level
Structure of the RBF	56	0.00
Non-linear function	32	0.00
Distance	27	0.00
Neurons	1.24	0.29
Resolution	1.01	0.37
Noise (%)	0.95	0.43

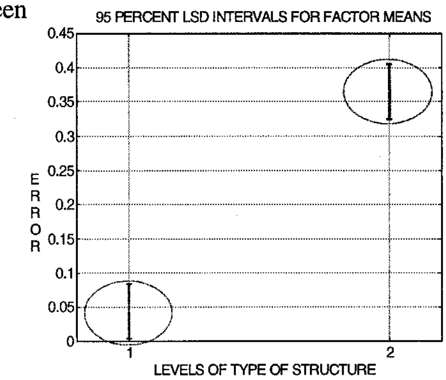

Fig. 4 Levels of variable structure of the RBF

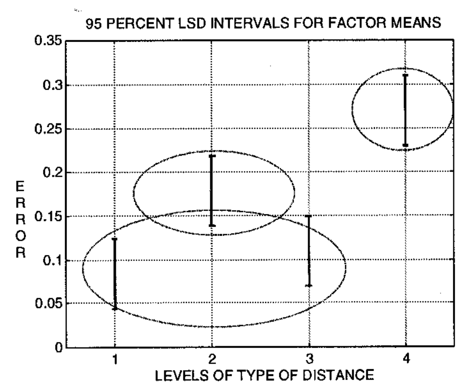

Fig. 5 Levels of variable type of distance

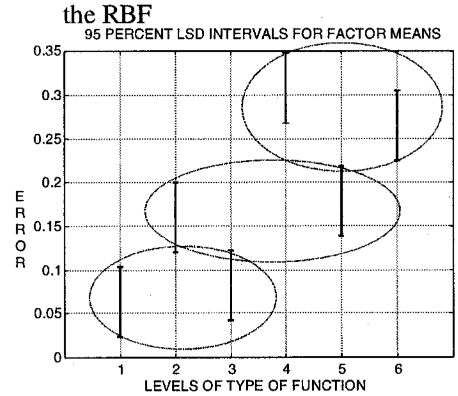

Fig. 6 Levels of variable type of function in the hidden layer

The worst values in the performance index used (the normalized mean squared error NMSE), were obtained when the weighted sum was used in conjunction with the distance λ= 3 and the cubic function in the hidden layer. The lowest error values in the output, and thus the best values for the performance index, correspond to the configurations using the weighted average, the Manhattan distance and gaussian function.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 1-6

6. Conclusion

The present statistical study was motivated by the great variety of alternatives that a designer has to take into account when developing an RBF. Thus, in addition to the existing intuitive knowledge, it is necessary to have a more precise understanding of the significance of the different alternatives.

An analysis was made of the principal functions required to design the neural network. The main goal of this study was to determine which variables are most influential on the response of an RBF. To carry out this analysis, the ANOVA statistical tool was used. Applying this methodology to a great variety of problems, it is possible to draw general conclusions about the most relevant factors defining the neural system. The most relevant factors are in this order: the structure of the RBFN (the output is the weighted sum or is the weighted average of the radial basis function), the distance used and the different alternatives for the nonlinear function in the hidden layer. Other, less significant factors, include the number of neurons in the hidden layer, the number of bits used for weight storage and the noise applitude used for corrupting the input pattern.

It should be emphasized that, using the techniques described, the designer of a neural network system is not only able to concentrate on the phases and basis primitives having greatest repercussion on the RBF design; it is also possible to create homogeneous groups of the levels of the analyzed factors with a similar performance.

REFERENCES

[1] M.Brown, C.Harris, "Neurofuzzy Adaptive Modelling and Control", Englewood Cliffs, NJ: Prentice-Hall, 1994.

[2] G.Casella, R.L.Berger, "Statistical Inference", *Duxbury Press*, 1990

[3] E.S.Chng, S.Chen, B.Mulgrew, "Gradiend radial basis function networks for nonliear and nonstarionary time series prediction", *IEEE Trans. on N..Networks*, vol.7, no.1, pp.190-194, January 1996.

[4] V.Cherkassky, D.Gehring, F.Mulier, "Comparison of Adaptive Methods for Function Estimation from Samples", *IEEE Tran. on N. Networks*, vol.7,no.4, pp.969-984, 1996.

[5] J.V.Hansen, R.D.Nelson, "Neural networks and traditional time series methods: A synergistic combination in state economic forecasts", *IEEE Trans. on N. Networks*, vol.8, no.4, pp. 863-873, July 1997.

[6] J.S.R.Jang, C.T.Sun, E.Mizutani, "Neuro-Fuzzy and soft computing", Prentice Hall, ISBN 0-13-261066-3, 1997

[7] W.Kaminski, P.Strumillo, "Kernel orthonormalization in radial basis function neural networks", *IEEE Trans. on Neural Networks*, vol.8, no.5, pp.1177-1183, September 1997.

[8] S. Makridakis, S. Wheelwright and R. Hyndman, "Forecasting: Methods and Applications", *To be published by Wiley in 1998.*

[9] Y.R.Park, T.J. Murray, C. Chen, "Predicting sunspots using a layered perceptron neural network", *IEEE Trans. on Neural Networks*, vol.7, no.2, pp.501-505, March 1996.

[10] M.A.S.Potts, D.S. Broomhead, "Time series prediction with a radial basis function neural network,", *SPIE Adaptive Signal Processing*, vol.1565, pp.255-266, 1991.

[11] I.Rojas, O.Valenzuela, A.Prieto,: "Statistical analysis of the main parameters in the definition of Radial Basis Function Networks". *Lecture Notes in Computer Science,* vol.1240. pp.882-891. Springer-Verlag 1997.

[12] B.A.Whitehead, Tinothy.D.Choate, "Cooperative-Competitive Genetic Evolution of Radial Basis Function Centers and Widths for Time Series Prediction", *IEEE Transaction on Neural Networks*, vol.7,no.4, pp.869-880, July, 1996.