ON THE CLASSIFICATION ENHANCEMENT OF RADIAL BASIS FUNCTION NETWORKS

Ö. Ciftcioglu, S. Durmisevic and S. Sariyildiz
TU Delft, Faculty of Architecture, Building Technology, Berlageweg 1,
2628 CR Delft, The Netherlands
e-mail: o.ciftcioglu@bk.tudelft.nl

Abstract: Artificial neural networks are powerful tools for analysing information expressed as data sets, which contain complex nonlinear relationships to be identified and classified. In particular radial basis function (RBF) neural networks have outstanding features for this. However, due to far reaching implications of the basis functions in the functionality of RBF networks they are still subject to study for best performance, in a general sense. One important parameter is the width of the radial basis functions. Here, we investigate the formation of a RBF neural network for its enhanced performance, which is closely related to the width parameter. For this aim, two key implementations are orthogonal least squares for training and multiresolutional decomposition of the sequence at the output of the network by *wavelets*.

Index Terms: RBF neural networks, orthogonal least squares, multiresolutional decomposition, wavelets

1. INTRODUCTION

RBF network represents the main structure of neuro-fuzzy systems and it has a feed-forward neural network form. Therefore, next to standard back-propagation algorithm, many other training algorithms are devised for efficient training and are reported in literature. Next to the training algorithms, there are a number of important issues on RBF network structure. The number of the hidden nodes, type of radial basis functions, width of the basis functions, centres of the basis functions are some of the examples on which numerous research works appeared in the literature. In contrast with this, however, there is remarkably only a few papers (Poggio & Girosi, 1990; Wong 1991; Borghese & Ferrari 1998) pointing out the functional approximation from the frequency domain viewpoint. They identify that basis functions basically behave as low pass filters. This basic observation already hints at the potential of the productive outcomes of frequency domain considerations of RBF networks. Wong hints the shortcomings of the RBF networks as over-filtering and difficult learning of high frequencies during the training. Borghese and Ferrari (1998) deal with the gaussian low-pass filters suggesting different gaussian units with different widths as hierarchical RBF networks. However, referring to this point, because of invariably low-pass filter properties of gaussians applied to the same network structure might limit the performance of the RBF network. Apparently, it is commonly accepted that change of the widths of gaussians do not play a major role on the performance of a RBF network (Lowe 1998). Referring to this ambiguity, in what follows, we are concerned with the effect of gaussian widths on the RBF performance and investigate this from the Fourier domain viewpoint to enhance the classification performance of the network.

The material is arranged as follows. In the following section we are concerned with the filtering effect of the gaussian radial basis function on the RBF network functionality. Based on this, in section three, wavelet transform is briefly explained. In section four a training algorithm of the network is identified together with explanation of the multiresolutional decomposition of the training set for enhanced performance. This is followed by experimental research and conclusions.

2. BASIS FUNCTIONS AS FILTERS

The architecture of a RBF network consists of an input layer, a hidden layer and an output layer. The hidden layer consists of a set of radial basis functions as nodes. Each node has a parameter vector **c** defining a cluster center dimension which is equal to the input vector. The hidden layer node calculates the Euclidean distance between the center and the network's input vector. The distance calculated is used to determine the radial base function output.

Conventionally, all the radial basis functions in the hidden layer nodes are the same type and usually gaussian. The response of the output layer node(s) can be seen as a map f: $R^n \rightarrow R$, of the form

$f(\mathbf{x}) = \Sigma\, w_i\, \Phi(||\mathbf{x} - \mathbf{c}_i||)^2$

Here the summation is over the number of training data N. $\mathbf{c}_i$ (i=1,2,….N) is the i-th centre which may be equal to the input vector $\mathbf{x}_i$ or may be determined in some other way. Once the basis function outputs are determined, the connection weights from hidden layer to the output are determined from a linear set of equations. As a result, accurate functional approximation is obtained. The complexity increases as the size of training data increases.

We consider a set of N data vectors $\{x_i\ ,\ i=1,...,N\}$ dimension of p in R^p and N real numbers $\{d_i, i=1,2,...,N\}$. We seek a function f(x): $R^p \rightarrow R^1$ that satisfies the interpolation conditions $f(x_i)=d_i$, i=1,2,...N. There are several methods as solutions for this interpolation problem, like Lagrange interpolation functions. We consider radial base functions (RBF) due to their suitability for multivariable interpolation. The characteristic feature of radial functions considered here is that their response decreases monotonically with distance from a central point. The RBF approach constructs a linear space using a set of radial basis functions $\phi(||\mathbf{x}-\mathbf{c}_i||)$ defined with a norm, which is generally Euclidean. The center described with a vector $\mathbf{c}_i$, a distance scale and the shape of the radial function are parameters of the model. By means of these base functions, we can model the function as

$$y = \sum_{i=1}^{N_h} h_i \phi_i [d(u, c_i)] \qquad (1)$$

$$d(u, c_i) = ||\, u - c_i\, ||_2 = \left(\sum_{i=1}^{N_i} (u_j - c_{ij})^2 \right)^{1/2} \qquad (2)$$

d(.) is a distance measure, usually taken to be the Euclidean norm. For gaussian radial basis functions, we can write

$$y(x) = \sum_{k=1}^{N} w_k\, g(x, c_k) = \sum_{k=1}^{N} w_k\, \exp((\frac{x - c_k}{\sigma})^2) \qquad (3)$$

where σ is the width parameter. In continuous form, we write

$$y(x) = \int_R w(c) g(x - c)\ dc \qquad (4)$$

where $(c_{k+1} - c_k) \rightarrow 0$. In the Fourier domain

$Y(\omega) \cong \bar{Y}(\omega) = W(\omega) G(\omega)$ and because of the interpolation condition in the discrete form, $Y(\omega) = W(\omega)$ and therefore, we obtain

$$\bar{Y}(\omega) = Y(\omega) G(\omega) \qquad (5)$$

It is clear that, if G(ω) has a flat spectrum, i.e., all pass filter, than $\bar{Y}(\omega) = Y(\omega)$ which corresponds to the case over fitting. This is not desirable because of the degradation of the generalization capability. G(ω) should have some low pass filter characteristic as gaussian radial basis function satisfies this. The cut off frequency is determined by the width parameter σ. However, to take a single σ value in a network cannot satisfy the space-frequency requirements of the fitting. That is spatially at some regions we need higher frequency to follow variations accurately in the approximation, and at some regions vice versa to avoid over fitting. To circumvent this, we consider (5) in the form

$$\bar{Y}(\omega) = Y_1(\omega) G_1(\omega) + Y_2(\omega) G_2(\omega) \qquad (6)$$

where $Y_1(\omega)$ *and* $Y_2(\omega)$ are the orthogonal decomposition of $Y(\omega)$; $G_1(\omega)$, $G_2(\omega)$ are suitable gaussian filters, matching $Y_1(\omega), Y_2(\omega)$ respectively. Now, the orthogonal decomposition of Y(ω) can be carried out by means of wavelet decomposition and suitable gaussian functions matching the decomposition can be determined together with, $Y_1(\omega), Y_2(\omega)$ matching the decomposition. Before explaining how this should be done, a brief description of wavelets is in order to facilitate the explanation.

3. WAVELET TRANSFORM

We first introduce wavelets and multiresolution analysis (Daubeshies, 1992; Mallat, 1999). Given a function f(x), wavelet transform provides coefficients, called 'wavelet' coefficients, which are resulted from the inner products of the signal and a family of 'wavelets'. With these coefficients and the associated wavelet functions the function f(x) can be expressed in different level of approximations. For continuous wavelet transform, the wavelet, corresponding to a scale ***a*** and the time location ***b*** is

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi(\frac{x - b}{a}) \qquad (7)$$

where $\psi_{a,b}(x)$ is the wavelet which is an element of the space $\mathbf{L}^2(\mathbf{R})$. Functions f(x) in this space must satisfy

$$\int_{-\infty}^{+\infty} |\, f(x)\,|^2\ dx \qquad (8)$$

The wavelet is subject to the following additional constraints

$$\int_{-\infty}^{+\infty} \psi(x)dx = 0 \qquad (9)$$

Both ψ and its Fourier transform Ψ must be window functions, that is, they must have a well defined center radius so that they are localized in both the time and frequency domains within the limits imposed by uncertainty principle. This implies that a wavelet function has general shape that decays rapidly and also exhibits some degree of oscillation. The integral wavelet transform $(W_\psi f)_{(a,b)}$ is defined as

$$(W_\psi f)_{(a,b)} = \mid a \mid^{-1/2} \int_{-\infty}^{+\infty} f(x)\psi_{a,b}\left[\frac{x-b}{a}\right]dx \qquad (10)$$

where $\boldsymbol{a,b} \in \mathbf{R}$ and $f \in \mathbf{L}^2\mathbf{(R)}$. Since, ψ is localisable in both time and frequency scale, the integral wavelet transform is also localised and gives information in both domains within the bounds of the uncertainty principle. $\psi_{a,b}(x)$ are the basis functions and can be viewed as contracted and shifted versions of the prototype basis function $\psi(x)$. The parameter *a* is the scale parameter, and the argument *b* is the location variable. If the scale parameter is large, then the basis function becomes a stretched version of the prototype, corresponding to a low frequency function. Similarly, small values of the scale parameter *a* corresponds to high frequencies. Since the windowing is contained in the basis functions, they handle the time-frequency window. The reconstruction of the function f(x) from its wavelet transform counterpart is possible and it is given by

$$f(x) = \frac{1}{C_\psi}\int_{-\infty}^{+\infty}\int_{0}^{\infty}(W_\psi f)_{(a,b)}a^{-1/2}\psi(\frac{x-b}{a})\frac{da\,db}{a^2} \qquad (11)$$

where

$C_\psi = \int_0^\infty \frac{|\Psi(\omega)|^2}{\omega}d\omega$ is the admissibility condition

$\Psi(\omega)$ *being Fourier transform of* $\psi(x)$. As seen from above f(x) can be described as a summation of the basis functions $\psi_{a,b}(x)$ with the weights of the wavelet coefficients $(W_\psi f)_{(a,b)}$. One can discretize the values of *a* and *b* and it is still possible to reconstruct the signal from its transform. For the discretization we define

$$a = a_o^m, b = n\,b_o\,a_o^m,\ m,n \in Z,\ a_o > 1,\ b_o \neq 0 \qquad (12)$$

For computational convenience b_o is conventionally taken equal to unity and a_o=2 so that the exact reconstruction can be achieved and the set of basis functions $\psi_{a,b}(x)$ form an orthogonal basis. Then the inverse WT becomes

$$f(x) = C_\psi \sum_{m\in Z}\sum_{n\in Z} d_{m,n}\,a_o^{-m/2}\psi(a_o^{-m/2}x - n) \qquad (13)$$

In the spatial decomposition, the continuous wavelet described above should be applied in discrete form..For functions in the $\mathbf{L}^2$(**R**), i.e., for square integrable functions , the functions

$$\phi_{m,n}(x) = 2^{-m/2}\phi(2^{-m}x - n) \qquad (14)$$

form an orthonormal basis. These are called scaling functions and for m=0, we basically write

$$\phi_{0,n}(x) = \phi(x - n) \qquad (15)$$

The function f(x can be expressed by these orthogonal base functions as approximation in such a way that

$$f(x) = \lim_{m\to\infty} f_m(x) \qquad (16)$$

where

$$f_m(x) = \sum_n < \phi_{m,n}(x),\, f(x) > \phi_{m,n}(x)$$

$$= \sum_n c_{m,n}\,\phi_{m,n}(x) \qquad (17)$$

where $< \phi_{m,n}(x),\, f(x) >$ is the inner product:

$$< \phi_{m,n}(x),\, f(x) > = \int_{-\infty}^{\infty} \phi_{m,n}(x)\, f(x)\, dx$$

Now, let $\psi(x) = \psi_{0,0}(x)$ be a basis function and

$$\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n) \qquad (18)$$

The functions $\psi_{m,n}(x)$ are identical to the wavelets, after the discretization, described before. There are strong relations between ϕ(x) and $\psi(x)$. The introduction of the wavelet functions enables us to write any function f(x) in $\mathbf{L}^2$(**R**) as a sum , of the form

$$f(x) = \sum_{j=-\infty}^{\infty} w_j(x)$$

where $w_j(x) = \sum_k < \psi_{j,k}(x),\, f(x) > \Psi_{j,k}(x)$ (19)

Considering a certain scale m, the function f(x) can be written as the sum of a low resolution part f_m(x) and the detail part which is constituted by the wavelets so that

$$f(x) = f_m(x) + \sum_{j=-\infty}^{m} w_j(x) \qquad = \sum_n < \phi_{m,n}(x),\, f(x) > \phi_{m,n}(x) + \sum_{j=-\infty}^{k}\sum_{k}^{m} < \psi_{j,k}(x),\, f(x) > \psi_{j,k}(x) \qquad (20)$$

which can be expressed as

$$f(x) = \sum_{n} c_{m,n}\phi_{m,n}(x) + \sum_{j=-\infty}^{m}\sum_{k} d_{j,k}\psi_{j,k}(x) \quad (21)$$

Above, the coefficients $\boldsymbol{d}_{j,k}$ are known as the ***wavelet coefficients***. From the preceding equation multiresolution decomposition is represented by an approximation i.e., the first term with $\phi_{m,n}(x)$ functions, and the detail part i.e., the second term with the $\psi_{j,k}(x)$ functions. The variable *m* indicates the scale and is called *scale factor* or *scale level*. If the scale level *m* is high, it indicates that the function is a coarse approximation of f(x), so the details are neglected. On the contrary, if the scale level is low, a detailed approximation of f(x) is achieved. Incidentally, referring the research reported here, the scale level is m=-3 and the detail levels are j=–2 and j=–1.

4. CLASSIFICATION ENHANCEMENT

In the RBF network if f(x) contains relatively high frequency components, than it is necessary to use relatively more basis functions for the same approximation accuracy at the cost of degradation of the generalization capability. That is, if the number of basis functions is relatively few with higher width (σ) than the approximation error is high for rapidly changing functions. This results in higher generalization error. Conversely if the number of basis functions is relatively high with smaller σ, than generalization error is again high due to redundancy in a slowly changing functions. These two different situations can occur at the same time at the different parts of the function being approximated. This is the case in general and therefore clearly explains the common observation that change of the widths of gaussians do not play a major role on the performance of a RBF network. The solution of this inconvenience is the multiresolutional decomposition of the function subject to approximation by the wavelet transform. By doing so, the function is decomposed into one or more functions with the same number of data points but corresponding to different regions in the Fourier domain. In these regions, the sequences are orthogonal to each other and they satisfy (8). This is the total energy piecewise delivered to respective Fourier domains at the output of the RBF network. In particular, for the approximation part i.e., the first term in (21), it is the variance after mean is removed, and for the detail parts, i.e., the rest in (21), it is the variance of the RBF output. These variances are accurately considered in a particular training algorithm, which is known as orthogonal least squares (OLS) (Chen et al., 1990). Essentially, OLS considers that a RBF network is the generalization of the linear regression model

$$d(x) = \sum_{i=1}^{N} \theta_i\, r_i(x) + \varepsilon(x) \quad (22)$$

where d(x) is desired output, $r_i(x)$ are the regressors which are some fixed functions of x. If the model is right, then ε is not correlated with the regressors. For a set of input-output pairs, this model, in matrix form can be expressed as

$$\mathbf{d} = R\,\theta + E \quad (23)$$

where **d**, is the desired output vector; R is the regression matrix that consists of regressor vectors r_i each of which has dimension of M ; θ is the parameter vector; E is the error vector. The regressor vectors r_i form a set of basis vectors, and the least squares estimation of θ provides that ($R\theta$) is the projection of **d** onto the space spanned by these vectors. The **OLS** method makes orthogonal decomposition of R so that (22) becomes **d**=W **q** + E, where W is a matrix with orthogonal columns providing computational convenience for training of RBF.

Powell (1992) has traditionally used the RBF method for strict interpolation in multidimensional space. Michelli(1986) required as many centers as data points (assigning all input data points as centers). Later Broomhead and Lowe (1988) removed this restriction and used less centers than data samples so that the case became approximation in multidimensional space. In the latter case, generally, selection of the centers is another issue in the RBF applications. However, in the OLS method used for training, the centers are selected as the "appropriate" data points. Each decomposed function component, i.e., approximation and detail parts are approximated by a separate RBF network with appropriate width (σ) parameter matching to the respective frequency bands. Explicitly, for a lower frequency band, a higher σ value with less number of bases functions and vice versa. Due to orthogonal transformation, the approximated functions are finally summed to obtain the estimation of the function subjected to decomposition.

5. EXPERIMENTAL RESEARCH

Verification of the theoretical considerations presented in the preceding section is performed by means of a set of architectural data subjected to RBF network modeling. The input space of the network is 43-dimensional. The output is two-dimensional.

Based on the data set applied to the input, classification is aimed at the output where categorically there are five distinct partitions at each output variable range. The data are obtained by means of a questionnaire. The questionnaire is not simply the gathering of facts but more importantly it is an instrument for gathering meaningful data to test the hypotheses. The hypotheses are consciously developed and tested through the questionnaire by means of words, questions and specific layout/format. The structure, purpose and the meaning of the researched topic is provided through hypotheses. Having firm hypotheses in mind they actually later on serve to form the questions and to omit the unnecessary questions that are not related to the hypothesis (Labaw, 1980).

In general, the way in which data is measured is called a level of measurement or the scale of measurement for variables. There are four kinds of scales in which variables can be measured: the *nominal* scale, the *ordinal* scale, the *interval* scale and the *ratio* scale (Dalen & Leede, 2000). In this research we have used an *ordinal scale measurement*, where such measurement involves placement of values in a rank order, in order to create an ordinal scale variable. The relationship between observations takes on a form of 'greater than 'and 'less than'. For this questionnaire a five-point scale was used. In such way, respondents have the opportunity to 'strongly agree' or 'strongly disagree' with a question or to strongly express their opinion regarding design issues.

For a case study, Blaak underground station in the Netherlands was chosen for which the questionnaire was designed. This is an important exchange station, which is situated in the center of Rotterdam. It is at the same time a tram, metro and a train station. Tram station is situated on a ground level. Metro platforms are one level below ground (at approximately -7m) and train platforms are two levels below ground (at approximately -14 meters). From 27th May till 30th May 2000, one thousand of questionnaires were handed out to the passengers visiting the station. The questionnaire covered aspects that are related to safety and comfort at the station. In total there were 43 aspects in input space each having five possible options and two aspects in the output space again having five possible options. The latter two are design variables being *safety* and *comfort*. The input aspects, which are identified to be related to safety are given in *table 1* and those related to comfort are presented in *table 2* (Ciftcioglu, Durmisevic, et al., 2001). Main purpose of the questionnaire was to provide information on user's perception regarding specific spatial characteristics of that station. The questions covered all aspects given in *table 1* and *table 2* and additional two final questions were related to user's perception of public safety and comfort at Blaak station.

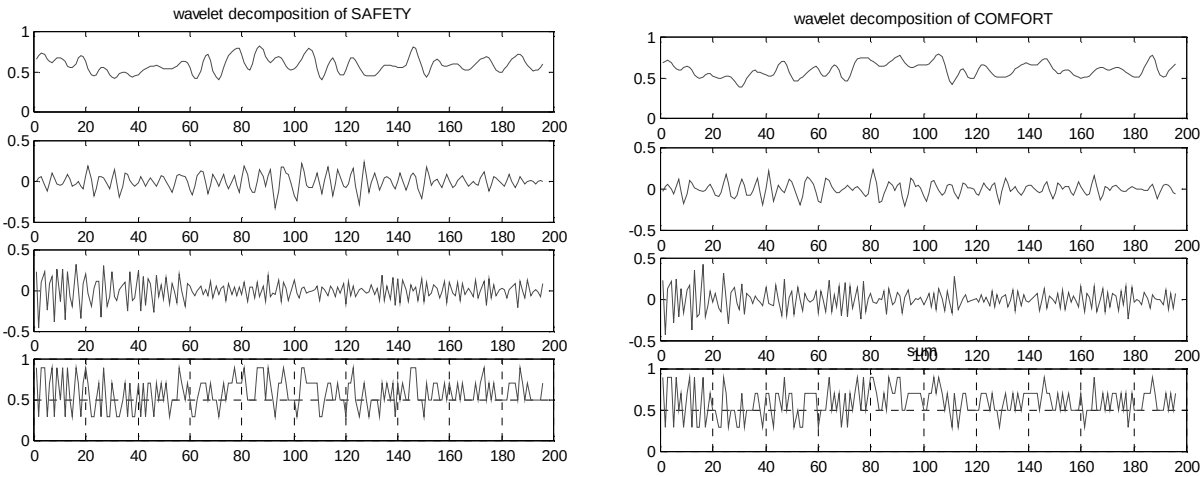
Overview	**Escape**	**Lighting**	**Presence of people**	**Safety surrounding**
Entrance	Possibilities	Entrance	Public control	Safety in surrounding
Train platform	Distances	Train platform	Few people daytime	
Metro platform		Metro platform	Few people night	
Exchange area		Exchange area		
		Dark areas		

Table1: Aspects related to safety (15 aspects)

Attractiveness	**Wayfinding**	**Daylight**	**Physiological**
Colour	To the station	Pleasantness	Noise
Material	In station	Orientation	Temperature winter
Spatial proportions	Placement of signs		Temperature summer
Furniture	Number of signs		Draft entrance
Maintenance			Draft platforms
Spaciousness entrance			Draft exchange areas
Spaciousness train platform			Ventilation entrance
Spaciousness metro platform			Ventilation platforms
Platform length			
Platform width			
Platform height			
Pleasantness entrance			
Pleasantness train platform			
Pleasantness metro platform			

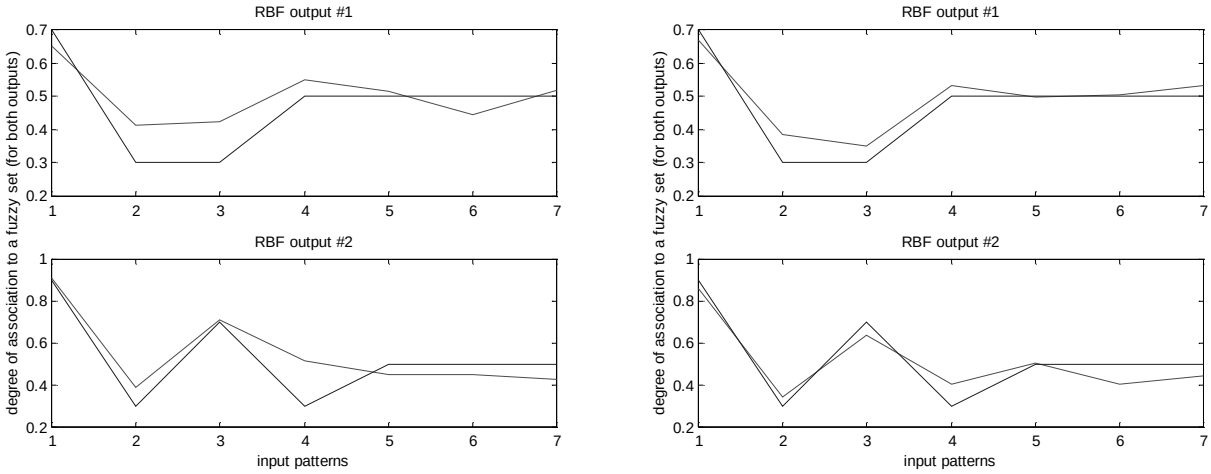
Table2: Aspects related to comfort (28 aspects)

For the analysis, the linguistic information is firstly converted to terms in fuzzy logic domain and after appropriate treatment, the data analyses are carried out and the results are expressed in most comprehensible form for design assessments. Such conversions are referred to as fuzzification and defuzzification, where the data are expressed in numerical form and therefore become convenient for mathematical treatment. Conventionally, wavelet transform is conveniently computed for a number of data points having a length equal to the power of two (Press, 1992). With this limitation, since we had only 196 input-output pairs, this number had to be reduced to 128. To avoid this loss of data in the training, a special wavelet transform algorithm is prepared for this study which can deal with any number of length of data for the output parameter safety. The wavelet decomposed data for two RBF outputs are shown in figure 1. In figure (a) and (b) the uppermost curve is the approximation and lower two are two detail levels in the multiresolutional decomposition, the average being zero. The sum of three curves makes the output data subjected to decomposition. Three separate RBF networks are trained by OLS algorithm using the sets of wavelet approximation and two wavelets details data. Here it is important to note that, the width parameter (σ) in each network is different matching the respective band in the Fourier domain. Explicitly for approximation σ=1.4 and for the detail components σ=1.1 and σ=1., respectively. The trained RBF networks are tested with the test data, which contained 7 test cases. Typical test results from estimations using a single RBF network trained by using OLS algorithm is shown in figure 2a. Using the same test data, but RBF networks trained by multiresolutional set of data is shown in figure 2b where the mean-square error criterion is approximately twice as much smaller than that obtained from single RBF network. Here considering the nature of the data, the estimations are satisfactory for the purpose. Namely, the input data set belong to architectural qualitative quantities dealt with fuzzy logic techniques, there are fuzzy imprecision on them. Due to this, there are fuzzy imprecision in the estimations and the differences between two lines are within the allowed tolerance limits. More importantly, here it is to note that, by means of multiresolutional RBF network, estimations are robust. That is, since the training information orthogonally expanded, each component is treated with a matching RBF network so that enhanced estimation is achieved as algebraic sum of each sub-network outcome. Added to this, since the networks are independent, the random errors due to various error sources have less influence on the final outcome. .



(a) *(b)*

Figure 1:Two wavelet decomposition as approximation and two detail levels, from the top, obtained from RBF network's a series of two outputs (in boxes with grids) used for multiresolutional training.



(a) *(b)*

Figure 2:Test results from single RBF network (a) and multiresolutional RBF network (b)

6. CONCLUSIONS

In RBF networks the width of the basis functions is an essential parameter and it imposes limitation in the accuracy of the classification and this is an important issue on RBF network functionality. The required essential solution to this inconvenience is obtained by orthogonal wavelet decomposition. For each orthogonal wavelet decomposition of every output space variable, a different RBF network is trained with appropriate widths. The final outcome is obtained simply by algebraically summing up the outcomes from respective RBF networks corresponding to approximation and details parts of the wavelet decomposition. A marked enhancement in the RBF classification is obtained at the cost of intensive computational efforts based on the *perfect reconstruction* properties of wavelet decomposition. However, more importantly, a robust enhancement on the generalization capability of RBF network is obtained that implies a robust estimation method by RBF networks. Theoretical details of the approach are presented and the effectiveness of this novel approach is demonstrated by experiments using actual data in the context of artificial intelligence in design.

7. REFERENCES

Borghese N.A. and Ferrari S., (1998), "Hierarchical RBF networks and local parameter estimate", *Neurocomputing* 19, pp.259-283

Broomhead, D.S. and D. Lowe (1988), "Multivariable Function Interpolation and Adaptive Networks", *Complex Systems, 21*

Chen S, C.F.N. Cowan and P.M. Grant, (1991), "Orthogonal Least Squares Algorithm for Radial Basis Function Networks", *IEEE Trans. on Neural Networks*, Vol.2, No.2, March

Ciftcioglu, Ö., Durmisevic, S. and Sariyildiz, S., (2001), "Multi-resolutional Knowledge Representation", Conf. proceedings EuropIA8, 25-27 April 2001, Delft, The Netherlands

Dalen, van J. and Leede, de E. (2000*) Statistisch Onderzoek met SPSS for Windows*, Uitgeverij Lemma BV, Utrecht

Daubechies I (1992), "Ten lectures on wavelets", *CBMS-NSF regional conference series in applied mathematics*, *SIAM* Vol.61

Labaw, P. J. (1980). *Advanced Questionnaire Design*. Abt Books, Cambridge, Massachusetts

Lowe D. (1998), "Characterising complexity by the degrees of freedom in a radial basis function network", *Neurocomputing 19*

Mallat S (1999), *Wavelet Tour of Signal Processing*, Academic Press, San Diego, London

Michelli, A.C., (1986), "Interpolation of scattered data:Distance matrices and conditionally positive definite functions", Construct. Approx., Vol.2,

Poggio,T and F Girosi.(1990), "Network for Approximation and Learning", *Proc.IEEE*, 78(9),

Powell, M.J.D. (1992), "Radial basis functions in 1990", Advanced Numerical Analysis,Vol.2.

Press, W.H., Teukolsky S.A., Vetterling, W.T., (1992), *Numerical Recipes in C*, Cambridge University Press

Wong Y. (1991), "How Gaussian Radial Basis Functions Work", *IJCNN*, Int'l Conference on Neural Networks, July 8-12, 1991, Seattle, WA