Performance analysis of pattern classification for the Handwritten English vowels with Back propagation & DG-RBF Feed forward Neural Networks

Naveen Kumar Sharma[1], S R Pande[2] and Manu Pratap Singh[3]*

Abstract

The purpose of this study is to analyze the performance of feed forward neural network for the pattern classification of hand written English vowels using conventional back propagation algorithm for multi layer feed forward neural network and decent gradient learning for radial basis function network. This analysis has been done with five different samples of hand written English vowels. These characters are presented to the neural network for the training. Adjusting the connection strength and network parameters perform the training process in the neural network. By using a simulator program, each algorithm is compared with five data sets of handwritten English language vowels. The 5 trials indicate the significant difference between the two algorithms for the presented data sets. The results indicate that the performance of the neural network is much efficient and convergence for the RBF network.

Key words : Pattern Classification, Feed forward neural network, Back propagation algorithm, Radial Basis Function Neural Network.

[1] CET-IILM-AHL, Knowledge Park II, Greater Noida, India, Email: nivan101@gmail.com

[2] Department of Computer Science, SSES Amti's Science College, Congressnagar, Nagpur, India Email: srpande65@rediffmail.com

[3] Department of Computer Science, ICIS, Dr. B. R. A. University, Khandari, Agra, India Emil: manu_p_singh@hotmail.com

*Corresponding author

1. Introduction

An artificial neural network (ANN) is a well-established technique for creating the artificial intelligence in the machine. This is an attempt to simulate the human behavior in the machine for the various pattern recognition tasks [1]. Neural networks consist of computer programmable objects called as neurons. These neurons are programmed to perform a simple mathematical function or to process a small portion of data. A neuron is interconnected with other neurons with the connection strength known as weight. These weights of the neural network are adjustable in nature to adept the behavior of input pattern information. Thus, by adjusting the weights of the network, the behavior of the neural network can be altered and controlled. This mechanism in neural network system is known as learning.

Neural networks have been used in a number of applications such as pattern recognition & classification [2,3,4,5], remote sensing [6], dynamic modeling and medicine [7]. The increasing popularity of the neural networks is partly due to their ability to learn and generalization. Particularly, feed forward neural network makes no prior assumption about the statistics of input data and can construct complex decision boundaries [8]. This property makes neural networks, an attractive tool to many pattern classification problems such as hand written curve scripts [3, 9, 10, 11].

This research has been focused on the recognition of handwritten English vowels in its most basic form i.e. individual character classification. The rationale for this study is to improve the efficiency of neural network for handwritten character classification task. In this paper we propose a more suitable and efficient learning method for feed forward neural networks when neural networks are used as a classifier for the hand written English vowels.

There are different types of architectures and designs for the neural networks, but here we discuss the most common one, i.e. feed forward manner. In a feed forward neural network the nodes are organized into layers; each "stacked" on each other. The neural network consists of an input layer of nodes, one or more hidden layers, and an output layer [12]. Each node in the layer has one corresponding node in the next layer, thus creating the stacking effect. The input layer's nodes consists with output functions those deliver data to the first hidden layers nodes. The hidden layer(s) is the processing layer, where all of the actual computation takes place. Each node in a hidden layer computes a sum based on its input from the previous layer (either the input layer or another hidden layer). The sum is then "compacted" by an output function (sigmoid function), which changes the sum down to more a limited and manageable range. The output sum from the hidden layers is passed to the output layer, which exhibits the final network result. Feed-forward networks may contain any number of hidden layers, but only one input and one output layer. A single-hidden layer network can learn any set of training data that a network with multiple layers can learn [13]. However, a single hidden layer may take longer to train.

In neural networks, the choice of learning algorithm, network topology, weight and bias initialization and input pattern representation are important factors for the network performance in order to accomplish the learning. In particular, the choice of learning algorithm determines the rate of convergence, computational cost and the optimality of the solution. The multi layer feed forward is one of the most widely used neural network architecture. The learning process for the feed forward network can consider as the minimization of the specified error (E) that depends on all the free parameters of the network. The most commonly adopted error function is the least mean square error. In the feed forward neural network with J processing units in the output layer and for the l^{th} pattern, the LMS is given by;

$$E^l = \frac{1}{2}\sum_{j=1}^{M}(d_j^l - y_j^l)^2 \qquad (1.1)$$

where $l = 1$ to L(total number of input-output pattern pairs of training set)

Here d_j^l and y_j^l are the desired and actual outputs corresponding to the l^{th} input pattern. Hence, due to the non-linear nature of E, the minimization of the error function is typically carried out by iterative techniques [14]. Among the various learning algorithms, the back propagation algorithm [15] is one of the most important and widely used algorithms and has been successfully applied in many fields. It is based on the steepest descent gradient and has the advantage of being less computationally expensive. However, the conventional back propagation learning algorithm suffers from short coming, such as slow convergence rate and fixed learning rate. Furthermore it can be stuck to a local minimum of the error.

There are numerous algorithms have been proposed to improve the back propagation learning algorithm. Since, the error surface may have several flat regions; the back propagation algorithm with fixed learning rate may be inefficient. In order to overcome with these problems, vogel et. al. [16] and Jacobs [17] proposed a number of useful heuristic methods, including the dynamic change of the learning rate by a fixed factor and momentum based on the observation of the error signals. Yu et. al. proposed dynamic optimization methods of the learning rate using derivative information [18]. Several other variations of back propagation algorithms based on second order methods have been proposed [19-23]. This method generally converges to minima more rapidly than the method based solely on gradient decent method. However, they require an additional storage and the inversion of the second-order derivatives of the error function with respect to the weights. The storage requirement and computational cost, increases with the square of the number of weights. Consequently, if a large number of weights are required, the application of the second order methods may be expensive.

In this paper, we consider the two neural networks architectures (NN1 & NN2). The NN1 is trained with the conventional back propagation learning algorithm with incorporation of momentum terms & Doug's Momentum descent term [24]. The NN2 network architecture has been implemented with the Radial basis function [25] in the single hidden layer. The performance of these two network architectures has been analyzed for the handwritten English vowels. Analysis has been conducted with the series of tests to determine which of two learning algorithms, back propagation or decent gradient with RBF, trained the feed forward neural network faster and more efficiently. The rate of convergence and the number of epochs for each pattern are important observation of this study. The simulated results are determined from the number of trails with five sets of handwritten characters of English vowels.

The next section presents the implementation of the neural network architecture with Radial basis function. The simulation design and algorithmic steps of the problem are represented in section 3. The experimental results and discussion are presented in section 4. Section 5 contents the conclusion of this paper and the future research directions.

2. Implementation of the Radial basis function

There are various methods for classification problems [26] like the handwritten English characters recognition and each of them has pros and cons. Table 1 presents a summary of the evaluation of the representative pattern classification methods based on the training time and classification time for the conventional back propagation algorithm and decent gradient with the Radial basis function (DG-RBF).

Table 1 : A summary of the evaluation based on the training time & classification time for DG-RBF & BP

Classifiers	Training time	Classification time
DG-RBF	Short	Short
BP	Long	Long

The table.1 shows that the DG-RBF network is having better than the BP classifiers. Therefore DG-RBF network could be a reasonable choice for those classification problems, which do not have any particular requirements.

The architecture and training methods of the RBF network are well known [27, 28, 29, 30, 31, 32] & well established. The Radial basis function network (RBFN) is a universal approximator with a solid foundation in the conventional approximation theory. The RBFN is a popular alternative to the MLP, since it has a simpler structure and a much faster training process. The RBFN has its origin in performing exact interpolation of a set of data points in a multidimensional space [33, 25]. The RBFN is having, network architecture similar to the classical regularization network [28], where the basis functions are the Green's functions of the Gram operator associated with the stabilizer. If the stabilizer exhibits radial symmetry, the basis functions are radially symmetric as well and an RBFN is obtained. From the viewpoint of approximation theory, the regularization network has three following desirable properties [34, 35]:

1. It can approximate any multivariate continuous function on a compact domain to an arbitrary accuracy, given a sufficient number of units.

2. The approximation has the best-approximation property since the unknown coefficients are linear.

3. The solution is optimal in the sense that it minimizes a functional that measures how much it oscillates.

An RBFN is a three layer feed forward network that consists of one input layer, one hidden layer and one output layer as shown in figure (1), each input neuron corresponds to a component of an input vector x. The hidden layer consists of K neurons and one bias neuron. Each node in the hidden layer uses an RBF denoted with $\phi(r)$, as its non-linear activation function.

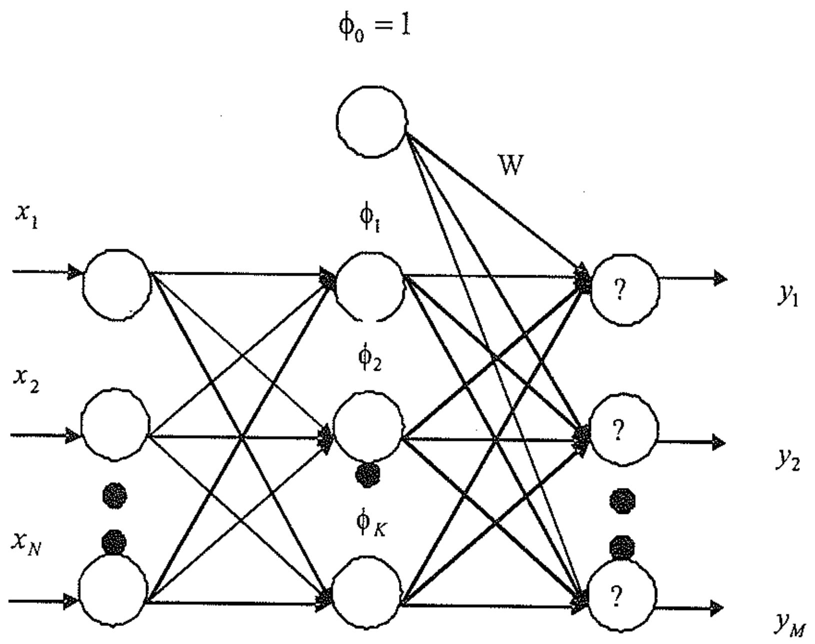

Figure 1 : Architecture of the RBFN. The input layer has N nodes; the hidden and the output layer have K and M neurons, respectively. $\phi_0(x) = 1$, corresponds to the bias.

The hidden layer performs a non-linear transform of the input and the output layer this layer is a linear combiner which maps the nonlinearity into a new space. The biases of the output layer neurons can be modeled by an additional neuron in the hidden layer, which has a constant activation function $\phi_0(r) = 1$. The RBFN can achieve a global optimal solution to the adjustable weights in the minimum MSE range by using the linear optimization method. Thus, for an input pattern x, the output of the j[th] node of the output layer can define as;

$$y_j(x) = \sum_{k=1}^{K} w_{kj}\phi_k(\|x_i - \mu_k\|) + w_{0j} \qquad (2.1)$$

for $j = (1,2,\ldots\ldots,M)$ where $y_j(x)$ is the output of the j^{th} processing element of the output layer for the RBFN, w_{kj} is the connection weight from the k[th] hidden unit to the j[th] output unit, μ_k is the prototype or centre of the k[th] hidden unit. The Radial Basis Function $\phi(.)$ is typically selected as the Gaussian function that can be represented as:

$$\phi_k(x_l) = \exp(-\frac{\|x_l - \mu_k\|^2}{2\sigma_k^2}) \quad \text{for}$$

$$k = (1,2,\ldots\ldots,K) \quad (2.2)$$

and 1 for $k = 0$ (bias neuron)

Where x is the N- dimensional input vector, μ_k is the vector determining the centre of the basis function ϕ_k and σ_k represents the width of the neuron. The weight vector between the input layer and the k[th] hidden layer neuron can consider as the centre μ_k for the feed forward RBF neural network.

Hence, for a set of L pattern pairs $\{(x_l, y_l)\}$, (2.1) can be expressed in the matrix form as

$$Y = w^T \phi \quad (2.3)$$

where $W = [w_1, \ldots\ldots\ldots\ldots w_m]$ is a KxM weight matrix, $w_j = (w_{0j}, \ldots\ldots\ldots\ldots w_{kj})^T$, $\phi = [\phi_0, \ldots\ldots\ldots\phi_k]$ is a K x L matrix, $\phi_{l,k} = [\phi_{l,1}, \ldots\ldots\ldots\phi_{l,k}]^T$ is the output of the hidden layer for the l[th] sample, $\phi_{l,k} = \phi(\|x_l - c_k\|)$, $Y = [y_1, y_{2,}\ldots\ldots y_m]$ is a M x L matrix and $y_{lj} = (y_{l1,}\ldots\ldots y_{lm})^T$.

The important aspect of the RBFN is the distinction between the rules of the first and second layers weights. It can be seen [24] that, the basis functions can be interpreted in a way, which allows the first layer weights (the parameters governing the basis function), to be determined by unsupervised learning. This leads to the two stage training procedure for RBFN. In the first stage the input data set {x[n]} is used to determine the parameters of the basis functions. The basis functions are then keep fixed while the second – layer weights are found in the second phase of training. There are various techniques have been proposed in the literature for optimizing the basis functions such as unsupervised methods like selection of subsets of data points [36], orthogonal least square method [37], clustering algorithm [28], Gaussian mixture models [38] and with the supervised learning method.

It has been observed [39] that the use of unsupervised techniques to determine the basis function parameters is not in general an optimal procedure so far as the subsequent supervised training is concerned. The difficulty with the unsupervised techniques arises due to the setting up of the basis functions, using density estimation on the input data and takes no consideration for the target labels associated with the data. Thus, it is obvious that to set the parameters of the basis functions for the optimal performance, the target data should include in the training procedure and it reflects the supervised training. Hence, the basis function parameters for regression can be found by treating the basis function centers and widths along with the second layer weights, as adaptive parameters to be determined by minimization of an error function. The error function has considered in equation (1.1) as the least mean square error (LMS). This error will minimize along the decent gradient of error surface in the weight space between hidden layer and the output layer. The same error will minimize with respect to the Gaussian basis function's parameter as defined in equation (2.2). Thus, we obtain the expressions for the derivatives of the error function with respect to the weights and basis function parameters for the set of L pattern pairs (x^l, y^l) as; where $l = 1$ to L.

$$\Delta w_{jk} = -\eta_1 \frac{\partial E^l}{\partial w_{jk}} \quad (2.4)$$

$$\Delta \mu_k = -\eta_2 \frac{\partial E^l}{\partial \mu_k} \quad (2.5)$$

and $\Delta \sigma_k = -\eta_3 \frac{\partial E^l}{\partial \sigma_k}$ (2.6)

here, $E^l = \frac{1}{2}\sum_{j=1}^{M}(d_j^l - y_j^l)^2$

and $y_j^l = \sum_{k=1}^{K} w_{jk}\phi_k(\|x^l - \mu_k^l\|)$ (2.7)

and $\phi_k(\|x^l - \mu_k^l\|) = \exp(-\frac{\|x^l - \mu_k^l\|^2}{2\sigma_k^2})$

Hence, from the equation (2.4) we have,

$$\Delta w_{jk} = -\eta_1 \frac{\partial E^l}{\partial w_{jk}} = -\eta_1 \frac{\partial E^l}{\partial y_j^l}.\frac{\partial y_j^l}{\partial w_{jk}} = -\eta_1 \frac{\partial E^l}{\partial y_j^l}.\phi_k\|x^l - \mu_k^l\|$$

or $\Delta w_{jk} = -\eta_1 \frac{\partial E^l}{\partial s_j^l(y_j^l)}.\frac{\partial s_j^l(y_j^l)}{\partial y_j^l}.\exp(-\frac{(\|x^l - \mu_k^l\|)^2}{2\sigma_k^2})$

$$=\eta_1 \sum_{j=1}^{M}(d_j^l - y_j^l)s_j^l(y_j^l).\sum_{k=1}^{K}\exp(-\frac{(\|x^l - \mu_k^l\|)^2}{2\sigma_k^2}$$

So, that $\Delta w_{jk} = \eta_1 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l)\exp(-\frac{(\|x^l - \mu_k^l\|)^2}{2\sigma_k^2})$ (2.8)

Now, from the equation (2.6) we have

$$\Delta \mu_{ki} = -\eta_2 \frac{\partial E^l}{\partial \mu_{ki}} = -\eta_2 \frac{\partial E^l}{\partial y_j^l}.\frac{\partial y_j^l}{\partial \mu_{ki}}$$

$$= -\eta_2 \frac{\partial E^l}{\partial y_j^l}.w_{jk}.\exp(-\frac{(\|x_i^l - \mu_{ki}^l\|^2)}{2\sigma_k^2}).(\frac{x_i^l - \mu_{ki}^l}{\sigma_k^2})$$

or $\Delta \mu_{ki} = \eta_2 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l)w_{jk}.\exp(-\frac{(\|x_i^l - \mu_{ki}^l\|)^2}{2\sigma_k^2})\ (\frac{x_i^l - \mu_{ki}^l}{\sigma_k^2})$ (2.9)

Now, from the equation (2.6) we have

$$\Delta\sigma_k = -\eta_3 \frac{\partial E^l}{\partial \sigma_k} = -\eta_3 \frac{\partial E^l}{\partial y_j^l}.\frac{\partial y_j^l}{\partial \sigma_k}$$

$$= -\eta_3 \frac{\partial E^l}{\partial y_j^l}.w_{jk} \exp\left(-\frac{(\|x_i^l - \mu_{ki}^l\|)^2}{2\sigma_k^2}\right)\frac{\|x_i^l - \mu_{ki}^l\|^2}{\sigma_k^3}$$

$$\text{or, } \Delta\sigma_k = \eta_3 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l).w_{jk}.\exp\left(-\frac{(\|x_i^l - \mu_{ki}^l\|)^2}{2\sigma_k^2}\right)\frac{\|x_i^l - \mu_{ki}^l\|^2}{\sigma_k^3} \qquad (2.10)$$

So that, we have from equations (2.8), (2.9) & (2.10) the expressions for change in weight vector & basis function parameters to accomplish the learning in supervised way. The adjustment of the basis function parameters with supervised learning represents a non-linear optimization problem, which will typically be computationally intensive and may be prove to finding local minima of the error function. Thus, for reasonable well-localized RBF, an input will generate a significant activation in a small region and the opportunity of getting stuck at a local minimum is small. Hence, the training of the network for L pattern pair i.e. (x^l, y^l) will accomplish in iterative manner with the modification of weight vector and basis function parameters corresponding to each presented pattern vector. The parameters of the network at the mth step of iteration can express as;

$$w_{jk}(m) = w_{jk}(m-1) + \eta_1 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l).\exp\left(-\frac{(\|x_i^l - \mu_{ki}^l\|)^2}{2\sigma_k^2}\right) \qquad (2.11)$$

$$\mu_{ki}(m) = \mu_{ki}(m-1) + \eta_2 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l).w_{jk}.\phi_k(x_i^l).\left(\frac{x_i^l - \mu_{ki}^l}{\sigma_k^2}\right) \qquad (2.12)$$

$$\sigma_k(m) = \sigma_k(m-1) + \eta_3 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l).w_{jk}.\phi_k(x_i^l).\frac{\|x_i^l - \mu_{ki}^l\|^2}{\sigma_k^3} \qquad (2.13)$$

where η_1, η_2 & η_3 are the coefficient of learning rate.

The discussed gradient decent approach for implementation of RBFNNs system is incremental learning algorithm in which the parameters update for each example (x^l, y^l). The RBFNNs trained by the gradient-decent method is capable of providing the equivalent or better performance compared to that of the multi layer feed forward neural network trained with the back propagation. The gradient decent method is slow in convergence since it cannot efficiently use the locally tuned representation of the hidden layer units. When the hidden unit receptive fields, controlled by the width σ_k are narrow for a given input only a few of the total number of hidden units will be activated and hence only these units need to be updated. Thus, there is no guarantee that the RBFNN remains localized after the supervised learning [28]. As a result the computational advantage of locality is not utilized. Indeed, in numerical simulations it is found that the subset of the basis functions may evolve to have very broad responses. It has been realized that some of the main advantages of the radial basis function network, is fast two-stage training and interpretability of the hidden unit representation.

Hence, among the neural network models, RBF network seems to be quit effective for pattern recognition task such as handwritten character recognition. Since it is extremely flexible to accommodate various and minute variations in data. Now, in the following subsection we are presenting the simulation designed and implementation details of redial basis function worked as a classifier for the handwritten English vowels recognition problem and compare the results with the back propagation algorithm for the MLP network.

3. Simulation Design and Implementation Details

The experiments described in this segment were designed to evaluate the performance of feed forward neural network when evolved with the back propagation algorithm for MLP & RBF network with decent gradient method.

3.1 Experiments

The parameters used for both experiments are described in Table 2 and 3.

Table 2: Parameters Used for Back propagation Algorithm.

Parameter	Value
Back propagation learning Rate (η)	0.1
Momentum Term (α)	0.9
Doug's Momentum Term $\left(\frac{1}{1-(\alpha)}\right)$	$\left(\frac{1}{1-(\alpha)}\right)$
Adaption Rate (K)	3.0
Minimum Error Exist in the Network $(MAXE)$	0.00001
Initial weights and biased term values	Randomly Generated Values Between 0 and 1

Table 3: Parameters Used for Decent Gradient -RBF Algorithm.

Parameter	Value
Back propagation learning Rate (η)	0.1
Momentum Term (α)	0.9
Doug's Momentum Term $\left(\frac{1}{1-(\alpha)}\right)$	$\left(\frac{1}{1-(\alpha)}\right)$
Adaption Rate (K)	3.0
Spread parameter σ	1.0
Mean of inputs(c)	Between maximum & minimum values
Minimum Error Exist in the Network ($MAXE$)	0.00001
Initial weights and biased term values	Randomly Generated Values Between 0 & 1

The task associated to the neural networks in both experiments was to accomplish the training of the handwritten English language vowels in order to generate the appropriate classification. For this, first we obtained the scanned image of five different types of samples of handwritten English language vowels as shown in figure (2). After collecting these samples, we partitioned an English vowel image in to four equal parts and calculated the density of the pixels, which belong to the central of gravities of these partitioned images of an English vowel. Like this, we will get 4 densities from an image of handwritten English language vowel, which we use to provide the input to the feed forward neural network. We use this procedure of generating input for a feed forward neural network with each sample of English vowel scanned images.

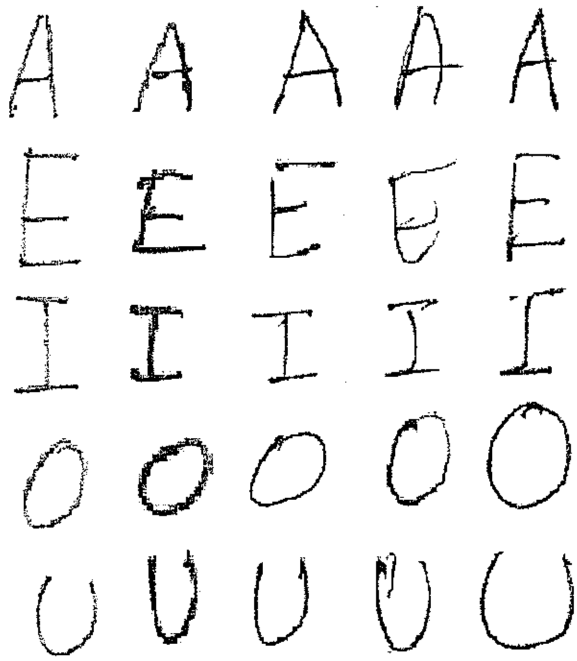

Figure 2: Scanned images of five different samples of handwritten English language vowels.

3.2. RBF Implementation in the Neural Network Architecture

The first neural network (NN1) structural design was based on feed forward multilayer generalized perceptron. Four input units have been used, with two numbers of hidden layers of six numbers of neurons and five numbers of neurons in output layer. The second neural network (NN2) structural design was also based on a completely connected feed forward multilayer generalized perceptron. But four input units have been used, with single hidden layer of six neurons and five neurons in output layer. The NN1 network is employing the sigmoid function for generating the output signal from the processing elements of all the hidden layers and output layer. The NN2 is using the same sigmoid function for the processing elements of output layer, but the Gaussian form of radial basis function is used for the hidden layer elements. The hidden layers were employed to investigate the effects with back propagation and decent gradient-RBF would have on the hyper plane. The MLP network has a single output layer with the following activation and output functions for the pattern vector (x_1, y_1)

$$y_j^l = \sum_{k=0}^{K} w_{kj} s_h^l(q_h^l) \tag{3.1}$$

and, $$s_j^l(y_j^l) = f[y_j^l] = f[\sum_{k=0}^{K} w_{kj} s_h^l(q_h^l)] \tag{3.2}$$

for $j = (1,2,.....,M)$ and $l = (1,2,....,L)$,where function $f[y_i^l]$ can define as,

$$s_j^l(y_j^l) = \frac{1}{1+e^{-Ky_j^l}} \tag{3.3}$$

Now, similarly, the output and activation value for the neurons of hidden layers and input layer can be written as,

$$q_h^l = \sum_{i=0}^{N} w_{ik} x_i$$

and, $$s_h^l(q_h^l) = f[q_h^l] = f[\sum_{i=0}^{N} w_{ik} x_i]$$, for $h = (1,2,......,K)$ (3.4)

$$s_h^l(q_h^l) = (\frac{1}{1+e^{-Kq_h^l}}) \tag{3.5}$$

In the Back propagation learning algorithm the change in weights are being done according to the calculated error in the network, after each, iteration of training. The change in weights and error in the network can be calculated as,

$$\Delta w_{kj}(t+1) = -\eta \sum_{k=1}^{K} \frac{\partial E^l}{\partial w_{kj}} + \alpha \Delta w_{kj}(t) + \frac{1}{1-(\alpha \Delta w_{kj}(t))} \tag{3.6}$$

$$\Delta w_{ik}(t+1) = -\eta \sum_{i=1}^{N} \frac{\partial E^l}{\partial w_{ik}} + \alpha \Delta w_{ik}(t) + \frac{1}{1-(\alpha \Delta w_{ik}(t))} \tag{3.7}$$

$$E^l = \frac{1}{2}\sum_{l=1}^{L}\left(d^l - s^l(y^l)\right)^2 \qquad (3.8)$$

Where $\left(d^l - s^l(y^l)\right)^2$ is the squared difference between the actual output value and the target output value of output layer for pattern l. Here, we have used the doug's momentum term [24] with momentum descent term for calculating the change in weights in eqn. (3.6) & (3.7). Doug's momentum descent is similar to standard momentum descent with the exception that the pre-momentum weight step vector is bounded so that its length cannot exceed 1 (one). After the momentum is added, the length of the resulting weight change vector can grow as high as 1 / (1 - momentum). This change allows stable behavior with much higher initial learning rates, resulting in less need to adjust the learning rate as training progresses.

Now, in the decent gradient learning for the RBF network the change in weights and basis function parameters can be computed as;

$$\Delta w_{jk}(t+1) = \eta_1 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l).s_j^l(y_j^l)\exp\left(-\frac{\|x_i^l - \mu_{ki}^l\|^2}{2\sigma_k}\right) + \alpha_1 \Delta w_{jk}(t) + \frac{1}{1-(\alpha_1 \Delta w_{jk}(t))} \qquad (3.9)$$

$$\Delta \mu_{ki}(t+1) = \eta_2 \sum\sum(d_j^l - y_j^l).s_j^l(y_j^l).w_{jk}.\exp\left(-\frac{\|x_i^l - \mu_{ki}^l\|^2}{2\sigma_k^2}\right).\left(\frac{x_i^l - \mu_{ki}^l}{\sigma_k^2}\right) + \alpha_2 \Delta \mu_{ki}(t) + \frac{1}{1-(\alpha_2 \Delta \mu_{ki}(t))}$$

(3.10) and

$$\Delta \sigma_k(t+1) = \eta_3 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l).s_j^l(y_j^l).w_{jk}.\exp\left(-\frac{\|x_i^l - \mu_{ki}^l\|^2}{2\sigma_k^2}\right).\left(\frac{x_i^l - \mu_{ki}^l}{\sigma_k^3}\right) + \alpha_3 \Delta \sigma_k(t) + \frac{1}{1-(\alpha_3 \Delta \sigma_k(t))}$$

(3.11)

Here, again we are using the Doug's Momentum Term with momentum decent term for calculating the change in weights and basis function parameters.

The redial basis function network has the single output and hidden layer with the following output functions for the pattern vector (x^l, y^l).

$$y_j^l = \sum_{k=1}^{K} w_{jk}\phi_k(\|x_i^l - \mu_{ki}^l\|) \qquad (3.12)$$

and $s_j^l(y_j^l) = f[y_j^l] = f[\sum_{k=1}^{K} w_{jk}\phi_k(\|x_i^l - \mu_{ki}^l\|)]$ (3.13)

for $j = (1,2,....,M)$ & $i = (1,2,...,N)$

Where function $f[y_j^l]$ can define as;

$$s_j^l(y_j^l) = \frac{1}{1+e^{-ky_j^l}}$$

and $$\phi_k(\|x_i^l - \mu_{ki}^l\|) = \exp(-\frac{\|x_i^l - \mu_{ki}^l\|^2}{2\sigma_k^2}) \quad (3.14)$$

4. Results

4.1 Results for First Trial of Simulation

Table 4: Results for Classification of Handwritten English Vowels using Back propagation for MLP and decent gradient with RBF network.

S. No.	Characters	Back propagation Epochs					DG-RBF Epochs				
		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	A	550	0.3	0.4	0.4	0.4	72	33	0.2	605	0.1
2	E	0.2	0.2	0.2	0.4	0.4	115	0.4	89	867	1525
3	I	0.4	0.2	0.4	0.4	0.4	5594	867	9232	0.4	0.3
4	O	0.4	0.1	0.4	564	0.4	89	0.1	0.4	1759	3179
5	U	0.4	0.4	0.4	0.4	0.4	674	933	97	29	8398

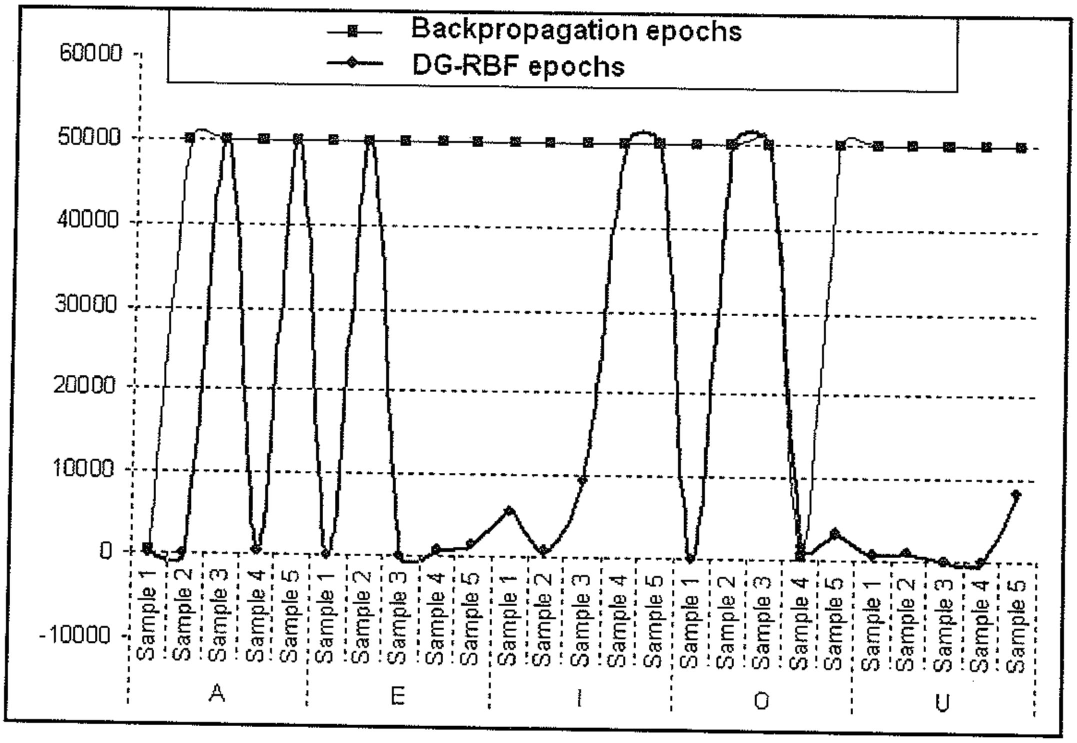

Figure 2: The Comparison Chart for Handwritten English Vowels Classification Epochs for two learning algorithms.

4.2 Results for Second Trial of Simulation

Table 5 : Results for Classification of Handwritten English Vowels using Back propagation for MLP and decent gradient with RBF network.

S. No.	Characters	Back propagation Epochs					DG-RBF Epochs				
		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	A	141	0.4	0.4	0.4	0.4	625	0.4	0.2	0.3	535
2	E	0.4	0.4	0.4	0.4	0.4	741	0.2	322	628	748
3	I	0.4	0.4	0.4	0.4	0.4	589	5339	72	1533	59
4	O	0.4	0.4	0.4	0.4	0.4	0.2	0.2	92	56	5208
5	U	0.4	0.4	0.4	0.4	0.4	78	0.3	41270	628	32

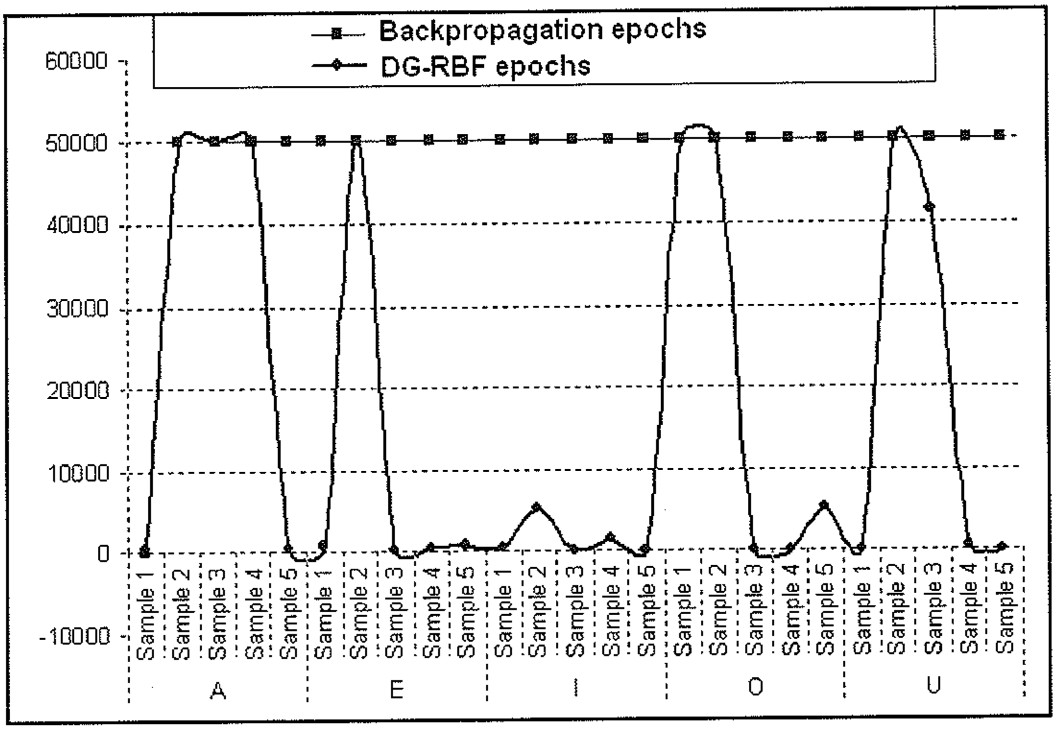

Figure 3: The Comparison Chart for Handwritten English Vowels Classification Epochs for two learning algorithms

4.3 Results for Third Trial of Simulation

Table 6: The Results for Classification of Handwritten English Vowels using Back propagation for MLP and decent gradient with RBF network.

S. No.	Characters	Back propagation Epochs					DG-RBF Epochs				
		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	A	180	0.4	0.4	0.4	0.4	0.2	748	1005	0.5	24748
2	E	15905	0.4	0.4	0.4	0.4	5611	0.2	0.3	0.3	166
3	I	8361	0.4	0.4	0.4	0.4	4135	34	412	0.2	0.3
4	O	0.4	0.4	0.4	0.4	0.4	826	0.1	0.2	11547	196
5	U	0.2	0.4	0.4	0.4	0.4	0.1	0.3	51	1143	0.3

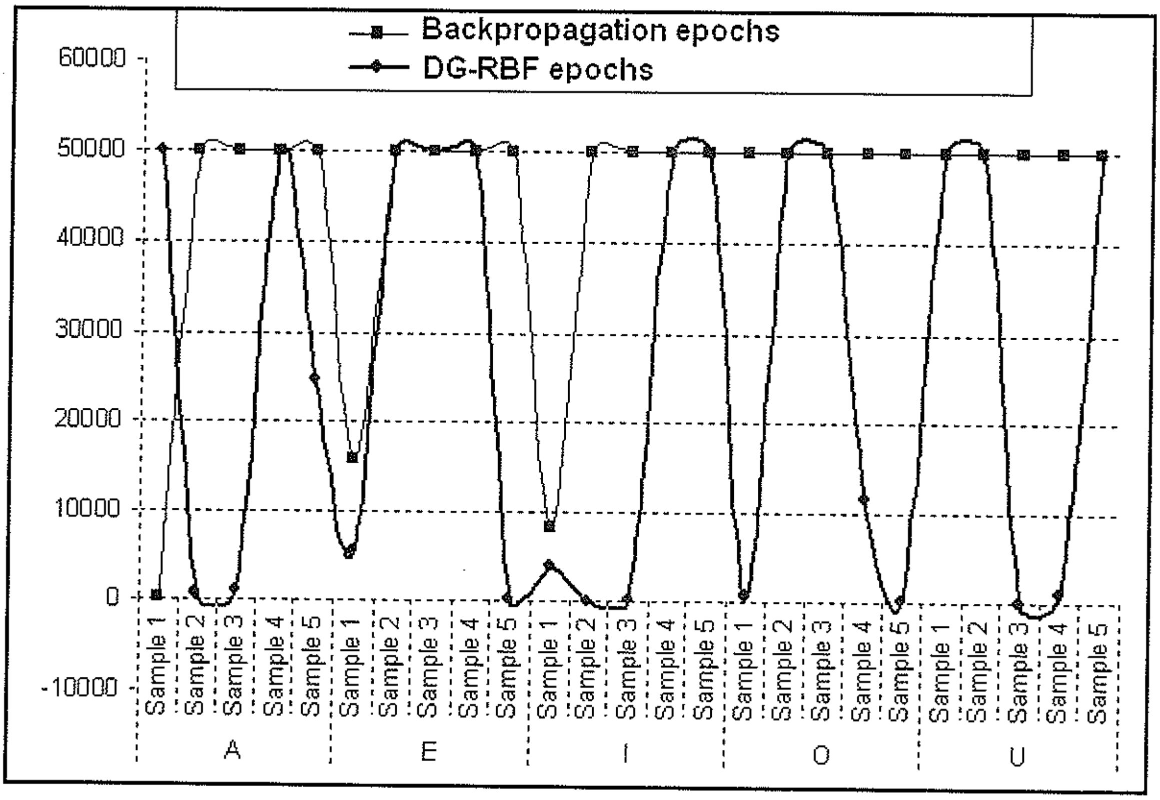

Figure 4: The Comparison Chart for Handwritten English Vowels Classification Epochs for two learning algorithms.

4.4 Results for Fourth Trial of Simulation

Table 7: The Results for Classification of Handwritten English Vowels using Back propagation for MLP and decent gradient with RBF network.

S. No.	Characters	Back propagation Epochs					DG-RBF Epochs				
		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	A	1696	0.4	0.4	0.4	0.4	4672	5134	0.2	0.1	0.3
2	E	12	0.4	0.4	0.4	0.4	97	0.4	3583	0.4	148
3	I	4909	0.4	0.4	0.4	0.4	532	232	0.3	0.2	0.2
4	O	1391	0.4	0.4	0.4	0.4	17248	0.4	0.1	2975	8282
5	U	147	1848	0.4	2369	0.4	5928	0.2	0.5	0.2	0.3

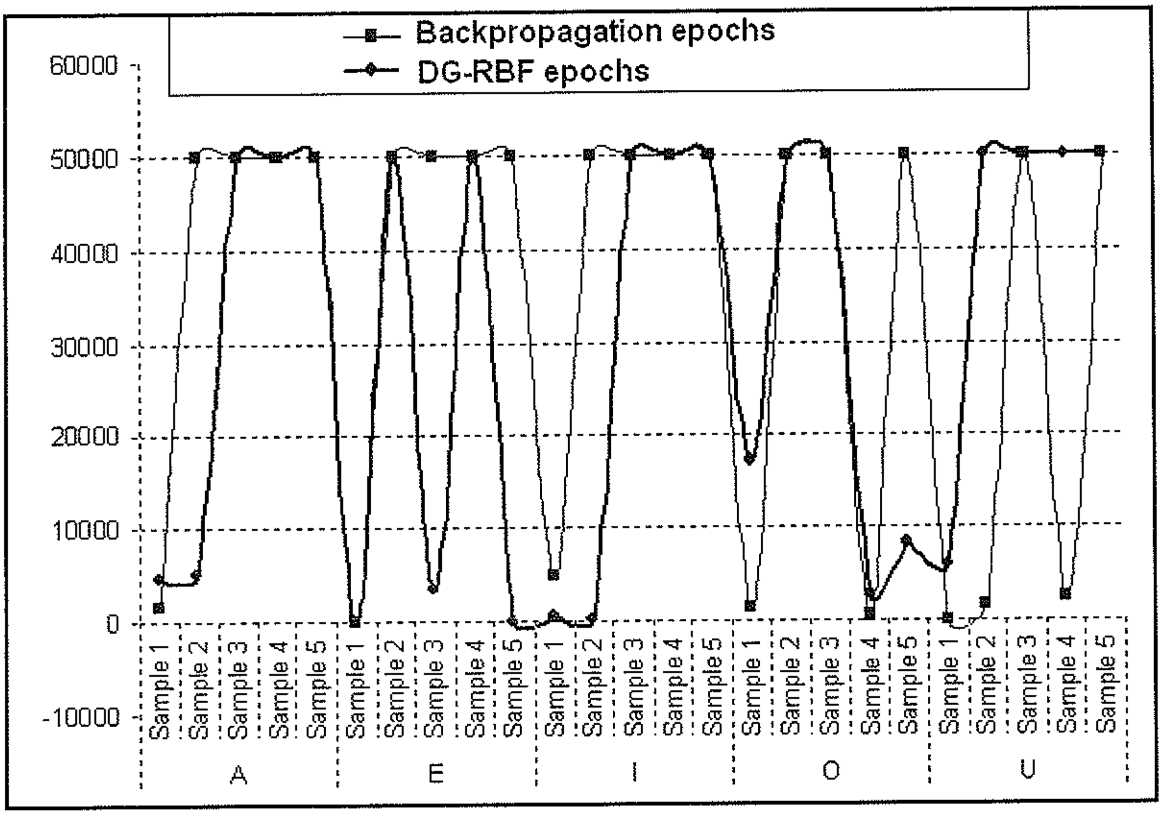

Figure 5: The Comparison Chart for Handwritten English Vowels Classification Epochs for two learning algorithms.

4.1 Results for Fifth Trial of Simulation

Table 8 : The Results for Classification of Handwritten English Vowels using Back propagation for MLP and decent gradient with RBF network

S. No.	Characters	Back propagation Epochs					DG-RBF Epochs				
		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	A	222	0.3	0.3	0.5	0.4	315	0.3	0.5	7067	391
2	E	0.1	0.1	0.1	0.4	0.4	0.1	124	88	0.3	0.2
3	I	35728	0.3	0.4	0.3	0.4	1231	2568	0.2	187	94
4	O	0.4	0.3	0.3	0.4	0.4	0.3	82	0.3	423	650
5	U	586	1848	0.4	2369	0.4	7296	18	39	1557	342

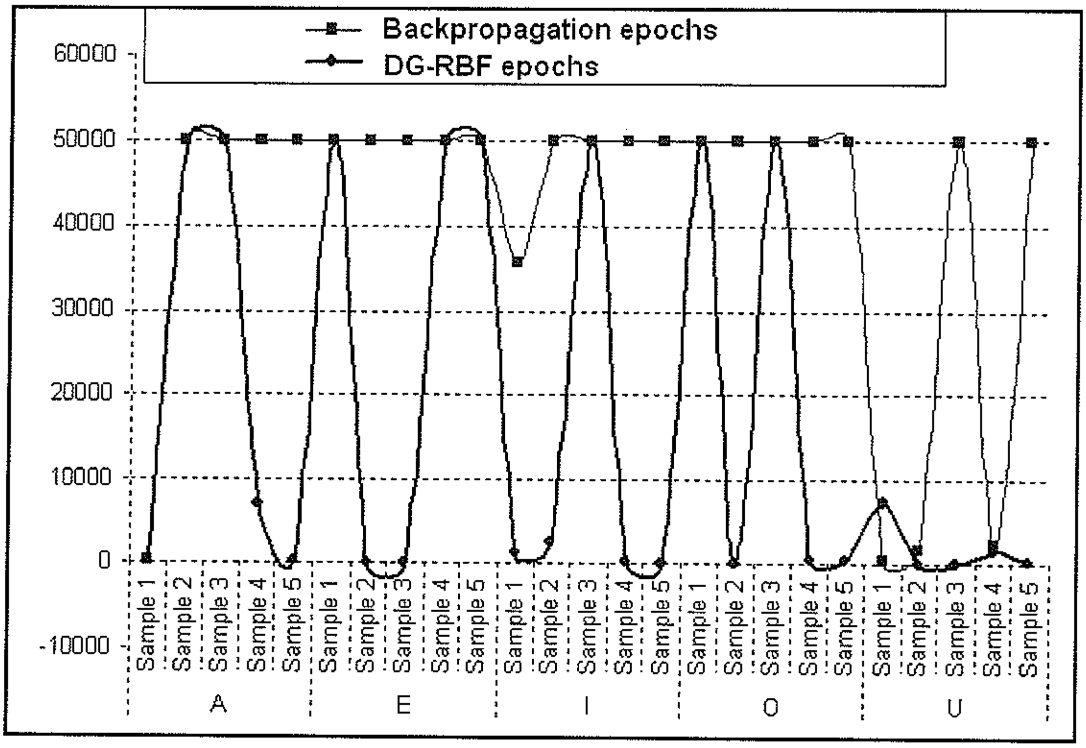

Figure 6: The Comparison Chart for Handwritten English Vowels Classification Epochs for two learning algorithms.

5. Discussion

The results presented in previous section are demonstrating the large significant difference exist between the performances of BPNN and DG-RBFN for handwritten English language vowels classification problem.

Tables 4, 5, 6, 7 & 8 are representing the results for handwritten English language vowels classification problem performed (5 times) with both algorithms up to the maximum limit of 50000 iterations. All the five results contain five different types of handwritten samples for each English vowel character. The training has been performed in such a way that repetition of same input sample for a character can not be happen simultaneously, i.e. if we have trained our network with a input sample of a character then next training can not be happen with the other input sample of the same character. This input sample will appear for training after other samples of other characters training.

Figures 2, 3, 4, 5& 6 are representing the comparison charts, designed on the basis of values available in the tables 4, 5, 6, 7& 8. These graphs are representing the graphical justification of the results shown in the tables.

It can observe from the results of tables and graphs that BPNN has converged conversing approximately for the 20 percent cases but the RBFN has converged for 75 percent cases.

The tables are showing some real numbers. These entries represents the error exit in the network after executing the simulation program up to 50000 iterations i.e. up to 50000 iterations the algorithm could not converge for a sample of a hand written English language vowels into the feed forward neural network.

The simulation program, which we have been developed in MATLAB 6.5, for testing these two networks for handwritten English language vowels classification problem, generates initial weights randomly through its random generator. So the epochs for the algorithms will be different every time with the same network structure and the same training data set.

6. Conclusions & Future Work

The results described in this paper indicate that, for the handwritten English language vowels classification problem, feed forward neural network trained with back propagation algorithm does not perform better in comparison to feed forward neural network trained with decent gradient with RBF. We found that, in each and every case, the DG-RBF network gives better results for the classification of English vowels, in comparison to the back propagation for the MLP network. It has been also observed that the RBF network has also stuck in local minima of error for some of the cases. The reason for this observation is quite obvious, because there is no guarantee that RBFNN remains localized after the supervised learning and the adjustment of the basis function parameters with the supervised learning represents a non-linear optimization, which may lead to the local minimum of the error function. But the considered RBF neural network is well localized and it provides that an input is generating a significant activation in a small region. So that, the opportunity is getting stuck at local minima is small. Thus the number of cases for DG-RBFNN to trap in local minimum is very low.

The direct application of DG-RBF to the handwritten character classification has been explored in this research. The aim is to introduce as alternative approach to solve the handwritten character classification problem. The results from the experiments conducted are quite encouraging and reflect the importance of radial basis function for the optimal classification to the given problem. Nevertheless, more work need to be done especially on the tests for large complex handwritten characters. Some future works should also be explored. For instances, current work is showing the performance of DG-RBF over back propagation for MLP network in classification process up to handwritten English language vowels but we can proceed further to use this idea for more complex handwritten character problems in which the character are not of the same size and may be not having the same rotation angle. The conjugate decent for the weights between hidden layer and output layer and for the parameters of basis function can also be calculated to increase the performance of the network for convergence as the extension work.

References

[1] M. Riedmiller, "*Rprop - Description and Implementation Details Technical Report*", University of Karlsruhe: W-76128 Karlsruhe (1994).

[2] K. Fukushima and N. Wake, "*Handwritten alphanumeric character recognition by the neocognitron*," IEEE Trans. on Neural Networks, 2(3) 355-365 (1991).

[3] Manish Mangal and Manu Pratap Singh, "*Analysis of Classification for the Multidimensional Parity-Bit-Checking Problem with Hybrid Evolutionary Feed-forward Neural Network.*" Neurocomputing, Elsevier Science, 70 1511-1524 (2007).

[4] D. Aha and R. Bankert, "*Cloud classification using error-correcting output codes*", Artificial Intelligence Applications: Natural Resources, Agriculture, and Environmental Science, 11(1) 13-28 (1997).

[5] Y.L. Murphey, Y. Luo, "*Feature extraction for a multiple pattern classification neural network system*", IEEE International Conference on Pattern Recognition, (2002).

[6] L. Bruzzone, D. F. Prieto and S. B. Serpico, "*A neural-statistical approach to multitemporal and multisource remote-sensing image classification*", IEEE Trans. Geosci. Remote Sensing, 37 1350-1359 (1999).

[7] J.N. Hwang, S.Y. Kung, M. Niranjan, and J.C. Principe, "*The past, present, and future of neural networks for signal processing*", IEEE Signal Processing Magazine, 14(6) 28-48 (1997).

[8] C. Lee and D. A. Landgrebe, "*Decision boundary feature extraction for neural networks,*" IEEE Trans. Neural Networks, 8(1) 75-83 (1997).

[9] C. Apte, et al., "*Automated Learning of Decision Rules for Text Categorization*", ACM Transactions for Information Systems, 12 233-251 (1994).

[10] Manish Mangal and Manu Pratap Singh, "*Handwritten English Vowels Recognition using Hybrid evolutionary Feed-Forward Neural Network*", Malaysian Journal of Computer Science, 19(2) 169 –187 (2006).

[11] Y. Even-Zohar and D. Roth, "A sequential model for multi class classification", In EMNLP-2001, the SIGDAT. Conference on Empirical Methods in Natural Language Processing, 10-19 (2001).

[12] Martin M. Anthony, Peter Bartlett, "*Learning in Neural Networks: Theoretical Foundations*", Cambridge University Press, New York, NY, (1999).

[13] M. Wright, "*Designing neural networks commands skill and savvy*", EDN, 36(25), 86-87 (1991).

[14] J. Go, G. Han, H. Kim and C. Lee, "*Multigradient: A New Neural Network Learning Algorithm for*

Pattern Classification", IEEE Trans. Geoscience and Remote Sensing, 39(5) 986-993 (2001).

[15] S. Haykin, "*Neural Networks*", A Comprehensive Foundation, Second Edition, Prentice-Hall, Inc., New Jersey, (1999).

[16] T. P. Vogl, J. K. Mangis, W. T. Zink, A. K. Rigler, D. L. Alkon "Accelerating the Convergence of the Back Propagation Method", Biol. Cybernetics, 59 257-263 (1988)

[17] R. A. Jacobs, "*Increased Rates of Convergence through Learning Rate Adaptation*", Neural Networks, 1 295-307 (1988).

[18] X.H. Yu, G.A. Chen, and S.X. Cheng, "*Dynamic learning rate optimization of the backpropagation algorithm*", IEEE Trans. Neural Network, 6 669 - 677 (1995).

[19] Stanislaw Osowski, Piotr Bojarczak, Maciej Stodolski, "*Fast Second Order Learning Algorithm for Feedforward Multilayer Neural Networks and its Applications*", Neural Networks, 9(9) 1583-1596 (1996).

[20] R. Battiti , "*First- and Second-Order Methods for Learning: Between Steepest Descent and Newton's Method*". Computing: archives for informatics and numerical computation, 4(2) 141-166 (1992).

[21] S. Becker, & Y. Le Cun, "*Improving the convergence of the back-propagation learning with second order methods*", In D. S. Touretzky, G E. Hinton, & T. J. Sejnowski (Eds.), Proceedings of the Connectionist Models Summer School. San Mateo, CA: Morgan Kaufmann, 29-37(1988).

[22] M. F. Moller, "A *Scaled Conjugate Gradient Algorithm for Fast Supervised learning*", Neural Networks, 6 525-533(1993).

[23] M.T. Hagan, and M. Menhaj, "*Training feed-forward networks with the Marquardt algorithm*", *IEEE Transactions on Neural Networks*, 5(6) 989-993(1999).

[24] P. Christenson, A.Maurer & G.E.Miner, "handwritten recognition by neural network", http://csci.mrs.umn.edu/UMMCSciWiki/pub/CSci4555s04/InsertTeamNameHere/handwritten.pdf, (2005).

[25] M. J. D. Powell, "*Radial basis functions for multivariable interpolation: A review*", in *Algorithms for Approximation of Functions and Data*, J. C. Mason, M. G. Cox, Eds. Oxford, U.K.: Oxford Univ. Press, 143-167 (1987).

[26] Guobin Ou, Yi Lu Murphey: "*Multi-class pattern classification using neural networks*". Pattern Recognition, 40 (1) 4-18 (2007).

[27] J. Moody and C.J. Darken. "*Fast learning in networks of locally-tuned processing units*", Neural Computation, 1(2) 281-294 (1989).

[28] T. Poggio, and F. Girosi, "*Regularization algorithms for learning that are equivalent to multilayer networks*", *Science*, 247 978-982 (1990b).

[29] M. T. Musavi , W. Ahmed , K. H. Chan , K. B. Faris , *D. M. Hummels*, "*On the training of radial basis function classifiers*", Neural Networks, 5(4) 595-603 (1992).

[30] D. Wettschereck, T. G. Dietterich, "*Improving the performance of radial basis function networks by learning center locations*". In J. E. Moody, S. J. Hanson, and R. P. Lippmann, (Eds.) *Advances in Neural Information Processing Systems*, San Mateo, CA: Morgan Kaufmann, *4* 1133-1140(1992).

[31] W. P. Vogt, "*Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences*", Thousand Oaks: Sage (1993).

[32] S. Haykin, "*Neural Networks*", Macmillan College Publishing Company, Inc, New York (1994).

[33] D. S. Broomhead and D. Lowe, "*Multivariable functional interpolation and adaptive networks*", Complex Syst, 2 321-355(1988).

[34] T. Poggio, F. Girosi, "*Networks for approximation and learning*", Proc IEEE **78**(9): 1481–1497(1990).

[35] F. Girosi, T. Poggio, "*Networks and the best approximation property*", Biol Cybern 63:169–176(1990).

[36] M.A. Kraaijveld and R.P.W. Duin, "*Generalization capabilities of minimal kernel-based networks*", *Proc. Int. Joint Conf. on Neural Networks (Seattle, WA, July 8-12)*, IEEE, Piscataway, U.S.A., I-843 - I-848(1991).

[37] S. Chen, C. F. N. Cowan, and P. M. Grant, "*Orthogonal least squares learning algorithm for radial basis function networks*", *IEEE Transactions on Neural Networks*, ISSN 1045-9227, 2 (2) 302-309(1991).

[38] C. M. Bishop, "*Novelty detection and neural network validation*", IEEE Proceedings: Vision, Image and Signal Processing. Special issue on applications of neural networks, 141 (4), 217–222(1994b).

[39] C. M. Bishop. "*Neural Networks for Pattern Recognition*", Oxford University Press, Oxford, New York, (1995).

Author's Biography

Dr. Manu Pratap Singh received his Ph.D. in computational physics from Kumaun University Nanital, Uthrakhand, India, in 2001. He has completed his Master of Science (M. S.) in Computer Science from Allahabad University, Allahabad in 1994. He is currently a Reader in Department of Computer Science, ICIS, Dr. B.R. Ambedkar University, Agra, UP, India since 2003. His research interests are focused on neuro-computing, neuroscience, neuro-informatics, soft-computing, etc. He is a member of technical committee of IASTED, Canada since 2004. He is being referee of various international/ national journals. In 2005, he received young scientist award from International Academy of Physical Sciences, Allahbad, UP, India. He has more than 45 papers in International and national journals and has supervised several doctoral and M.Tech theses.

Mr. Naveen Kumar Sharma has submitted his thesis of Ph.D. in Computer Science in the Department of Computer Science, ICIS, Dr. B.R. Ambedkar University, Agra, UP, India, in 2008. He has completed his Integrated M.Sc in Computer Science & Mathematics from Dr. B.R. Ambedkar University, Agra, UP, India, in 2001. He is currently working as an Assistant Professor in Department of Computer Science & Engineering, CET IILM AHL, Greater Noida, affiliated to UPTU, Lucknow, U.P., India since 2006. His research interests are focused on neuro-computing, neuroscience, pattern recognition, etc. He has 5 papers in International and national journals and has supervised several projects for the B.Tech & MCA students.

S. R. Pande is currently Associate Professor of Computer Science and Head, Department of Computer Science, S.S.E.S. Amt's, Science College,R.T.M. Nagpur University, Nagpur, Maharashtra state, India. He is currently pursuing his Ph.D in Computer Science from R.T.M. Nagpur University, Nagpur. He has about 19 years of experience in teaching and research. His areas of interest are Data structure, Discrete mathematics, Theory of Computations, Neural Network, Fuzzy logic, Computer Graphics, Image Processing, Relational DBMS,C++.