IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol-6, Special Issue-4, May 2018 E-ISSN: 2347-2693

Evolution of Feed Forward Network for solving Classification and Prediction Problems

M. Sornam[1*], P. Balamurugan[2]

[1*]Dept. of Computer Science, University of Madras, Chennai, India
[2]Dept. of Computer Science, University of Madras, Chennai, India

[*]*Corresponding Author: madasamy.sornam@gmail.com, Tel.: +91-044-25399628*

Available online at: www.ijcseonline.org

***Abstract*—**Over the past decade, ANN is used in many fields including Engineering and Medical electronics. ANN is also been applied to solve many problems of classification and prediction. Depending on the problem space and complexity various approaches were proposed to solve the problems in an efficient way. Multi-layer Feed forward network is one of the network architecture predominantly used to solve classification and prediction problems. The objective of this paper is to study the various methods available in the literature for solving those problems. The study starts with a simple feed forward network for image classification, then continued to investigate the methods to improve the classification accuracy using various wavelets and dimensionality reduction techniques. The various improvements were proposed for Backpropagation algorithm including complex BP were analysed. For performance improvement, methods of evolutionary algorithms and Pruning techniques were studied briefly. Finally the improved RBF network for complex numbers was analysed. This paper gives an overall idea of how the feed forward network was evolved with various approaches for solving classification and prediction problems.

***Keywords*—**Feed Forward, Classification, Predication, Backpropogration, PSO, RBF, Complex valued ANN.

I. INTRODUCTION

In recent years, the scope Artificial Neural Networks (ANN) has been extended to solve many application related problems. ANN has an ability to learn and the learning is used to solve problems like classification, recognition, prediction etc. Feed Forward network is one of the widely used neural network model. Over the last decade lot of research work has been done over feed forward network to solve the problems more accurately and effectively. The various approaches includes network optimization, usage of feature extraction mechanisms, new activation functions, extending the networks to use complex values etc. The aim of this paper is to study the available methods and their limitations.

Rest of the paper is organized as follows, Section I contains the introduction about the survey paper , Section II briefly explains about the literature study of various approaches applied in Feed Forward Network for solving classification and prediction problems, Section III contains the proposed plan of work based on the survey, Section IV concludes with the summary of the study.

II. SURVEY OF FEED FORWARD NETWORKS

A. Simple Feed Forward Network

In the past, various machine learning classifiers such as Support Vector Machine, K-Nearest neighbor etc., were used for solving image classification problems. A simple Feed Forward network was employed using Backpropagation algorithm with Logistic Sigmoid Activation Function for image classification [1]. The network is trained to classify the images of different classes. The image pixels are defined with statistical attributes such as Mean, Variance, Skewness, Kurtosis, Maximum Probability, Correlation, Contrast, Uniformity, Homogeneity and Entropy. These statistical attributes are used as Feature Vector for the given image. The primary step is to extract the features from the given image. Better accuracy was achieved with the proposed approach. Though the accuracy is increased when compared to machine learning classifiers, there are some drawbacks also. The computational complexity is increased, more time is taken to train the network. The network failed to classify an image contain more than once class.

B. Radial Basis Function and Wavelet Network

Radial Basis Function (RBF) Networks has a simple network architecture and provides lot of benefits in terms of convergence speed, usage of non-linear function [2]. RBF

© 2018, IJCSE All Rights Reserved

based networks are employed for solving various problems including classification, predication etc. Face Recognition Based on Wavelet Packet coefficients using Radial Basis Function Neural Network is experimented [2]. Averaged wavelet packet coefficients are used to represent the face image. Level-2 decomposition is used. At level-1, the image is decomposed into approximation coefficients and detailed coefficients using Wavelet decomposition. At level-2, the approximation coefficients are again decomposed into approximation coefficients and detailed coefficients. As a result, four Wavelet Packet Coefficients (WPC) are obtained viz., approximation, horizontal details, vertical details and diagonal details. The obtained wavelet coefficients are averaged. Two methods were discussed for averaging the wavelet coefficients. (i) Average Wavelet Packet coefficients and (ii) Averaged WPC for classes of face images. The averaged WPC is fed to RBF network for training andrecognition. Gaussian function was used as a basis function in the hidden layer. The experiment was conducted with 3 different datasets – ORL, JAFFE and Essex Face. The experiment concluded with promising results. The computational complexity is reduced because of the usage of reduced dimension with improved recognition rate.

Pattern recognition using Radial Basis Network with reduced features using Wavelet Transform and Discriminative Common Vector (DCV) was discussed [3]. In the given input pattern, wavelet transform is applied to get Wavelet coefficients which are then used for feature extraction. Using the obtained wavelet coefficients, DCV is computed using within-class scatter matrix method. The computed DCV is classified by using Radial Basis Network. Wavelet transform is accurate in representing the image with lighting changes and capturing significant face expressions. Since approximation coefficients exhibits the change in pose and scale, only that coefficients are considered for feature set. The feature set may have irrelevant information. So the discriminant features are extracted using Discriminative Common Vector (DCV). For DCV computation within-class scatter matrix method was used. With the reduced feature set, the result shows that the computational complexity is reduced along with increased recognition rate.

Artificial Neural Network was employed predominantly to solve classification problems. In many applications especially in medical electronics, there are high rate of False Positive (FP) and False Negative (FP). Wavelet Neural Network (WNN) reduces FP and FN cases significantly [5]. WNN is based on wavelet transform theory and an alternate to the feed forward neural networks for approximating arbitrary nonlinear functions. In WNN, the neurons are called as wavelons,which is a combination of affine transformation and a multidimensional wavelet.

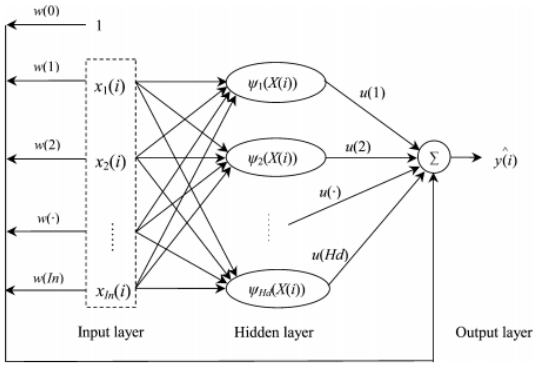

Figure 1. A typical wavelet network with single hidden layer [4]

Neural networks have extended to solve medical electronics problems. Cancer diagnosis and biopsy classification is one such problem to consider. WNN based approach was applied to solve such problem [5]. Morlet, Shannon and Polywog wavelets were used to train WNN. In their proposed method, results were compared with Feed forward network using Levenberg-Marquardt backpropagation algorithm. From the result, it is inferred that WNN shall also be applied to solve pattern classification and function approximation problems. It is also observed that WNN gives better convergence than feed forward networks. Also WNN is able to train the network much faster than feed forward networks

C. Document Analysis and Recognition

The other area in which ANN gets popularized is document analysis and recognition field. This is more like image classification and pattern recognition, but the key challenge here is to recognize the characters, words or the document. This kind of recognition is essential for application like machine translation, and paraphrasing. Also this problem is language specific. A method of recognition of Tamil word available in PDF format using single hidden layer feed forward neural network was designed [6]. The PDF documents are converted to JPEG image and Wiener filter is applied to reduce high frequency noise and then the image is binarized using Niblack's approach. With the binary vector, the skew angle was corrected to provide font independence. The next stage is to extract the feature set using segmentation technique. The structural features of the words such as number of loops, horizontal lines, vertical lines etc. are extracted. The method also handles touching characters elegantly using separating procedure. With the extracted feature set, the network is trained. Each character in the Tamil alphabets are assigned with a unique code (character code). The training dataset is converted to unique character code sequence after segmenting and passed as input to the network for training. The proposed method

© 2018, IJCSE All Rights Reserved

experimented with Backpropagation algorithm. The trained network is able to recognize Tamil accurately with no deviation. The proposed method is not dependent on any language expect the character code mapping with Tamil alphabets. To apply this method to other languages, only the character code mapping has be modified to corresponding language's alphabet list.

The following table list the recognition rate for the approaches discussed above:

Table 1. Recognition Rate of various ANN Approaches

Method	Application	Recognition Rate
Simple Feed Forward Network with BP	Image Classification	97.2%
Wavelet	Face Recognition	97%
Wavelet + DCV	Face Recognition	98%
WNN	Cancer Diagnosis & Biopsy classification	98%
Segmentation + unique character code	Document Analysis	100%

Note: The application and the dataset considered for each methods are different. This table just gives an idea about the level of recognition produced with ANN

D. Backpropagation Algorithm

Backpropagationis the most widely used learning or training algorithm in neural networks. At the output layer, the error value is calculated with actual and desired outputs and the error is back propagated to the network for synaptic weight update. There are lot of applications solved with the usage of backpropagation. In a sound based system, echo is an unwanted noise that disturbs the pure speech signal. A backpropagation based feed forward neural network with unipolar sigmoid activation function was used to suppress the echo was discussed by Soumava et al [7]. The aim is to create a filter to remove the echo from pure speech signal. For identifying the correlation between echo and pure speech. The network was trained with pure and echo signals. The trained ANN model is expected to estimate the amount of echo or noise present. Based on the estimation the filter shall suppress the echo. From the result it was concluded that the trained network is able to correlate echo and pure speech signal effectively.

The trained ANN model can also be used for predication. ANN was employed to predict traffic accident casualties for the given year by Galal et al [8]. The ANN model is trained with past historical data and based on the trained data, the network is expected to predict. From the collected data set, a few independent variables along with year are selected as input and the dependent variable number of casualties is calculated based on the inputs. With the help of regression analysis, the independent variables were plotted over the plane and curve-fitting equations was arrived. And that equation was used to predict traffic causalities. The predicted results seems to be very favourable. The difference between the estimated and actual are almost to a near range.

The major limitation of Backpropagation algorithm is slow convergence and local minima. To solve these limitations, an improved three-term optical backpropagation algorithm was proposed [9]. In output layer based on error, a linear function on error term is generated and used as a 3rd term in the weight update formula. The error gets adjusted by using a non-linear function and is then backpropagated to hidden neurons. Along with other parameters (learning rate and momentum) the new third term is used to change the synaptic weight. The improved algorithm is tested with standard benchmarked problems such as XOR, symmetry detection, Rotation (Geometric transformation), similarity transformation, and function approximation. It was observed that the number epochs used to train the network was reduced considerably when compared to OBP. The implementation of this new improved algorithm is very simple, robust and can be applied with any kind of activation functions.

E. Evolutionary Approach

The efficiency of the network is determined by the parameters such as number of hidden layer nodes, activation function, the learning rule etc. used in the network. To improve the efficiency, many researches considered to use optimized parameters to design the network architecture. Particle Swarm Optimization (PSO) is a heuristic optimization algorithm trying to optimize the problem iteratively by improving the local solution with respect to the given fitness function. So it is apparent that we can use PSO to optimize the network parameters which in turn improve the efficiency of the network. For classifying the educational dataset, an optimized neural network architecture was designed by Devika et al [10] using PSO algorithm. Before training the network, the parameters are optimized using PSO. Then the optimized parameters are used to evaluate the performance of the optimized network. Three different activation functions are considered, uni-polar sigmoid, Bi-polar sigmoid and hyperbolic tangent function. The maximum hidden neurons are fixed based on number of input and output neurons. After the experiment, it is inferred that bi-polar sigmoid activation function with Learning Rate either 0.7 or 0.8 and hidden neurons as 4 or 5 or 6 gives an accuracy of 92.1671% and precision of 86.6681% when compared to others considered parameters. The limitation with this approach is the optimal parameters applicable to the concerned application.

As discussed earlier, RBF network outperforms for prediction problems. Even for RBF based networks, the parameters are initialized randomly. The random value has direct impact on convergence, number of iterations. So instead of using the random numbers, usage of optimized

© 2018, IJCSE All Rights Reserved

initial parameter values shall improve the efficiency of the system. The RBF network optimized with PSO was discussed for Internet traffic prediction [11]. Network operators shall predict the traffic for the purpose of network management, security and network control. The prediction accuracy was improved by optimizing the parameters used in RBF neural network. The parameters used in the RBF network are optimized using PSO. The optimized parameters used in the RBF network are then used for traffic prediction. The results shows that the computational complexity was reduced significantly with the improvement in prediction accuracy. This method shall be extended to have a prediction in complex networks.

PSO is a simple algorithm with no gradient information. Though PSO carries lot of benefits, it also has some short comings such as the speed to derive the optimal values are slow and so the convergence. In order to overcome the short comings of PSO a new improved version of PSO was suggested by Xiaobing et al [12]. Tabu search was used to improve the operating efficiency of PSO. During search process, Tabu search applied to reduce the convergence time. Tabu search avoids local minima by accepting the worst moves and avoid coming back to previously visited solutions by remembering the visited solution for short duration of time. Tabu search algorithm is used to complete the remaining down search and the search start with feasible solutions and find neighbourhood solutions. The performance of the algorithm is also depends on neighbourhood norms.

F. Pruning Algorithms

The other technique to optimize the network architecture is to use Pruning algorithms. The pruning algorithm is a process of trimming a network with unwanted parameters, neurons to improve the performance and efficiency of the network. There are various pruning technique available in the literature. A comparative study of existing pruning algorithms were done in [13]. The pruning approaches are broadly classified into brute force, constructive and destructive. In Penalty method, a penalty value is added with objective functions for weight updates and the smaller values tends to 0. Redundant weights are removed. Effects of other weights are also considered. Gauss Schmidt used to determine principal nodes and are given a chance for weight updation. In cross validation method, the dataset is divided into training and cross-validation set. At each phase of pruning, the cross validation set is used to validate the pruned network and decide upon the optimality of the network. This improves the generalization capacity of the networks. In Magnitude based, the assumption is the lesser weights are irrelevant and remove them. But it may be possible that those weights are important for some cases. In Mutual Information based, singular value decomposition (SVD) is used to compute covariance matrix for hidden layer. The rank of the covariance matrix is used to determine the number of nodes to be used in Hidden node. Here irrelevant units in both input and hidden layers are pruned. In Evolutionary method, Genetic Algorithm (GA) is used for pruning. Different networks are created based on GA operations - mutations, reproduction and cross-over. In Sensitivity based method, the weights or nodes that have least effect on objective function are pruned. Optimal Brain Damage (OBD) and Optimal Brain Surgeon (OBS) are famous sensitivity based pruning techniques. Early stopping procedure is employed to stop the learning process when the error starts to increase. In Significance measure based method, the parameters with the significance value less than threshold value are pruned.

A novel pruning algorithm was proposed by Gethsiyal et al [14] for optimizing feed forward networks for classification problem set. N2PS - Neural Network Pruning by Significance. The significant value is computed by using sigmoidal activation value. The parameters which are to be eliminated are determined by the significant value. If the significant value is less than threshold,then eliminate. The significant value is computed for hidden and input layers. Various pruning algorithms N2PS, VNP, Xing-Hu's method, MBP, OBD and OBS are applied over iris, cancer, hepatitis and diabetes datasets for evaluating the performance.

Table 2. Result of comparison of N2PS with Xing-Hu Approach

Dataset	Original		Xing-Hu		N2PS	
	NN	Acc	NN	Acc	NN	Acc
Iris	4-10-3	96	3-2-3	98.67	3-3-3	98.67
Cancer	9-10-2	95.4	3-3-2	96.78	3-2-2	97.1
Hepatitis	19-25-2	80.2	3-8-2	84.62	2-3-2	86.4
Diabetes	8-40-2	68.6	6-8-2	74.22	5-3-2	70.3

From the table [14], it is evident that N2PS shows better result. And the number of nodes in input and hidden layers are considerably reduced. Also the generalization ability is also increased.

G. Complex Valued Artificial Neural Network

ANN has extended its scope to various engineering domains such as signal processing, communication engineering, and medical imaging. Typically these domains use complex numbers to represent the attribute values. So it is essential for ANN to learn processing values in complex valued domain. The complex valued neural network are the networks whose input, output, thresholds, weights and bias are complex values. Enhancing artificial neural networks using complex numbers was discussed in [15]. The implementation is based on simple perceptron model. The given input is first mapped to specific range of input. The complex valued activation functions uses the magnitude of the complex value for computation. The magnitude is compared with the threshold value for output neuron activation. The output shall be in real valued domain. The weight updation is based on new

© 2018, IJCSE All Rights Reserved

learning rule which is partial derivative. The partial derivative was used because of the nature of activation function. The bias used in the network can either be real or complex. An experiment was conducted over Boolean logic gates and the results are promising over real valued network. The advantages of complex valued network are reduced network size, fast learning, single layer is enough to solve linearly non-separable problems with no additional cost.

Tohru Nitta [16] proposed Complex Backpropagation algorithm for complex values in neural networks. In complex valued neural network, the geometric transformation is applied over a point P(x, y) to P'(x', y'). The network was trained and has given correct transformation. Real BP fails to give correct transformation. The behaviour of the complex BP resembles the Identity Theorem of complex analysis.

Identity Theorem: Let F and G be regular functions over the complex domain D. If F(z) = G(z) on a given line then F(z) = G(z) over D identically. The training on the straight line in complex function F. The network approximates F based on the training point and results in other complex function G. Hence G(z) = F(z) which satisfies the Identity Theorem.

As briefed above, it was proven that ANN can be used to solve prediction related problem elegantly. For now the problem solving approach was applied to predict output values for the given input based on the training. But there are certain cases where the output is known and it is expected to predict the input values. Here the word prediction is used because for an output value there may be a possibility of more than one input values, so the correct input value need to be predicted. Such application are used in robotics and mechanics problem solving. In literature, the inversion procedure was applied to many problems with real valued domain. ANN has a capability to learn complex number patterns naturally. Hence inversion was applied over complex valued neural networks with the usage of complex backpropagation algorithm [17]. There are 3 classes of inversion algorithms viz., exhaustive search, multi-component evolutionary method and single element search method. Nearest inversion approach was used to find the inverse solution lying to the nearest point. Split sigmoid activation function was employed in the inversion process. The process was experimented with the standard benchmark problems – XOR, similarity transformation and anti-clockwise rotation. For all these standard problems, the inversion solution was able to be computed. The inversion values are close to the actual input values.

In the prior discussion, RBF based networks on real valued domain works efficiently to solve various application related problems. When RBF is used in engineering and medical problems with complex numbers, real valued RBF is not relevant. So a new approach of fully complex-valued radial basis function (FC-RBF) network was proposed [18]. The FC-RBF uses complex valued activation function (*sech*) and its complex-valued gradient descent learning algorithm. The proposal was evaluated using standard problem sets – XOR problem, synthetic function approximation and two-spiral classification. Compared to the existing complex-valued RBF such as CRBF, CMRAN and CELM, the FC-RBP provides better convergence, approximation and classification ability.

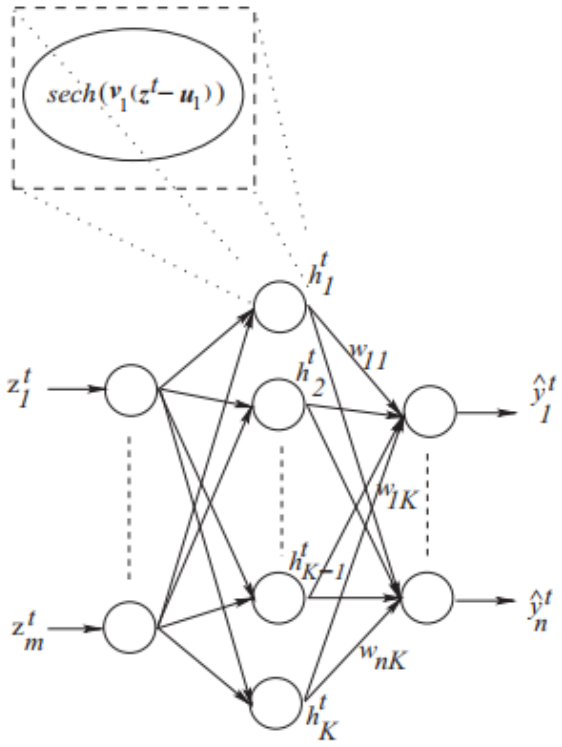

Figure 2. The fully complex-valued RBF network [19]

III. WORK PLAN

Based on the literature study, it is evident that ANN can be applied to solve classification and prediction problems efficiently. At end of this study, the following works are planned: (1) design an optimal complex valued network using optimization techniques such as PSO or Pruning algorithm (2) propose an improvised optimization technique (3) generate a new activation function for solving classification and prediction problems. The planned work can be evaluated by solving problems from medical electronics or speech synthesis or network security domains.

IV. CONCLUSION

The literature study gives an overview of how feed forward networks have been proposed with new methods for solving problems of various fields. The initial proposals focused on classification rate with the trade-off of performance,

© 2018, IJCSE All Rights Reserved

efficiency and generalization. After reaching the desired accuracy, the focus got diverted to solve the limitations by using evolutionary methods, Pruning techniques etc. There was a paradigm shift from real valued networks to complex valued networks for solving engineering and medical application problems. The literature has lot of design alternatives, this paper has studied a few of them. From the study it is evident that ANN can be applied to solve any kind of problems efficiently and effectively with respond to the environment changes.

REFERENCES

[1] K. Verma, L. K. Verma and P. Tripathi, "*Image Classification using Backpropogation Algorithm",* JOURNAL OF COMPUTER SCIENCE AND SOTWARE APPLICATION, Vol. **1**, No.**2**, 2014

[2] T. Kathirvalavakumar, J. B. Vasanthi, "*Face Recognition Based on Wavelet Packet Coefficients and Radial Basis Function",* Journal of Intelligent Learning Systems and Applications, 2013

[3] T. Kathirvalavakumar, J. B. Vasanthi, "*Features Reduction using Wavelet and Discriminative Common Vector and Recognizing Faces using RBF",* International Journal of Computer Applications, Vol. **74**, No.**5**, 2013

[4] Y. Xie and Y. Zhang, "*A Wavelet Network Model for Short-Term Traffic Volume Forecasting"*, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, Vol. **10**, No.**3**, pp.**141-150**, 2006

[5] A. Sinha, M. Pavithra and B. J. Pandian, "*A Wavelet Neural Network based approach in Cancer Diagnosis and Biopsy Classification",* International Journal of Advancements in Research & Technology, Vol. **2**, Issue.**8**, 2013

[6] M. Karthigaiselvi, T. Kathirvalavakumar, "*Recognition of Words in Tamil Script Using Neural Network"* , Int. Journal of Engineering Research and Application, Vol.**7** Issue.**3**, (Part-6), pp.**62-70**, 2017

[7] S. K. Roy, C. F. Rodrigues, "*Echo Canceller Using Error Back Propogration Algorithm",* International Conference on Soft Computing and Machine Intelligence, 2014

[8] G. A. Ali and A. Tayfour, "*Characteristics and Prediction of Traffic Accident Casualties in Sudan Using Statistical Modelling and Artificial Neural Networks",* International Journal of Transportation Science and Technology, Vol.**1** No.**4**.**202**, pp.**305-317**, 2012

[9] M. Sornam and P. Thangavel, "*An improved three-term optical backpropogation algorithm",* Int. J. Artificial Intelligence and Soft Computing, Vol.**2**, No.**4**, pp.**321-333**, 2011

[10] D. Chhachhiya, A. Sharma and M. Gupta, "*Designing optimal architecture of neural network with particle swarm optimization techniques specifically for educational dataset",* 7th International Conference on Cloud Computing, Data Science & Engineering, 2017

[11] T. He, T. Dan, Y. Wei, H. Li, X. Chen, G. Qin, "*Particle Swarm Optimization RBF Neural Network Model for Internet Traffic Prediction",* International Conference on Intelligent Transportation, Big Data & Smart City, 2016

[12] X. Liu, X. Jiao, Y. Li and X. Liang, "*Improved New Particle Swarm Algorithm Solving Job Shop Scheduling Optimization Problem",* 3rd International Conference on Computer Science and Network Technology, 2013

[13] M. G. Augasta, and T. Kathirvalavakumar, "*Pruning algorithms of neural networks – a comparative study",* Cent. Eur. J. Comp Sci. Vol. **3(3)**, pp.**105-115**, 2003

[14] M. G. Augasta, and T. Kathirvalavakumar, "*A Novel Pruning Algorithm for Optimizing Feedforward Neural Network of Classification Problems",* Springer, 2011

[15] T. Nitta, "*A Back-propagation Algorithm for Complex Numbered Neural Networks",* Internal Joint Conference on Neural Networks, 1993

[16] H. E. Michel and A.A.S. Awwal, "*Enhanced Artificial Neural Network using Complex Numbers",* IEEE, 1999

[17] A. S. Gangal, P.K. Kalra, and D.S. Chauhan, "*Inversion of Complex Valued Neural Networks using Complex Back-propagation Algorithm",* International Journal of Mathematics and Computers in Simulation, Vol.**3**, Issue.**3**, 2009

[18] R. Savitha, S. Suresh and N. Sundararajan, "*A Fully Complex-valued Radial Basis Function Network and its Learning Algorithm",* International Journal of Neural Systems, Vol.**19** No.**4, 99**, pp.**253-267**, 2009

[19] R. Savitha, S. Suresh, N. Sundararajan andH. J. Kim, "*A fully complex-valued radial basis function classifier for real-valued classification"*, Neurocomputing, Vol.**78**, Issue.**1**, pp.**104-110**, 2012

Authors Profile

*Dr M. Sornam*received her MSc in Mathematics from the University of Madras in the year 1987, Master's Degree in Computer Applications from the University of Madras in the 1991 and received her Ph.D in the year 2013 from the University of Madras. Since 1991–1996, she worked as a Lecturer in Computer Science at Anna Adarsh College, Chennai. Later, from 1996 to 2000 she worked as a Lecturer in Computer Science at T.S. Narayanasami College of Arts and Science, Chennai. Since 2001, she has been working in the Department of Computer Science, University of Madras. At present, she is working as an Associate Professor in Computer Science at the University of Madras. Her area of interest includes artificial intelligence, artificial neural networks, image processing, data mining, pattern recognition and applications.

P. Balamurugan received his M.S in Software Engineering from Birla Institute of Technology and Science, Pilani and currently pursuing M.Phil in Department of Computer Science at University of Madras, Chennai, India. His area of interest includes artificial intelligence, artificial neural networks, evolutionary computing, cryptography and image processing.

© 2018, IJCSE All Rights Reserved