Proc. of Int. Conf. on Computer Electronics Electrical Mechanical and Civil

Improved Classification Performance in Imbalanced Dataset using Projection based Learning Algorithm with Fuzzy Radial Basic Function

S. Padma [1] and R. Pugazendi [2]
[1]Research Scholar, Research and Development Center, Bharathiyar University
Coimbatore, Tamilnadu, India
padmacde@gmail.com
[2]Asst Prof, Department of Computer Science, Government Arts College
Salem, Tamilnadu, India

***Abstract*—The article gives an overview of the Radial Basis Function which is combined with Fuzzy C-Means algorithms and its learning process made by Projection Based Learning which is pointed out as PBL-FRBF The Projection Based Learning decreases the learning time, finds optimum output weight by its energy function and it prefers small amount of samples for testing. The performance of a classification content not only depends on learning algorithm selected but it also depends on the selection of dataset. Performance analysis is evaluated by benchmark datasets for classification problem from the UCI machine learning repository of the four datasets two of them are termed to be as balanced and the remaining as imbalanced dataset. The performance of the proposed PBL-FRBF has produced superior results compared with Fuzzy Radial Basis Function and Radial Basis Function for classification problems.**

***Index Terms*— C-Means, Projection Based Learning, Radial Basis Function, Classifications, Neural Networks, balanced dataset, imbalanced dataset.**

I. INTRODUCTION

Neural Networks extensively applied for classification and employed to solve real word problems[1]. Classification works with an objective of assesing both the input and output features to provide a suitable class label. Radial Basis Function Neural Network[2] has been widely used for solving classification problems and function approximations[3]. The RBF neural network[4] can be trained and by means of two stages, first, by finding the centers of the hidden neurons c_i and width σ^2 and second, by finding the weight $w_i$between hidden and output neurons. Each Gaussian basis function is considered by distance between data points and the centers. This suggests that the centers should be distributed in the range of input data sets[5]. One simple procedure for selecting the basis function centers c_i is to set them equal to a random subset of the input vectors from the training set. For the width the parameter σ can be simply set as some constant in the same scale of the data points. Radial Basis Function[4] parameters are center, width of the hidden neuron and weight between hidden and output layer. For finding the center of the hidden neuron in Radial Basis function, Fuzzy C-Means[5, 6] that has been produced high accuracy compare with random initialization and its weight updating

Grenze ID: 02.CEMC.2017.3.2

process has been done by standard linear regression in classification problem. Haralambos Sarimveis et al. proposed Fuzzy Radial Basis Functions[7]method the calculation of the hidden node centers is based on the Fuzzy C- Means clustering algorithm, while the connection weights are obtained using standard linear regression. Symmetry-based Fuzzy C-Means Clustering[8,9,10] algorithm used for selecting center and width of Radial Basis Function hidden layer which has been tested alongside the standard Radial Basis Function network and the networks called standard Fuzzy C-means Clustering Radial Basis Function network in forecasting.

The proposed the K-Nearest Neighbour technique[16, 17, 18] (KNN) is used for the initialization of the radius of each RBF and its weight optimize based on Singular Value Decomposition (SVD). Easily lead to decreasing ability of weight learning process in standard Radial Basis Functions, to avoid this deficiency to that has to be solved by using genetic algorithm[19] based Radial Basis Function in that improve the efficient of real time problems to raise more performance. A statistical linear regression[20] technique which is based on the orthogonal least squares (OLS) algorithm for improving the generalization performance (stopping early) of a radial basis function (RBF) neural network. Recently, many research paper applied projection based learning algorithms for updates networks weights that finds the optimal output based on its energy function that is defined by the hinge loss error for classification and times series analysis data. The Meta-cognitive Radial Basis Function Network (McRBFN) and its Projection Based Learning (PBL) algorithm[21, 22, 23] has applied to classification problems for fine-tuning the its network weights. The modified Meta-Cognitive Radial Basis Function Network (McRBFN+) and its Projection Based Learning (PBL) [24, 25, 26]algorithm for classification problems.From the literature, the Projection Based Learning algorithm helps to find the optimum weights and improves the accuracy of the networks. Hence, the proposed method is applied for classification problems. Classification performance of data deals with the type of dataset. A balanced data set is one which has equal number of sample in each classes. As a result of this the classification performance is also improved. An imbalanced dataset is one which has unequal number of samples in each classes. The classification performance deviates as the training of the machine is done by means of imbalanced dataset.

Especially the medical datasets are mostly imbalanced. Through the effective usage of machine learning, diseases at the early stages may be identified. The proposed algorithm performs well in both balanced and imbalanced datasets compared to other machine learning algorithms.

II. METHOD

A. Projections Based Learning-Fuzzy Radial Basis Functional (PBL-FRBF)

Radial Basis Function is a feed forward neural networks model with good performance and global approximation. In RBF have three layers, such as input layer, hidden layer, output layer. In hidden layer, it contains node is called Radial Basis Function[3] units that is Gaussian function node. An RBF[14,15]neural network has two key parameters that describe location function center and width of the RBF unit. The data points may not be evenly distributed to the input space when the center and width of the RBF neuron selected by random. By using clustering techniques for selecting center and width of hidden neurons in RBF, that reflects more accurately in distribution in the data points. Fuzzy C-Means[11, 12, 13] is a fuzzy clustering algorithm which is used for finding the center and width of the Radial Basis Function rather than choosing randomly.

Radial Basis Function network performs a nonlinear mapping from input space R^n to the output space R^m . R^n is an input vector space that is denoted by x_i (for i=1, 2, 3….n) and R^m is output vector space that is denoted by y (for i=1, 2….m). The j^{th} hidden neuron of the Radial Basis Function, which is computes a Gaussian function as below

$$Z_i(x) = \exp(-\frac{\left\| x - c_j \right\|}{2\sigma_i^2}) \quad j = 1, 2, ...m \tag{1}$$

Where x is input feature vector with n dimension. c_j is the center of Gaussian vector of i and σ_i is width of the hidden layer. Center of an every hidden node is calculated by unsupervised method such as Fuzzy C-Means as followed by equation (3).

The width of the hidden layer σ_i is calculated by

$$\sigma_j = \sqrt{\frac{1}{m_j}\sum_{i=1}^{m_j} d^2(c_j - x_i)} \quad (2)$$

When a samples is used to update the output weight parameter by Projection Based Learning algorithm[22, 23] as follows

$$\frac{\partial J(W_K^t)}{\partial w_{pj}} = \frac{\partial J(W_K^t)}{\partial w_{pj}} + \frac{\partial J_t(W_K^t)}{\partial w_{pj}} = 0, p = 1, \ldots..K; \; j = 1, \ldots, n \quad (3)$$

With respect to zero, equating first partial derivative and re-arranging (6) and get as follows,

$$(A^{t-1} + (h^t)^T W_k^t - (B^{t-1} + (h^t)^T (y^t)^T)=0 \quad (4)$$

By substitute $B^{t-1} = A^{t-1} W_K^{t-1} A^{t-1} + (h^t)^T h^t = A^t$ and adding or subtracting the term $(h^t)^T h^t W_K^{t-1}$ on both sides Eq. (4) is reduced to

$W_K^t = (A^t)^{-1}\left(A^t W_K^{t-1} + (h^t)^T (y^t)^T - h^t W_K^{t-1}\right)$ (5) In conclusion, the output weight updated as

$$W_K^t = W_K^{t-1} + (A^t)^{-1} (h^t)^T (e^t)^T \quad (6)$$

This study is utilize the hinge lose error function[26] for estimating error rate between input and output relationship that minimizing an energy function. Projection Based Learning algorithms, linear problems into system of linear equation and that provides a solution for the optimal weights corresponding to the minimum energy function.

III. Performance Evaluation

The proposed PBL-FRBF is evaluated on well known benchmark classification problems and it's all simulations work are conducted with help of MATLAB 2015 and datasets have been collected from UCI machine learning repository[27]. The performance of proposed system is compared with Radial Basis Function (RBF)[7] and Fuzzy Radial Basis Function (FRBF)[12].

A. Description of Datasets

- Fisher's iris data set is consists of three different species of iris flower with four attributes and 150 samples. The distributions of instances are equal such as 50 instances per class which is balanced dataset.
- Wine dataset is consists of 178 samples and 3 different types of class with 13 features. This is an imbalanced dataset, the instances are distributed as follows for
  Class 1 – 59 instances
  Class 2 – 71 instances
  Class 3 – 100 instances
- Wisconsin breast cancer dataset, which consists of 683 samples with 2 class that characterized by 9 features. This dataset is also imbalances where 357 instances belong to the class being and 226 instances belongs to the class maligant
- Glass, which consists of 214 samples and 6 different types of glasses. Each type of class has 9 features. This is an light imbalanced dataset with 163 instances in one class and 51 instances in other class

B. Performance Measures

The overall and average classification efficiencies are used as quantitative evaluation measures in this study. Class level performance and overall performance is obtained by confusion matrix of Q for various algorithms. Class level performance is measured by the percentage of classification η_j is defined as follows

$$\eta_j = \frac{q_{ij}}{N_j} \times 100\ \% \quad (7)$$

Where a total number of its correctly classified samples in class j and is the total number of samples belonging to a class j in the training/testing data set.

Average per-class classification accuracy calculated by

$$\eta_a = \frac{1}{n}\sum_{j=1}^{n} \eta_j \quad (8)$$

An overall classification accuracy calculated by

$$\eta_0 = \frac{\sum_{j=1}^{n} q_{ij}}{N} \times 100\% \quad (9)$$

C. Discussions

The class-wise performance measures of average and overall testing efficiencies and samples used for PBL-FRBF, FRBF and RBF are reported in table 1. In table 2, which is contains average performance of the testing on all 4 datasets. A proposed PBL-FRBF algorithm performs well among present's algorithms on all 4 datasets that obtained from the UCI machine learning repository for classification problems. Advantage of proposed system is needs the least amount of samples to learn the assessment and expand compact architecture to increase generalization performance.

TABLE I: PERFORMANCE ANALYSIS OF PBL-FRBF, FRBF AND RBF FOR CLASSIFICATION

Datasets	**PBL_FRBF**		**FRBF**		**RBF**	
	η_a	η_0	η_a	η_0	η_a	η_0
Iris	**96.75**	**96.03**	93.89	93.81	90.65	91.09
Wine	**98.04**	**98.39**	96.81	95.89	95.89	94.14
Breast Cancer Wisconsin	**89.56**	**91.06**	85.19	86.97	81.94	82.10
Glass	**92.45**	**92.56**	91.56	90.78	90.56	90.04

In terms of iris datasets, proposed network, the performance of classification testing efficiency is improved by 2 % than fuzzy radial basis function and 5 % of efficiency improved than radial basis function. Efficiency of Wine datasets is improved below 2 % than the Fuzzy Radial Basis Functions and 3 % is improved than Radial Basis Function. 4 % of efficiency improved than Fuzzy Radial Basis Function and 8 % of efficiency is improved for Breast Cancer datasets. For applying Glass dataset, the efficiency is improved by 1% than Fuzzy Radial Basis Function and improved by 2 % than Radial Basis Function. Figure 1 is represented performance of the average per class classification accuracy for all used datasets. Its graphical representation is demonstrated that proposed methods is produced superior results than other used algorithms like Fuzzy Radial Basis Function and Radial Basis Function. Figure 2 is represented performance of the overall classification accuracy for all used datasets. It is demonstrated that proposed methods is produced superior results than other used algorithms for all datasets.

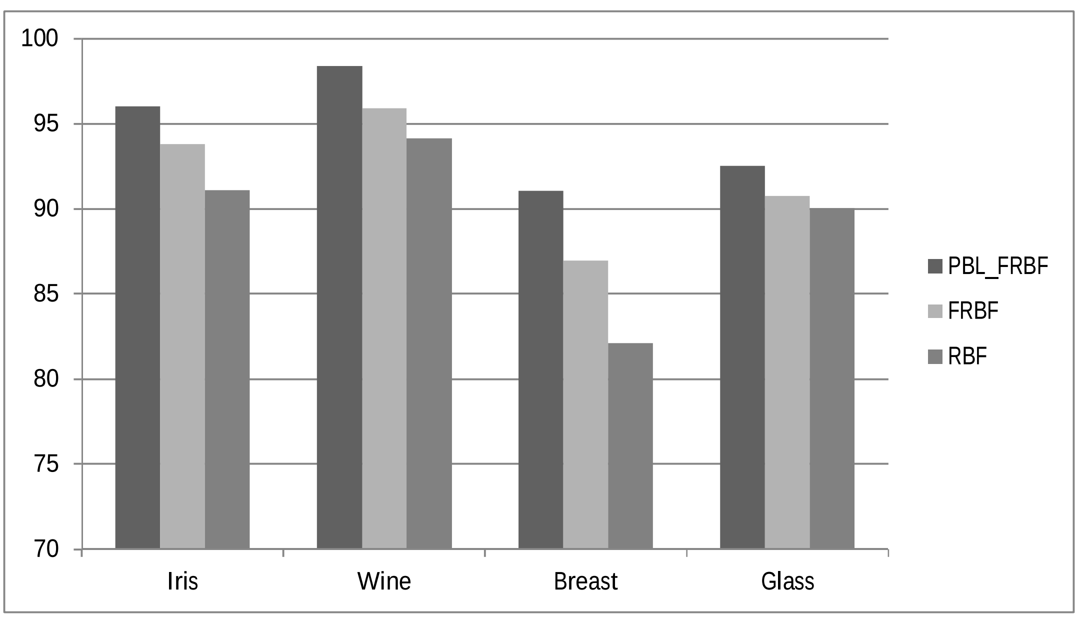

Figure 1 : Performance of an overall classification accuracy

IV. CONCLUSION

This article projects that the feed forward neural network using Fuzzy Radial Basis Function that output weight learned by projection based learning algorithm to reduce the run time with least testing samples and improved generalization performance of neural network architecture. The same has been proved with both imbalanced and balanced datasets. Projection Based Learning accurately estimates the output weight by hinge loss error in order to minimization of the misclassification error in classification problem. The proposed system of PBL-FRBF has been evaluated with standard datasets and demonstrated 4 % to 5 % improvement on imbalanced datasets for class-level as well average classification accuracy.

REFERENCE

[1] Zhang, G.P., 2000. Neural networks for classification: a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 30, 451-462.

[2] Sivanandam, S., Deepa, S., 2006. Introduction to neural networks using Matlab 6.0. Tata McGraw-Hill Education.

[3] Sivanandam, S., Deepa, S., 2007. PRINCIPLES OF SOFT COMPUTING (With CD). John Wiley & Sons.

[4] Tyagi, K., Cai, X., Manry, M.T., 2011. Fuzzy C-means clustering based construction and training for second order RBF network, Fuzzy Systems (FUZZ), 2011 IEEE International Conference on. IEEE, pp. 248-255.

[5] Czarnowski, I., Jędrzejowicz, P., 2016b. Kernel-Based Fuzzy C-Means Clustering Algorithm for RBF Network Initialization, Intelligent Decision Technologies 2016. Springer, pp. 337-347.

[6] Kayhan, G., Ozdemir, A.E., Eminoglu, İ., 2013. Reviewing and designing pre-processing units for RBF networks: initial structure identification and coarse-tuning of free parameters. Neural Computing and Applications. 22, 1655-1666.

[7] Sarimveis, H., Doganis, P., Alexandridis, A., 2006. A classification technique based on radial basis function neural networks. Advances in Engineering Software. 37, 218-221.

[8] Lim, E.A., Zainuddin, Z., 2008. An improved fast training algorithm for RBF networks using symmetry-based fuzzy C-means clustering. Matematika. 24, 141-148.

[9] Byun, H.-G., 2011. An Identification Technique Based on Adaptive Radial Basis Function Network for an Electronic Odor Sensing System. Journal of Sensor Science and Technology. 20, 151-155.

[10] Esmaeili, A., Mozayani, N., 2009. Adjusting the parameters of radial basis function networks using particle swarm optimization, 2009 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications. IEEE, pp. 179-181.

[11] Mitra, S., 2004. Fuzzy radial basis function network: a parallel design. Neural Computing & Applications. 13, 261

[12] Mitra, S., Basak, J., 2001. FRBF: a fuzzy radial basis function network. Neural Computing & Applications. 10, 244-252.

[13] Roh, S.-B., Oh, S.-K., Identification of Plastic Wastes by Using Fuzzy Radial Basis Function Neural Networks Classifier with Conditional Fuzzy C-Means Clustering.

[14] Kaushik, A., Soni, A., Soni, R., 2013. Radial basis function network using intuitionistic fuzzy C means for software cost estimation. International Journal of Computer Applications in Technology. 47, 86-95.

[15] Czarnowski, I., Jędrzejowicz, P., 2015b. An approach to RBF initialization with feature selection, Intelligent Systems' 2014. Springer, pp. 671-682.

[16] Kaminski, W., Strumillo, P., 1997. Kernel orthonormalization in radial basis function neural networks. IEEE Transactions on Neural Networks. 8, 1177-1183.

[17] Awad, M., Pomares, H., Ruiz, I.R., Salameh, O., Hamdon, M., 2009. Prediction of Time Series Using RBF Neural Networks: A New Approach of Clustering. Int. Arab J. Inf. Technol. 6, 138-143.

[18] Czarnowski, I., Jędrzejowicz, P., 2015a. An Approach to RBF Initialization with Feature Selection. in: Angelov, P., Atanassov, K.T., Doukovska, L., Hadjiski, M., Jotsov, V., Kacprzyk, J., Kasabov, N., Sotirov, S., Szmidt, E., Zadrożny, S. (Eds.), Intelligent Systems'2014: Proceedings of the 7th IEEE International Conference Intelligent Systems IS'2014, September 24□26, 2014, Warsaw, Poland, Volume 1: Mathematical Foundations, Theory, Analyses. Springer International Publishing, Cham, pp. 671-682.

[19] Jia, W., Zhao, D., Shen, T., Su, C., Hu, C., Zhao, Y., 2014. A new optimized GA-RBF neural network algorithm. Computational intelligence and neuroscience. 2014, 44.

[20] Lin, C.-L., Wang, J., Chen, C.-Y., Chen, C.-W., Yen, C., 2009. Improving the generalization performance of RBF neural networks using a linear regression technique. Expert Systems with Applications. 36, 12049-12053.

[21] Babu, G.S., Savitha, R., Suresh, S., 2012. A projection based learning in meta-cognitive radial basis function network for classification problems, The 2012 international joint conference on neural networks (IJCNN). IEEE, pp. 1-8.

[22] Babu, G.S., Suresh, S., 2013a. Meta-cognitive RBF network and its projection based learning algorithm for classification problems. Applied Soft Computing. 13, 654-666.

[23] Babu, G.S., Suresh, S., 2013b. Sequential projection-based metacognitive learning in a radial basis function network for classification problems. IEEE transactions on neural networks and learning systems. 24, 194-206.
[24] Subramanian, K., Suresh, S., Cheng, R.P., 2015. A modified projection based learning algorithm for a meta-cognitive radial basis function classifier, Cognitive Computing and Information Processing (CCIP), 2015 International Conference on. IEEE, pp. 1-6.
[25] Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences. 10, 191-203.
[26] Suresh, S., Sundararajan, N., Saratchandran, P., 2008. Risk-sensitive loss functions for sparse multi-category classification problems. Information Sciences. 178, 2621-2638.
[27] Blake, C., Merz, C.J., 1998. {UCI} Repository of machine learning databases.