Journal of Multidisciplinary Modeling and Optimization 3(2) (2020),57-69.

A New Conjugate Gradient Method for Learning Fuzzy Neural Networks

Hisham M. KHUDHUR[1] and Khalil K. ABBO[2]

[1]Mathematics Department, College of Computer Science and Mathematics,
University of Mosul, Mosul, Iraq
hisham892020@uomosul.edu.iq

[2]Department of Mathematics, College of Basic Education,
University of Telafer, Tall'Afar, Iraq
dr.khalil-khuder@uotelafer.edu.iq

Received: 02.01.2021, Accepted: 20.01.2021, Published: 25.03.2021

Abstract — *In this paper, we suggest a conjugate gradient method, which belongs to the optimization methods for learning a fuzzy neural network model that is based on Takagi Sugeno. A new algorithm based on the Polak–Ribière–Polak (PRP) method is introduced to overcome the slow convergence of Polak–Ribière–Polak (PRP) and Liu-Storey (LS) methods. The numerical results indicate the efficiency of the developed method for classifying data as shown in the Table (2) where the new method outperforms above mentioned methods in terms of average training time, average training accuracy, average test accuracy, average training MSE, and average test MSE.*

Keywords: Algorithm, Classification, Fuzzy neural networks, Techniques, Optimization.
Mathematics Subject Classification: 65K10, 90C26, 68T07.

1 Introduction

Fuzzy modeling is to create a large number of local input and output relationships. The purpose of this relation is to define a rule and to make clear a nonlinear manner instead of the classical modeling schemes which may use different equations. [1]. Therefore, by using the given input-output (I-O), a process identification data would become practically a different equivalent problem that concentrates on the description of a fuzzy model[2]. In general, the description of a fuzzy logic method or fuzzy neural (neuro-fuzzy) network method model covers chiefly two phases: construction description and parameter description [3].

Structure identification, In general, the determination of the construction of any fuzzy problem requires, in each law, the number of fuzzy regulations and the membership functions of the premise and consequent fuzzy sets. A variety of techniques is proposed for structure recognition. For the sake of extracting rules from the available input-output dataset to construct the initial rule base, one of these approaches is to use clustering

Cite as: *H. M. Khudhur and K. K. Abbo, A new Conjugate Gradient Method for learning Fuzzy Neural Networks, Journal of Multidisciplinary Modeling and Optimization 3(2) (2020), 57-69.*

algorithms. Typically, multiple clustering devices such as the K-means algorithm can be used to obtain the initial fuzzy rule base of a fuzzy logic system (FLS). [4] , (FCM) fuzzy c-means[5] [6] plus mountain clustering technique[7]. Furthermore, there are other methods for clustering, such as the so-called FPCM , PCM and PFCM in[4] [8] [9]. The fundamental view of the clustering method-based structure identification is to collect the specified samples and position them in various clusters linked by one cluster to a law. The number of laws is also equivalent to the number of clusters. The data must be obtained in advance through of clustering method-based structure recognition. Consequently, online structure recognition is not sufficient. In several experiments, however, scientists stress the use of fumigated neural networks to dynamically model the system.[10] [11]. It is also proposed that the Bayesian TSK fuzzy model in[12] [13], which can classify the number of fuzzy laws without returning to the knowledge of the previous expert. In this paper the researchers concentrate on the clustering of method-based structure recognition as a tool for resolving problems with static regression and classification. The "Gradient based Neuro-Fuzzy learning algorithm" is widely used to characterize the neuro-fuzzy system's feedback, similar to the Neural Network training feedback. [2] [3] [14] [15]. Inspired by the GNF for neuro-fuzzy structures, a GNF update, MGNF, is proposed in[16]. The error function type is revised by considering independent variables in the reciprocal widths of Gaussian membership to prevent singularity. Thus, The weight sequence update formulas are easily modified. This adjustment will help to evaluate the MGNF algorithm converging. In[16] The T-norm product, but the firing strength can be very low for the product, even for a moderate amount of inputs. While any atomic precedent clause can very well be fulfilled. One approach to this issue with other T criteria such as minimum standards [17] [18] [19]. Unfortunately, this is not differential; we want to use gradient-based procedures for T-norm differentiability. As a result, this paper uses a softer variant of the minimum, softmin, to calculate the value of the firing capacity. Softmin's purpose is distinguishable and can manage the Specimen with a wide number of features. [20] [21] [19]. The latter performs much better in terms of both efficiency and acceleration of convergence in general, compared to the common gradient descent technique with conjugate gradient (CG) techniques [22]. The first linear conjugate gradient (CG) technique is implemented in[23], The linear problems can be solved with positive definite coefficient matrices, which can be treated as an optimization algorithm. In addition to the above, the conjugate gradient (CG) method shown in [24] An effective way to solve large-scale nonlinear optimization problems has been found to be an effective tool. Hestenes-Stiefel and for (HS)[23] and Fletcher-Reeves (FR)[24], Another traditional Conjugate gradient (CG) technique Polak–Ribière–Polak (PRP) [25] The alternative direction of the descent is then suggested. Successfully, Conjugate gradient approaches can be extended to the training of neuro-fuzzy networks. [26] [27]. Eight methods of the conjugate gradient (CG) are described in[26] As they are used to equip the fuzzy logic system type-1 to solve the classification issue. The results of learned simulation in [26] Explain that the techniques of conjugate gradient (CG) converge more rapidly than the process of gradient descent (GD). Also, Compared to the ones generated by the optimized fuzzy logic system (FLS) using the gradient descent (GD) process, the classification results derived from conjugate gradient (CG) based fuzzy logic system (FLS) are the best. In [27], Recently, Ahmad et al. [21] developed a new numerical method for solutions of coupled burgers' equations. Also, Ahmad et al. [28] To obtain the numerical solutions of certain nonlinear PDEs such as KdV, mKdV and combined KdV-mKdV equations, a new modification of the variational iteration algorithm-II was proposed.

The goal of this paper is to develop a new Polak–Ribière–Polak (PRP) based algorithm for learning a fuzzy-neural network model to obtain the lowest average training error.

This paper is organized as follows: In Section 2 inference method for Zero-order Takagi-Sugeno is introduced. In Section 3 we present new conjugate gradient (CG) techniques and show that our algorithm satisfies descent and global convergence conditions. Section 4 presents numerical experiments and comparisons.

2 Inference Method for Zero-Order Takagi-Sugeno (TS)

A fuzzy inference scheme that is used as an adaptive network is the neuro-fuzzy model. The neuro-fuzzy model adopted in this article is the zero-order Takagi-Sugeno inference method. Its topological structure can be seen in Fig.1. It is a four-layer network with m-input nodes $x = (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$ and one output node y.

Let us first describe the inference method of the zero-order Takagi-Sugeno.

The basis of the fuzzy rule is defined as follows [29] [30] [31] [14] [32] [33].

$$\text{Rule } i\text{: IF } x_1 \text{ is } A_{1i} \text{ and } x_2 \text{ is } A_{2i} \text{ and } \dots \text{ and } x_m \text{ is } A_{mi} \text{ THEN } y \text{ is } y_i, \tag{1}$$

where $i\ (i = 1,2,\dots,n)$ Matches with the ith fuzzy rule, n is the number of the fuzzy rules, yi is a real number, Ali is a fuzzy subset of xl, and $Ali(xl)$ It means the role of Gaussian membership of the fuzzy judgment ‘‘xl is Ali” defined by

$$A_{li} = \frac{\exp(-(x_l - a_{li})^2}{\sigma_{li}^2} \tag{2}$$

where ali is the center of $Ali(xl)$, and rli is the width of $Ali(xl)$.

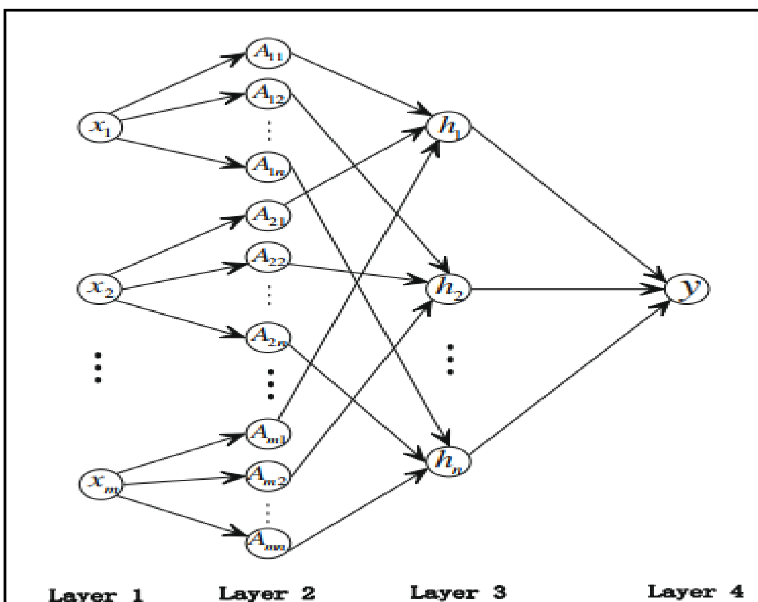

Figure 1: Topological structure of the zero-order takagi–sugeno inference system

For a stated observation $x = (x_1, x_2, \dots, x_m)$ the functions of the nodes in this model are as follows, according to the zero-order Takagi-Sugeno inference method:

Layer1: (input layer): In this layer, each neuron represents one input variable and the input variables are directly passed to the next layer.

Layer2: (membership layer): Each node in this layer represents the membership function of a linguistic variable and serves as a memory unit. Here, the Gaussian functions(2) are adopted as membership functions for the nodes. The weights connecting Layer1 and Layer2 can be interpreted as the Gaussian membership function's centers and widths, respectively.

Layer3: (rule layer): Nodes are referred to as rule nodes in this layer, and each of them denotes a term with a rule. For $i = 1,2,\dots,n$, Agreement on the ith Previous section is estimated by

$$h_i = h_i(x) = A_{1i}(x_1)A_{2i}(x_2)\dots A_{mi}(x_m) = \prod_{l=1}^{m} A_{ii}(x_l) \tag{3}$$

The connecting weights of layers 2 and 3 are set as constant 1.

Layer4: (output layer): This layer performs the summed-weight defuzzification process. The final product of this layer is y, which is a linear combination of the implications of Layer3:

$$y = \sum_{i=1}^{n} h_i y_i \tag{4}$$

The yi relation weights of the output layer are often referred to as conclusion parameters.

Remark 1. In original neuro-fuzzy models [29] [34] [32] [35], the final consequence y is calculated by using the gravity method as follows:

$$y = \frac{\sum_{i=1}^{n} h_i y_i}{\sum_{i=1}^{n} h_i} \tag{5}$$

A popular method is to achieve the fuzzy effect without measuring the center of gravity for ease of learning. Hence, the denominator in (5) is omitted [30] [31] [14] [33]. A further advantage of this operation is the rapid deployment of hardware. [36]. We therefore take the form of (4) in our discussions.

We then take the form of (4) in our debates.

The error function is defined as

$$E(\mathbf{W}) = \frac{1}{2}\sum_{j=1}^{J}(y^j - O^j)^2$$

where O^j is the desired output for the jth training pattern x^j, y^j is the corresponding fuzzy reasoning result, J is the number of training patterns.

The purpose of network learning is to find W^* such that $E(\mathbf{W}^*) = minE(\mathbf{W})$

To solve this optimization problem, the gradient descent approach is sometimes used [37] [38] [39].

3 New Conjugate Gradient (CG) Techniques

Development of new optimization algorithm Based on algorithm Polak–Ribière–Polak (PRP) for learning fuzzy neural networks in the field of data classification and comparison with other optimization algorithms

$$w_{k+1} = w_k + \alpha_k d_k, \;\; k \geq 1, \tag{6}$$

where α_k is step-size obtained by a line search and d_k is the direction of search specified by

$$d_{k+1} = \begin{cases} -g_1, & k = 1 \\ -g_{k+1} + \beta_k d_k, & k \geq 1' \end{cases} \tag{7}$$

where β_k is a parameter. $\beta^{LS} = \frac{-g_{k+1}^T y_k}{g_k^T d_k}$, see [40] and $\beta^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}$, see [25] where $g_k = \nabla E(w_k)$, denotes the gradient of the function of error $E(w)$ in regard to w, k the number of iterations denotes the, and let $y_k = g_{k+1} - g_k$.

Now we suggest a new conjugate gradient algorithm for classifying data depend basically on Polak–Ribière–Polak (PRP) algorithm so we get a new formula:

$$-\theta g_{k+1} + \beta_k d_k = -\gamma g_{k+1} + \beta_k^{PRP} d_k$$

$$-\theta g_{k+1}^T g_{k+1} + \beta_k g_{k+1}^T d_k = -\gamma g_{k+1}^T g_{k+1} + \beta_k^{PRP} g_{k+1}^T d_k$$

$$\beta_k^{NEW} = \begin{cases} \dfrac{(\theta-\gamma) g_{k+1} g_{k+1}}{g_{k+1}^T d_k} + \beta_k^{PRP}, & if \quad g_{k+1}^T d_k \neq 0 \\ \beta_k^{PRP}, & if \quad g_{k+1}^T d_k = 0 \end{cases}$$

where $\theta < \gamma$ and $\theta, \gamma \in [0,1]$.

$$d_{k+1} = -g_{k+1} + \left(\frac{(\theta-\gamma) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \frac{y_k^T g_{k+1}}{g_k^T g_k}\right) d_k \tag{8}$$

3.1 The Descent Property of a Conjugate Gradient (CG) Technique

Below we have to demonstrate the descending property for our proposed new conjugate gradient scheme, denoted by β_k^{NEW}. In the following part

Theorem 1. *The search direction d_{k+1} and β_k^{NEW} given in equation*

$$d_{k+1} = -g_{k+1} + \beta_k^{NEW} d_k$$

$$\beta_k^{NEW} = \begin{cases} \dfrac{(\theta-\gamma) g_{k+1} g_{k+1}}{g_{k+1}^T d_k} + \beta_k^{PRP}, & if \quad g_{k+1}^T d_k \neq 0 \\ \beta_k^{PRP}, & if \quad g_{k+1}^T d_k = 0 \end{cases}$$

$$d_{k+1} = -g_{k+1} + \left(\frac{(\theta-\gamma) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \frac{y_k^T g_{k+1}}{g_k^T g_k}\right) d_k \,,$$

where $\theta < \gamma$ and $\theta, \gamma \in [0,1]$. It will hold for all $k \geq 1$.

Proof. The proof is by using inducement mathematical

1- If $k = 1$ then $g_1^T d_1 < 0$, $d_1 = -g_1 \rightarrow < 0$.

2- Let the relation $g_k^T d_k < 0$ for all k.

3- We are going to prove that the relationship is true when $k = k + 1$ by multiplying the equation (8) in g_{k+1} we obtain

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \left(\frac{(\theta-\gamma) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \frac{y_k^T g_{k+1}}{g_k^T g_k}\right) g_{k+1}^T d_k$$

Let $\tau = \frac{(\theta-\gamma) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k}, v = \frac{y_k^T g_{k+1}}{g_k^T g_k}$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + (\tau + v) g_{k+1}^T d_k$$

Let $g_{k+1}^T d_{k+1} > 0$ and $\tau > v$ the $g_{k+1}^T d_{k+1} \geq 0$.

3.2. Global Convergence

We will display that conjugate gradient (CG) method with β_k^{NEW} convergences globally. For the convergence of the proposed new algorithm, we need a certain assumption.

Assumption 1. [41][42]

1- Assume E in the level set is bound below $S = \{w \in R^n : E(w) \le E(w_0)\}$; In some Initial point.

2- E is continuously differentiable and its gradient is Lipchitz continuous, there exist $L > 0$ such that[43]:

$$\| g(x) - g(y) \| \le Lx - y \| \forall x, y \in N \tag{9}$$

On the other hand, under Assumption(1), it is clear that there exist positive constants B such

$$\|w\| \le B, \forall w \in S \tag{10}$$

$$\| \nabla E(w) \| \le \overline{\gamma}, \forall x \in S \tag{11}$$

Lemma 1. *Assume that Assumption (1) and equation (10) hold. take into consideration any conjugate gradient method in from (6) and (7), where d_k is a decent direction and α_k is obtained by the S.W.L.S. If*

$$\sum_{k>1} \frac{1}{\| d_{k+1} \|^2} = \infty$$

then we have

$$\lim_{k\to\infty} inf \, \| g_k \| = 0$$

more details can be found in [44][45][46].

Theorem 2. *Assume that Assumption (1) and equation (6) and the descent condition hold. Consider a conjugate gradient scheme in the form*

$$d_{k+1} = -g_{k+1} + \beta_k^{NEW} d_k,$$

where α_k is computed from strong Wolfe line search condition for more details see [47] [48] [49] [50] , If the objective function is uniformly on set S, then

$$lim_{n\to\infty}(inf \, \| g_k \|) = 0 \, .$$

Proof.

$$d_{k+1} = -g_{k+1} + \beta_k^{NEW1} d_k$$

$$\beta_k^{NEW} = \begin{cases} \dfrac{(\theta - \gamma) g_{k+1} g_{k+1}}{g_{k+1}^T d_k} + \beta_k^{PRP}, & if \quad g_{k+1}^T d_k \ne 0 \\ \beta_k^{PRP}, & if \quad g_{k+1}^T d_k = 0 \end{cases}$$

$$\| d_{k+1} \| = \| -g_{k+1} + \left(\frac{(\theta - \gamma) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \frac{y_k^T g_{k+1}}{g_k^T g_k}\right) d_k \|$$

$$\| d_{k+1} \| \le \| g_{k+1} \| + \| \frac{(\theta - \gamma) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \frac{y_k^T g_{k+1}}{g_k^T g_k}) \| d_k \|$$

$$\| d_{k+1} \| \le \| g_{k+1} \| + \| \frac{(\theta - \gamma) \| g_{k+1} \|^2}{\| d_k \| \| g_{k+1} \|} + \frac{\| y_k^T \| \| g_{k+1} \|}{\| g_k \|^2} \| \| d_k \|$$

$$\| d_{k+1} \| \leq (1 + \| \frac{(\theta - \gamma) \| d_k \|}{\| d_k \|} + \frac{\| d_k \| \| y_k^T \|}{\| g_k \|^2}) \| g_{k+1} \|$$

$$\psi = \| \frac{(\theta - \gamma) \| d_k \|}{\| d_k \|} + \frac{\| d_k \| \| y_k^T \|}{\| g_k \|^2} \|$$

$$\| d_{k+1} \| \leq (1 + \psi) \| g_{k+1} \|$$

$$\sum_{k \geq 1} \frac{1}{\| d_{k+1} \|^2} \geq (\frac{1}{(1+\psi)^2}) \frac{1}{\gamma^2} \sum 1 = \infty.$$

3 Numerical Examples

The conjugate gradient algorithm developed to teach the fuzzy neural networks described in Part Two is evaluated by comparing it with related algorithms such as LS and PRP to classify the data given by the following classification problems (Iris, Thyroid, Glass, Wine, Breast Cancer and Sonar) [51], The developed algorithm NEW showed high efficiency in data classification compared to LS and PRP algorithms as shown in the following table and graphs, The simulation was carried out using Matlab 2018b, running on a Windows 8 HP machine with an Intel Core i5 processor, 4 GB of RAM and 500 GB of hard disk drive.

Table 1: Problems in Real-World Classification [51]

	Classification dataset	*Data size*	*No. of training samples*	*No. of testing samples*
1	*Iris*	150	90	60
2	*Thyroid*	215	129	86
3	*Glass*	214	107	107
4	*Wine*	178	89	89
5	*Breast Cancer*	253	127	126
6	*Sonar*	208	104	104

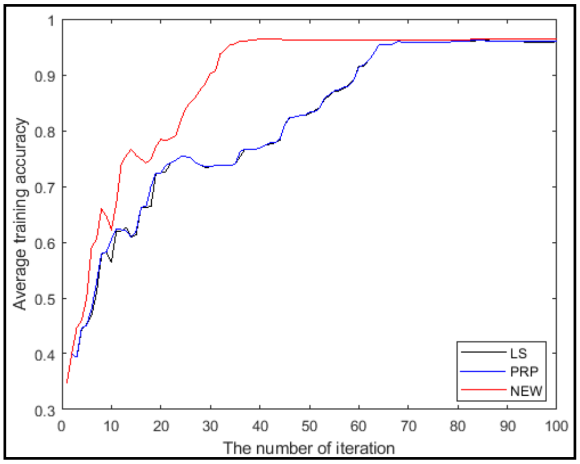

Figure 2: The average training accuracy for Iris

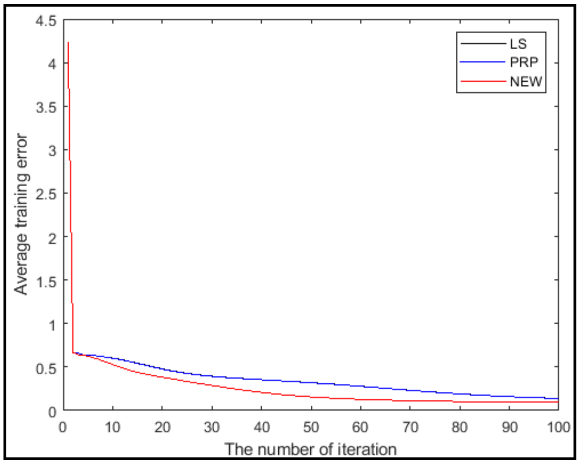

Figure 3: The average training error results for Iris

Table 2: Average Performance Comparison for Classification Problems for NEW

Datasets	*Algorithms*	*No. of training iteration*	*Average training time*	*Average training accuracy*	*Average test accuracy*	*Average training MSE*	*Average test MSE*
Iris	*LS*	*100*	***0.2883***	*0.9600*	*0.9367*	*0.1392*	*0.1541*
	PRP	*100*	*0.4328*	*0.9600*	*0.9367*	*0.1391*	*0.1541*
	NEW	*100*	*0.6028*	***0.9644***	***0.9500***	***0.0932***	***0.1110***
Thyroid	*LS*	*100*	*0.1702*	*0.6930*	*0.7070*	*0.3899*	*0.3874*
	PRP	*100*	*0.1703*	*0.8961*	*0.9000*	*0.1627*	*0.1632*
	NEW	*100*	***0.1686***	***0.9271***	***0.9163***	***0.1356***	***0.1370***
Glass	*LS*	*100*	*0.6123*	*0.3121*	*0.2617*	*0.7841*	*0.7944*
	PRP	*100*	***0.6059***	*0.3121*	*0.2673*	*0.7807*	*0.7990*
	NEW	*100*	*0.6565*	***0.5664***	***0.4636***	***0.5828***	***0.6414***
Wine	*LS*	*100*	*0.4195*	*0.4854*	*0.4427*	*0.6324*	*0.6568*
	PRP	*100*	***0.4147***	*0.9528*	*0.9258*	*0.1340*	*0.1827*
	NEW	*100*	*0.4230*	***0.9730***	***0.9348***	***0.1131***	***0.1597***
Breast Cancer	*LS*	*100*	*1.7727*	*0.4677*	*0.4619*	*0.9235*	*0.9347*
	PRP	*100*	***1.7482***	*0.4646*	*0.4619*	*0.9142*	*0.9257*
	NEW	*100*	*1.8412*	***0.6630***	***0.6349***	***0.6235***	***0.6727***
Sonar	*LS*	*100*	*2.1704*	*0.5442*	*0.5288*	*0.6071*	*0.6078*
	PRP	*100*	***2.1506***	*0.5115*	*0.5000*	*0.6063*	*0.6071*
	NEW	*100*	*2.1819*	***0.6558***	***0.5942***	***0.4296***	***0.4873***

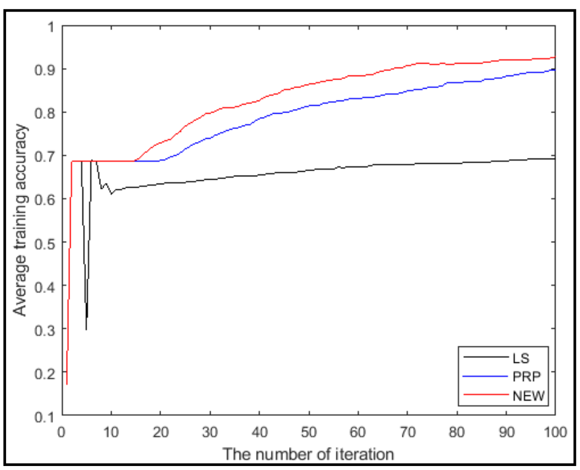

Figure 4: The average training accuracy for Thyroid

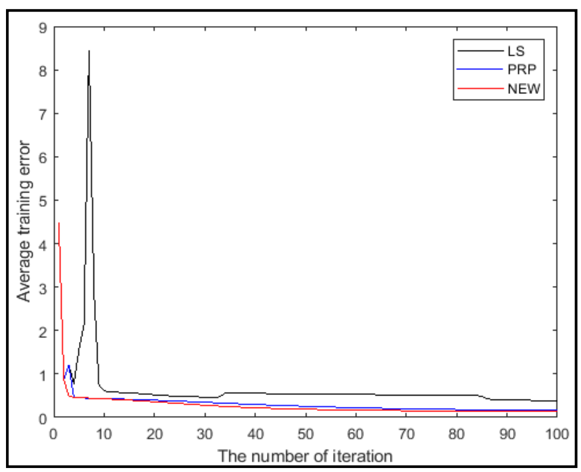

Figure 5: The average training error results for Thyroid

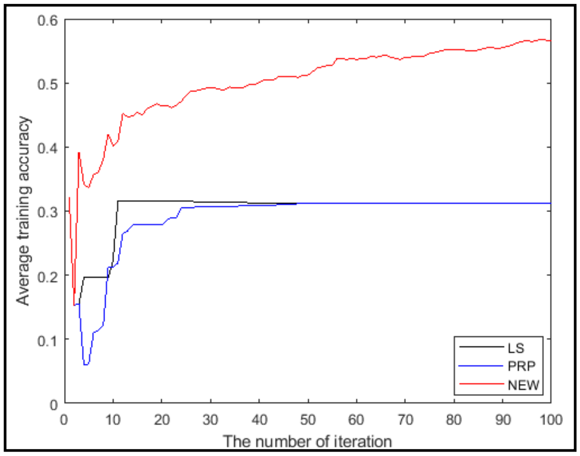

Figure 6: The average training accuracy for Glass

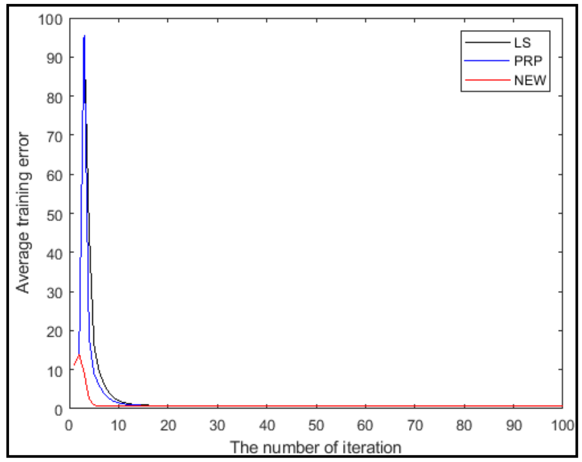

Figure 7: The average training error results for Glass

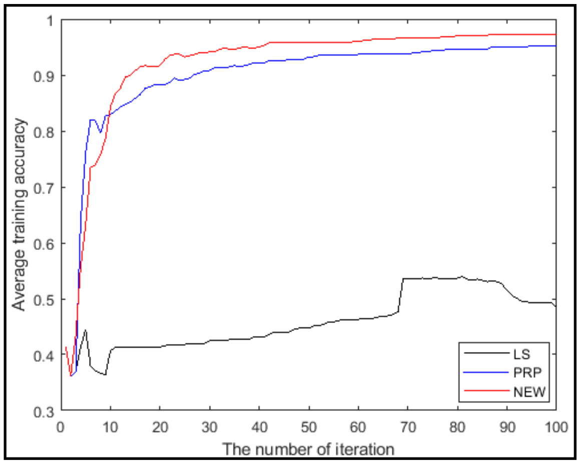

Figure 8: The average training accuracy for Wine

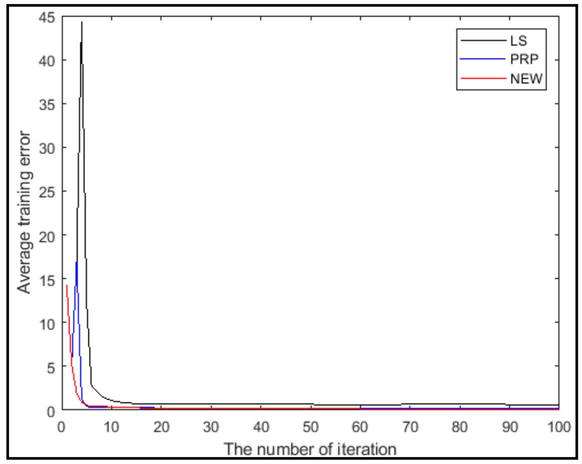

Figure 9: The average training error results for Wine

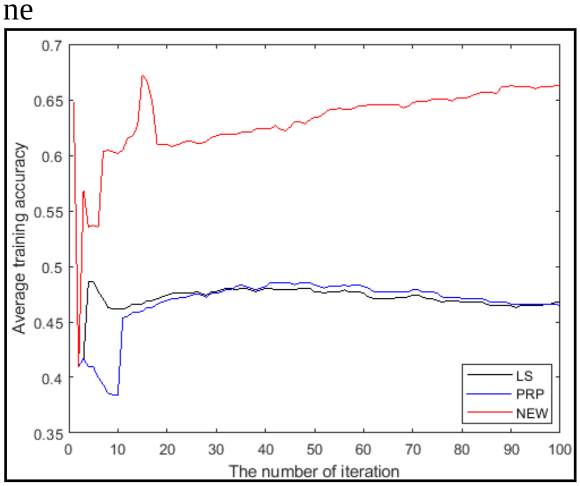

Figure 10: The average training accuracy for Breast Cancer

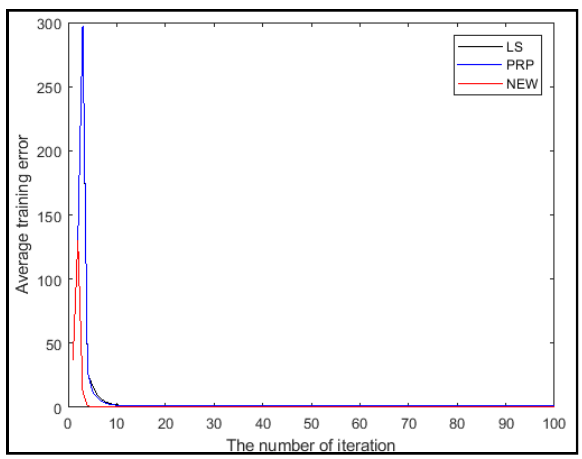

Figure 11: The average training error results for Breast Cancer

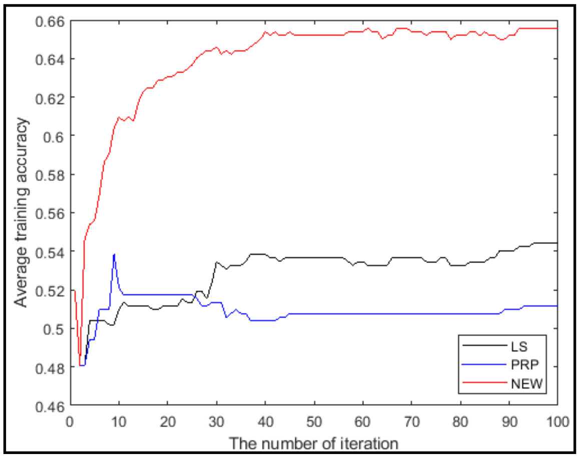

Figure 12: The average training accuracy for Sonar

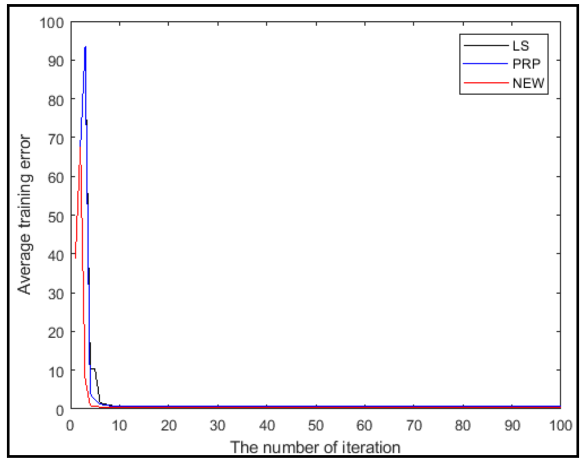

Figure 13: The average training error results for Sonar

4 Conclusion

Our Conjugate gradient technique is a good option to a gradient descent method for its faster convergence speed via looking for a conjugate descent path with adaptive learning coefficients. An updated conjugate gradient approach has been proposed in this paper to train the fuzzy neural network system of the 0-th order Takagi-Sugeno (TS). Numerical simulations shown that new algorithm has a better generalization efficiency than its current counterparts. Also, the simulations observed endorse the converging behavior of the suggested algorithm is very well. We also conclude that the proposed technique can solve the optimization functions and can be used in training artificial neural networks.

Acknowledgments

The authors are grateful to Dr. Tao Gao, Dr. Jian Wang at the College of information, Control Engineering and College of Science at the China University of Petroleum, China, for their precious support.

Conflict of Interest Declaration

The authors declare that there is no conflict of interest statement.

Ethics Committee Approval and Informed Consent

The authors declare that there is no ethics committee approval and/or informed consent statement.

References

[1] S. Chakraborty, A. Konar, A. Ralescu, and N. R. Pal, A fast algorithm to compute precise type-2 centroids for real-time control applications, *IEEE Trans. Cybern.*, 45(2) 2014, 340-353.

[2] M. Sugeno and G. T. Kang, Structure identification of fuzzy model, *Fuzzy Sets Syst.*, 28 (1), 1988, 15–33.

[3] T. Takagi and M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Trans. Syst. Man. Cybern.*, 1 1985, 116–132.

[4] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal, Fuzzy models and algorithms for pattern recognition and image processing, Springer Science & Business Media, vol. 4. 1999.

[5] J. C. Bezdek, Objective function clustering, in Pattern recognition with fuzzy objective function algorithms, Springer, 1981, 43–93.

[6] X. Gu and S. Wang, Bayesian Takagi–Sugeno–Kang fuzzy model and its joint learning of structure identification and parameter estimation, *IEEE Trans. Ind. Informatics*, 14(12) 2018, 5327–5337.

[7] R. R. Yager and D. P. Filev, Generation of fuzzy rules by mountain clustering, *J. Intell. Fuzzy Syst.*, 2(3) 1994, 209–219.

[8] R. Krishnapuram and J. M. Keller, A possibilistic approach to clustering, *IEEE Trans. fuzzy Syst.*, 1(2) 1993, 98–110.

[9] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, A possibilistic fuzzy c-means clustering algorithm, *IEEE Trans. Fuzzy Syst.*, 13(4) 2005, 517–530.

[10] C.-F. Juang and C.-T. Lin, An online self-constructing neural fuzzy inference network and its applications, *IEEE Trans. Fuzzy Syst.*, 6(1) 1998, 12–32.

[11] H. Shahparast, E. G. Mansoori, and M. Z. Jahromi, AFCGD: an adaptive fuzzy classifier based on gradient descent, *Soft Comput.*, 23(12) 2019, 4557–4571.

[12] X. Gu, F.-L. Chung, H. Ishibuchi, and S. Wang, Imbalanced TSK fuzzy classifier by cross-class Bayesian fuzzy clustering and imbalance learning, *IEEE Trans. Syst. Man, Cybern. Syst.*, 47(8) 2016, 2005–2020.

[13] X. Gu, F.-L. Chung, and S. Wang, Bayesian Takagi–Sugeno–Kang fuzzy classifier, *IEEE Trans. Fuzzy Syst.*, 25(6) 2016, 1655–1671.

[14] H. Ichihashi and I. B. Türksen, A neuro-fuzzy approach to data analysis of pairwise comparisons, *Int. J. Approx. Reason.*, 9(3) 1993, 227–248.

[15] J. M. Mendel, General type-2 fuzzy logic systems made simple: a tutorial, *IEEE Trans. Fuzzy Syst.*, 22(5) 2013, 1162–1182.

[16] W. Wu, L. Li, J. Yang, and Y. Liu, A modified gradient-based neuro-fuzzy learning algorithm and its convergence, *Inf. Sci. (Ny).*, 2010, doi: 10.1016/j.ins.2009.12.030.

[17] H. Ahmad, T. A. Khan, P. S. Stanimirović, Y. M. Chu, and I. Ahmad, Modified variational iteration algorithm-II: Convergence and applications to diffusion models, *Complexity*, 2020, doi: 10.1155/2020/8841718.

[18] A. Ghosh, N. R. Pal, and J. Das, A fuzzy rule based approach to cloud cover estimation, *Remote Sens. Environ.*, 100(4) 2006, 531–549.

[19] N. R. Pal and S. Saha, Simultaneous structure identification and fuzzy rule generation for Takagi–Sugeno models, *IEEE Trans. Syst. Man, Cybern. Part B*, 38(6) 2008, 1626–1638.

[20] H. Ahmad, A. Akgül, T. A. Khan, P. S. Stanimirović, and Y. M. Chu, New perspective on the conventional solutions of the nonlinear time-fractional partial differential equations, *Complexity*, 2020, doi: 10.1155/2020/8829017.

[21] H. Ahmad, T. A. Khan, and C. Cesarano, Numerical solutions of coupled burgers' equations, *Axioms*, 2019, doi: 10.3390/axioms8040119.

[22] J. Wang, W. Wu, and J. M. Zurada, Deterministic convergence of conjugate gradient method for feedforward neural networks, *Neurocomputing*, 74(14–15) 2011, 2368–2376.

[23] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, 49(1). NBS Washington, DC, 1952.

[24] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, Comput. J., 7(2) 1964, 149–154, doi: 10.1093/comjnl/7.2.149.

[25] E. Polak and G. Ribiere, Note sur la convergence de méthodes de directions conjuguées, ESAIM Math. Model. Numer. Anal. Mathématique Anal. Numérique, 3(1) 1969, 35–43.

[26] E. P. de Aguiar *et al.*, EANN 2014: a fuzzy logic system trained by conjugate gradient methods for fault classification in a switch machine, Neural Comput. Appl., 27(5) 2016, 1175–1189.

[27] T. Gao, J. Wang, B. Zhang, H. Zhang, P. Ren, and N. R. Pal, A polak-ribière-polyak conjugate gradient-based neuro-fuzzy network and its convergence, IEEE Access, 2018, doi: 10.1109/ACCESS.2018.2848117.

[28] H. Ahmad, A. R. Seadawy, and T. A. Khan, Study on numerical solution of dispersive water wave phenomena by using a reliable modification of variational iteration algorithm, Math. Comput. Simul., 2020, doi: 10.1016/j.matcom.2020.04.005.

[29] I. Del Campo, J. Echanobe, G. Bosque, and J. M. Tarela, Efficient hardware/software implementation of an adaptive neuro-fuzzy system, *IEEE Trans. Fuzzy Syst.*, 16(3) 2008, 761–778.

[30] K. T. Chaturvedi, M. Pandit, and L. Srivastava, Modified neo-fuzzy neuron-based approach for economic and environmental optimal power dispatch, *Appl. Soft Comput.*, 8(4) 2008, 1428–1438.

[31] H. Ichihashi, Iterative fuzzy modelling and a hierarchical network, 1991.

[32] C.-J. Lin and W.-H. Ho, An asymmetry-similarity-measure-based neural fuzzy inference system, *Fuzzy Sets Syst.*, 152(3) 2005, 535–551.

[33] M. Tang, K. jun Wang, and Y. Zhang, A research on chaotic recurrent fuzzy neural network and its convergence, in *2007 International Conference on Mechatronics and Automation*, 2007, 682–687.

[34] J.-S. Jang, ANFIS: adaptive-network-based fuzzy inference system, *IEEE Trans. Syst. Man. Cybern.*, 23(3) 1993, 665–685.

[35] X. G. Luo, D. Liu, and B. W. Wan, An adaptive fuzzy neural inferring network, *Fuzzy Syst. Math.*, 12(4) 1998, 26–33.

[36] C.-F. Juang and J.-S. Chen, Water bath temperature control by a recurrent fuzzy controller and its FPGA implementation, *IEEE Trans. Ind. Electron.*, 53(3) 2006, 941–949.

[37] A. Sahiner , N. Yilmaz and S. A. Ibrahem, Smoothing approximations to non-smooth functions, *J. Multidiscip. Model. Optim.*, 1(2) 2018, 69-74.

[38] A. S. Ahmed, Optimization Methods For Learning Artificial Neural Networks, University of Mosul, 2018.

[39] A. Sahiner and S. A. Ibrahem, A new global optimization technique by auxiliary function method in a directional search, *Optim. Lett.*, 2019, doi: 10.1007/s11590-018-1315-1.

[40] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms, part 1: theory, *J. Optim. Theory Appl.*, 69(1) 1991, 129–137.

[41] K. K. Abbo and H. M. Khudhur, New A hybrid Hestenes-Stiefel and Dai-Yuan conjugate gradient algorithms for unconstrained optimization, *Tikrit J. Pure Sci.*, 21(1) 2015, 118–123.

[42] Y. A. Laylani, K. K. Abbo, and H. M. Khudhur, Training feed forward neural network with modified Fletcher-Reeves method, *J. Multidiscip. Model. Optim.*, 1(1) 2018, 14–22.

[43] K. K. Abbo, Y. A. Laylani, and H. M. Khudhur, Proposed new Scaled conjugate gradient algorithm for unconstrained optimization, *Int. J. Enhanc. Res. Sci. Technol. Eng.*, 5(7) 2016.

[44] Z. M. Abdullah, M. Hameed, M. K. Hisham, and M. A. Khaleel, Modified new conjugate gradient method for Unconstrained Optimization, *Tikrit J. Pure Sci.*, 24(5) 2019, 86–90.

[45] H. M. Khudhur, Numerical and analytical study of some descent algorithms to solve unconstrained Optimization problems, University of Mosul, 2015.

[46] K. K. Abbo, Y. A. Laylani, and H. M. Khudhur, A new spectral conjugate gradient algorithm for unconstrained optimization, *Int. J. Math. Comput. Appl. Res.*, 8 2018, 1–9.

[47] M. Al-Baali, Descent property and global convergence of the Fletcher—Reeves method with inexact line search, *IMA J. Numer. Anal.*, 5(1) 1985, 121–124.

[48] L. Zhang and W. Zhou, Two descent hybrid conjugate gradient methods for optimization, *J. Comput. Appl. Math.*, 216(1) 2008, 251–264.

[49] K. K. Abbo and H. M. Khudhur, New A hybrid conjugate gradient Fletcher-Reeves and Polak-Ribiere algorithm for unconstrained optimization, *Tikrit J. Pure Sci.*, 21(1) 2015,124–129.

[50] H. N. Jabbar, K. K. Abbo, and H. M. Khudhur, Four--Term Conjugate Gradient (CG) Method Based on Pure Conjugacy Condition for Unconstrained Optimization, *Kirkuk Univ. J. Sci. Stud.*, 13(2) 2018, 101–113.

[51] T. Gao, Z. Zhang, Q. Chang, X. Xie, P. Ren, and J. Wang, Conjugate gradient-based Takagi-Sugeno fuzzy neural network parameter identification and its convergence analysis, *Neurocomputing*, 2019, doi: 10.1016/j.neucom.2019.07.035.

Hisham M. Khudhur, ORCID: https://orcid.org/0000-0001-7572-9283
Khalil K. Abbo, ORCID: https://orcid.org/0000-0001-5858-625X