Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Learning Fuzzy Neural Networks by Using Improved Conjugate Gradient Techniques

Hisham M. Khudhur[1], Khalil K. Abbo[2]
[1]Deparment of Mathematics , College of Computers Sciences and Mathematics , University of Mosul , Iraq,
hisham892020@uomosul.edu.iq.
[2]Department of Studies and planning, Presidency of telafer university, University of Telafer, Tall'Afar, Iraq.

Submitted : May 14, 2022 | **Accepted** : May 23, 2022 | **Published** : July 1, 2022

Abstract: One of the optimal approaches for learning a Takagi Sugeno-based fuzzy neural network model is the conjugate gradient method proposed in this research. For the PRP and the LS approaches, a novel algorithm based on the Liu-Storey (LS) approach is created to overcome the slow convergence. The developed method becomes descent and convergence by assuming some hypothesis. The numerical results show that the developed method for classifying data is more efficient than the other methods, as shown in Table (2), where the new method outperforms the others in terms of average training time, average training accuracy, average test accuracy, average training mean square error (MSE), and average test mean square error (MSE).

Keywords: classification; conjugate gradient; Liu-Storey; fuzzy neural networks; numerical; optimization.

INTRODUCTION

To understand a process, fuzzy modeling attempts to discover a set of local interactions between inputs and outputs. It is more effective at expressing nonlinear processes than standard modeling approaches that employ differential equations, which are more accurate at expressing linear processes [1]. The identification process based on input-output data (I-O) data is transformed into a nearly identical issue concentrating on the identification of a fuzzy model as a consequence of this transformation [2]. For the most part, there are two steps to the process of developing a fuzzy logic system (FLS) or fuzzy neural network system (neuro-fuzzy network system) model: Determination of the structural and parametric characteristics [3].
Identifying the structure of a fuzzy model, the number of fuzzy rules, and the membership functions of the premise and following fuzzy sets in each rule is the primary focus of structure identification. There are a number of methods for determining the structure of an object. In order to begin building the first rule base, clustering methods may be used to extract rules from an input-output dataset. The K-means algorithm is one example of several clustering techniques [4], fuzzy c-means (FCM) [5], and the mountain clustering method [6], are often used to obtain a FLS's first fuzzy rule foundation. Other clustering techniques, such as PCM, FPCM, and PFCM, are also mentioned in this study [7]. The main concept underlying the clustering method-based structure identification is that the provided samples are separated into multiple clusters, each of which corresponds to a rule in the structure identification process. In other words, the number of rules is the same as the number of clusters in the system. This approach of structure identification requires the collecting of data in advance, which makes it inappropriate for use in real-time structure identification applications. In a number of studies, researchers have concentrated on the modeling of dynamic systems using fuzzy neural networks [8] Furthermore, the Bayesian TSK fuzzy model suggested in [9][10] without the requirement for previous expert knowledge, it is possible to specify the number of fuzzy rules. In order to prevent singularity, the error function's form has been modified to include the reciprocals of Gaussian membership function widths as independent variables, rather than only the widths of the functions themselves. Because of this, the weight sequence formulae have been changed in an easy way. The convergence analysis of the MGNF algorithm is made simpler as a result of this conversion. In [11], The product T-norm is used in this instance. We have shown that even with a reasonable number of inputs, even with a high number of atomic antecedent clauses met, the firing strength of the product may be quite low when it is used. Using other T-norms, such as minimum and maximum, may assist in resolving this issue [12], We require the T-norm to be differentiable, however, if we are to employ gradient-based approaches, and this is not the case since it is not differentiable. The value of the firing strength is calculated in

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442
e-ISSN : 2541-2019
p-ISSN : 2541-044X

this research using a softer form of the minimum function, known as softmin, as a consequence of this. Softmin is a differentiable function that can deal with a sample that has a high number of different characteristics [13]. Conjugate gradient (CG) approaches often surpass the common gradient descent method in terms of efficiency and convergence speed when compared to the common gradient descent method [14]. The linear conjugate gradient (CG) method, which was first introduced in [15], is an optimization method for solving linear problems with positive definite coefficient matrices. Furthermore, the conjugate gradient (CG) method was shown to be a strong technique for solving large-scale nonlinear optimization problems in [16]. Aside from Hestenes-Stiefel (HS) [15] and Fletcher-Reeves (FR) [16], Polak-Ribiere-Polyak (PRP) [17] has been proposed as a typical conjugate gradient (CG) technique based on a different descent direction option. The use of conjugate gradient methods to train neuro-fuzzy networks has been successful [18] [19]. In[18], The type-1 fuzzy logic system is equipped with eight conjugate gradient techniques, which are utilized to solve the classification issue in the form of a classification problem solving algorithm. in [25] A Polak–Ribière–Polak (PRP) technique has been proposed as a typical conjugate gradient (CG) technique. The conjugate gradient method Polak–Ribière–Polak (PRP) has been successfully used to train neuro-fuzzy networks. The aim of this paper is to create a new Liu-Storey (LS) based algorithm for learning a fuzzy-neural network model with the smallest average training error possible.

TAKAGI–SUGENO INFERENCE SYSTEM WITH ZERO-ORDER

A neuro-fuzzy model, which can be considered as an adaptive network, represents a fuzzy inference system. The zero-order Takagi–Sugeno inference system is the neuro-fuzzy model we'll be looking at in this dissertation. Figure(1) depicts the topological structure of the network, which is a four-layer network with m input nodes $x = (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$ and a single output node y.
To begin, we'll go over the zero-order Takagi–Sugeno inference system. We're concerned with a hazy rule set, which is defined as follows: [20], [21]:

Rule i: IF x_1 is A_{1i} and x_2 is A_{2i} and … and x_m is A_{mi} THEN y is y_i, (1)

where i $(i = 1,2,\dots,n)$ n is the number of fuzzy rules, yi is a real number, Ali is a fuzzy subset of xl, and $Ali(xl)$ is a Gaussian membership function of the fuzzy judgment ''xl is Ali'' defined by

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

Where ali denotes $Ali(xl)$ is and rli denotes Ali's width (xl).

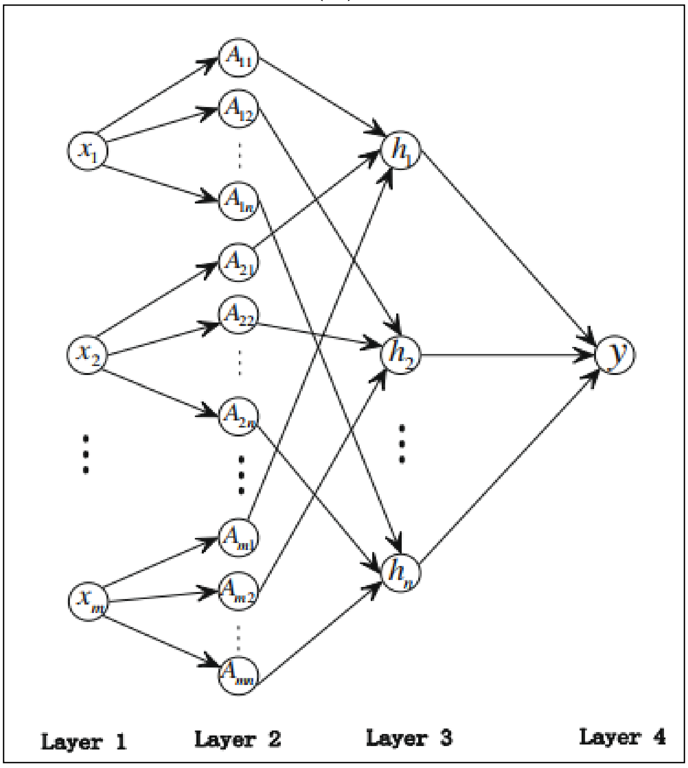

Fig. 1 The Topological Structure of the Takagi–Sugeno Inference System with Zero Order

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442
e-ISSN : 2541-2019
p-ISSN : 2541-044X

For a given observation $x = (x1, x2, ..., xm)$, The functions of the nodes in this model, according to the zero-order Takagi–Sugeno inference system, are as follows:
Layer1 (input layer): One input variable is represented by each neuron in this layer. And the input variables are passed through to the next layer without modification.
Layer2 (membership layer): Each node in this layer represents a linguistic variable's membership function and serves as a memory unit. As membership functions for the nodes, the Gaussian functions (2) are used. The weights connecting Layers 1 and 2 can be understood as the Gaussian membership function's centers and widths, respectively.
Layer3 (rule layer): The nodes in this layer are known as rule nodes, and each one represents a rule word. The agreement of the ith antecedent portion is computed for $i = 1,2, ..., n$ by

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

The link weights between layers 2 and 3 are set to a constant of one.
Layer4 (output layer): The summed-weight defuzzification operation is carried out by this layer. This layer's output is the final consequence y, which is a linear combination of the Layer3 consequences:

$$y = \sum_{i=1}^{n} h_i y_i \quad (4)$$

The output layer's link weights yi are also known as conclusion parameters.
Remark 1. The final result y is computed using the gravity method in original neuro-fuzzy models[20] as follows:

$$y = \frac{\sum_{i=1}^{n} h_i y_i}{\sum_{i=1}^{n} h_i} \quad (5)$$

A common approach for learning is to achieve the fuzzy result without computing the center of gravity, so the denominator in (5) is omitted [22]–[24]. Another benefit of this operation is its straightforward hardware implementation [25], [26] . As a result, we'll use the form (4) throughout our conversation.

NEW CONJUGATE GRADIENT (NEW2) METHOD

Development of new optimization algorithm Based on algorithm Liu-Storey (LS) for learning fuzzy neural networks in the field of data classification and comparison with other optimization algorithms

$$w_{k+1} = w_k + \alpha_k d_k, \quad k \geq 1 \quad (6)$$

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

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

Where β_k is a parameter.

$$\beta^{LS} = \frac{-g_{k+1}^T y_k}{g_k^T d_k}, \text{ see [27]}$$

$$\beta^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \text{ see [17]}$$

Where $g_k = \nabla E(w_k)$, signifies the gradient of the error function $E(w)$ with respect to w, k denotes the number of iterations, and let $y_k = g_{k+1} - g_k$ signify the number of iterations. The novel conjugate gradient approach for classifying data is based on the Liu-Storey (LS) algorithm, and as a result we get a new formula for classification data.

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

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

$$\beta_k^{NEW2} = \begin{cases} \dfrac{(\theta_k - \gamma_k) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \beta_k^{LS}, & if \quad g_{k+1}^T d_k \neq 0 \\ \beta_k^{LS}, & if \quad g_{k+1}^T d_k = 0 \end{cases}$$

Where $\theta > \gamma$ and $\theta, \gamma \in [0,1]$.

$$d_{k+1} = -g_{k+1} + \beta_k^{NEW2} d_k \quad (8)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442
e-ISSN : 2541-2019
p-ISSN : 2541-044X

A NEW CG METHOD'S DESCENT PROPERTY

The descent property of our proposed new conjugate gradient scheme, denoted by the abbreviation β_k^{NEW2}, must be shown in the next section. The following will be discussed:

Theorem (1)

The search direction d_{k+1} and β_k^{NEW2} are given in equation:

$$d_{k+1} = -g_{k+1} + \beta_k^{NEW2} d_k$$

$$\beta_k^{NEW2} = \begin{cases} \dfrac{(\theta_k - \gamma_k) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \beta_k^{LS}, & if \quad g_{k+1}^T d_k \neq 0 \\ \beta_k^{LS}, & if \ \ g_{k+1}^T d_k = 0 \end{cases}$$

Where $\theta > \gamma$ and $\theta, \gamma \in [0,1]$.

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

Where $\theta > \gamma$ and $\theta, \gamma \in [0,1]$.

Will be valid for all $k \geq 1$

Proof:-

The proof is based on the use of mathematical inducement

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

2- Let the relation $g_k^T d_k < 0 \ \forall \ k$.

3- When $k = k + 1$, we show that the relationship is true multiplication of the formula (8) by g_{k+1} we obtain

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

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

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

Let $g_{k+1}^T d_{k+1} > 0$ and $\sigma + \zeta$

Then

$$g_{k+1}^T d_{k+1} \leq 0.$$

GLOBAL CONVERGENCE STUDY

We will demonstrate how the conjugate gradient (CG) approach with β_k^{NEW}convergences works on a larger scale. We need a certain assumption in order for the suggested new method to achieve convergence.

Assumption (1)[28],[29],[30],[31]

1- Assume E in the level set is bound below $S = \{w \in R^n : E(w) \leq E(w_0)\}$; In some Initial point.
2- Because E is continuously differentiable and its gradient is Lipchitz continuous, there exists $L > 0$ such that E is not continuously differentiable [32]:

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

According to Assumption(1), on the other hand, it is unambiguous that there are positive constants B such that

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

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

Lemma (1)

*name of corresponding author

SinkrOn JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Assume that Assumption (1) and Equation (10) are true in this situation. In the case of any conjugate gradient approach in (6) and (7), where $d_k$is a reasonable direction and $\alpha_k$is produced by the S.W.L.S., the method should be considered. If

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

then we have

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

More information may be found in the following papers [33],[34],[32],[33] .

Theorem (2)

In this case, assume that Assumption (1), equation (6), and the descent condition are correct. Take, for example, a conjugate gradient scheme of the type

$$d_{k+1} = -g_{k+1} + \beta_k^{NEW2} d_k$$

Where α_k is determined using a strong Wolfe line search condition (for more information, see [36] [37] [30] [38]), If the objective function behaves evenly on set S, then

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

Proof

$$d_{k+1} = -g_{k+1} + \beta_k^{\boldsymbol{NEW2}} d_k$$

$$\beta_k^{\boldsymbol{NEW2}} = \begin{cases} \dfrac{(\theta_k - \gamma_k) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \beta_k^{LS}, & if \quad g_{k+1}^T d_k \neq 0 \\ \beta_k^{LS}, & if \quad g_{k+1}^T d_k = 0 \end{cases}$$

Where $\theta > \gamma$ and $\theta, \gamma \in (0,1)$.

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

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

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

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

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

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

NUMERICAL EXAMPLES

The conjugate gradient algorithm developed to teach the fuzzy neural networks described in Part Three 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) [39], The developed algorithm NEW2 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 the hard disk drive.

ATA:- Average Training Accuracy

ATE:- Average Training Error

Table(1) Real-World Classification Problems [39]

No.	*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***

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table. 2. An average comparison of classification problems for the New2 algorithm

Datasets	*Algorithm*	*No. of training iteration*	*Average training time*	*Average training acc*	*Average test acc*	*Average training MSE*	*Average test MSE*
Iris	*LS*	*100*	*0.1130*	*0.8867*	*0.8833*	*0.2193*	*0.2291*
	PRP	*100*	*0.0955*	*0.8867*	*0.8833*	*0.2194*	*0.2292*
	NEW2	*100*	*0.1021*	*0.9333*	*0.9433*	*0.1337*	*0.1379*
Thyroid	*LS*	*100*	*0.5069*	*0.6713*	*0.6767*	*0.4365*	*0.4252*
	PRP	*100*	*0.4885*	*0.8930*	*0.8744*	*0.1632*	*0.1829*
	NEW2	*100*	*0.7582*	*0.9225*	*0.9140*	*0.1323*	*0.1507*
Glass	*LS*	*100*	*0.7398*	*0.3252*	*0.3065*	*0.7710*	*0.7881*
	PRP	*100*	*0.7507*	*0.3159*	*0.3103*	*0.8260*	*0.8484*
	NEW2	*100*	*0.7968*	*0.4561*	*0.3626*	*0.6763*	*0.7235*
Wine	*LS*	*100*	*0.4370*	*0.6494*	*0.5573*	*0.4265*	*0.4610*
	PRP	*100*	*0.4282*	*0.9438*	*0.8944*	*0.1490*	*0.1832*
	NEW2	*100*	*0.4350*	*0.9596*	*0.9213*	*0.1356*	*0.1666*
Breast Cancer	*LS*	*100*	*1.5598*	*0.5339*	*0.5222*	*0.6884*	*0.6972*
	PRP	*100*	*1.5122*	*0.5339*	*0.5222*	*0.6851*	*0.6919*
	NEW2	*100*	*1.5703*	*0.6441*	*0.6365*	*0.5593*	*0.5653*
Sonar	*LS*	*100*	*2.1963*	*0.5442*	*0.5077*	*0.6005*	*0.6049*
	PRP	*100*	*2.1944*	*0.5635*	*0.5288*	*0.5990*	*0.6032*
	NEW2	*100*	*2.1991*	*0.7192*	*0.6346*	*0.3826*	*0.4721*

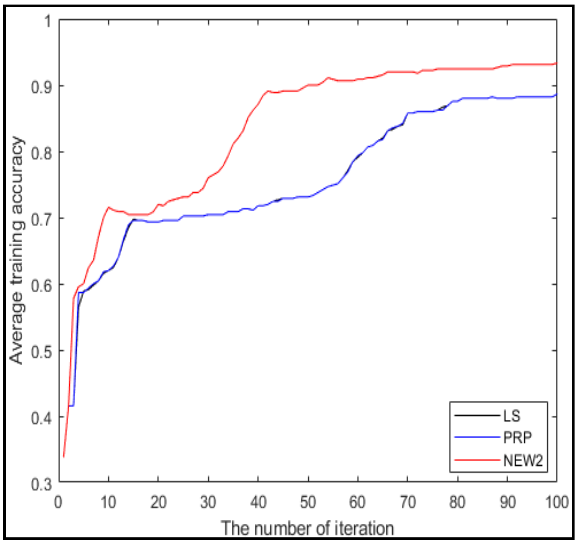

Fig. 2 ATA for Iris

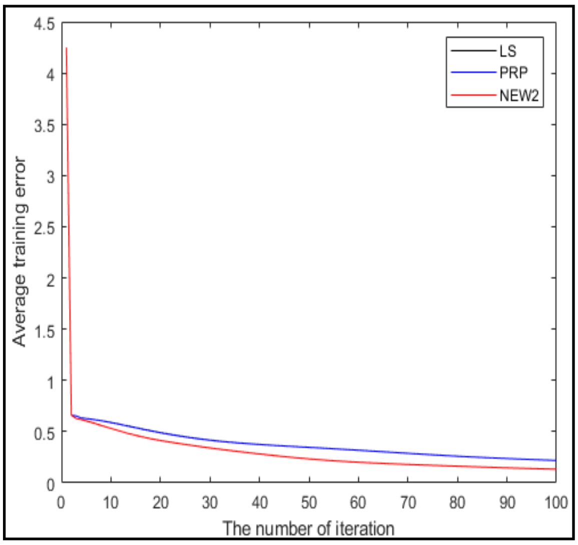

Fig. 3 ATE results for Iris

*name of corresponding author

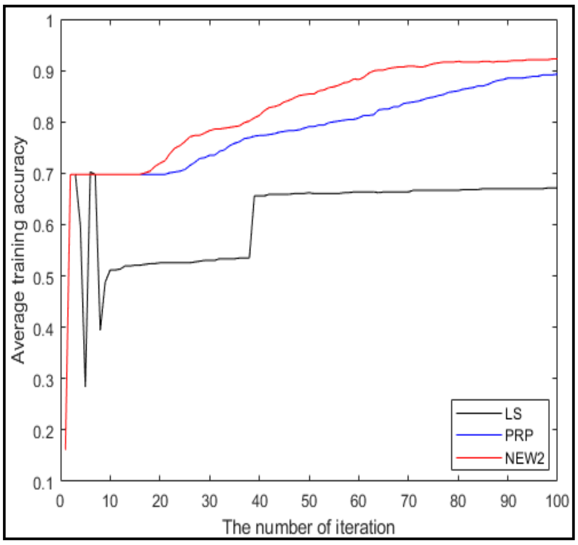

Fig. 4 ATA for Thyroid

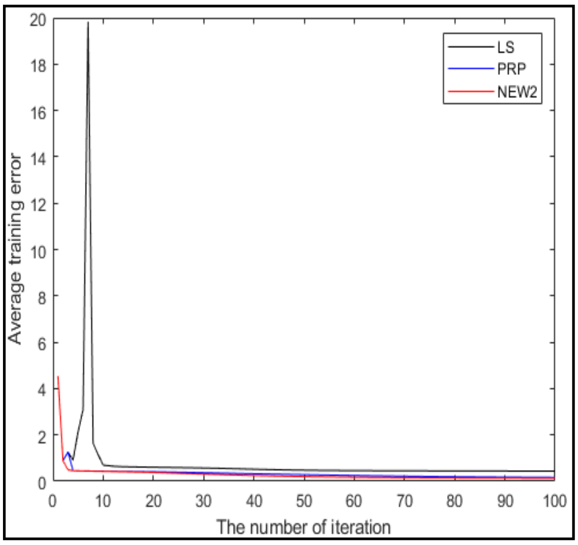

Fig. 5 ATE results for Thyroid

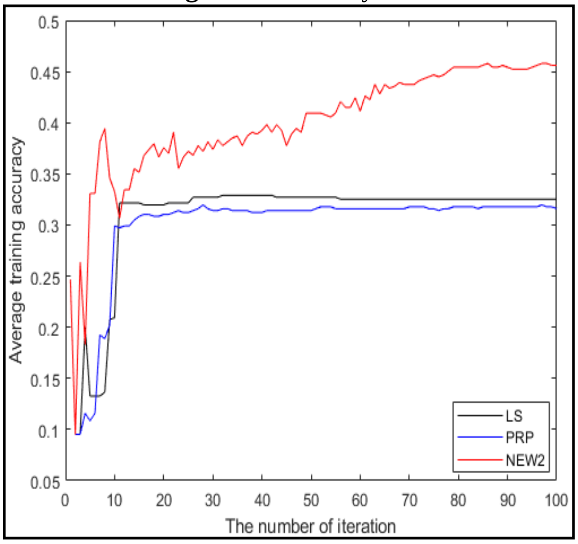

Fig. 6 ATA for Glass

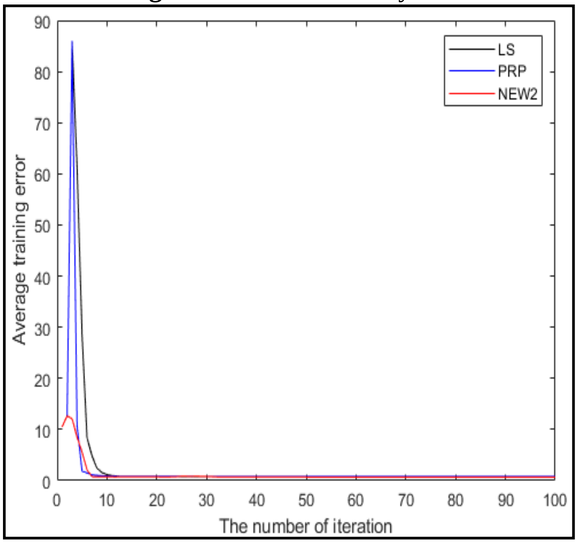

Fig. 7 ATE results for Glass

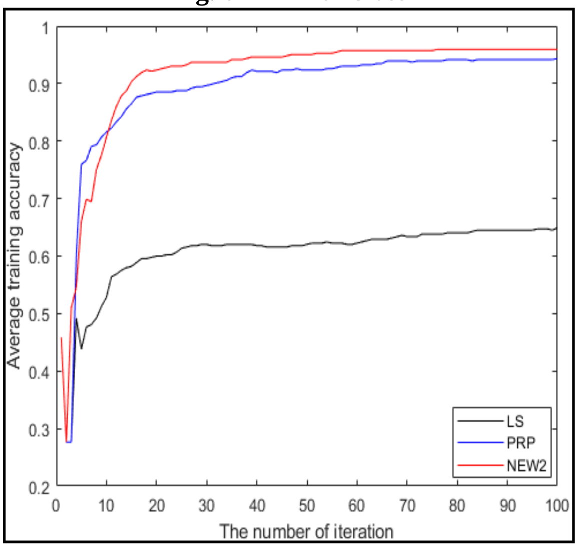

Fig. 8 ATA for Wine

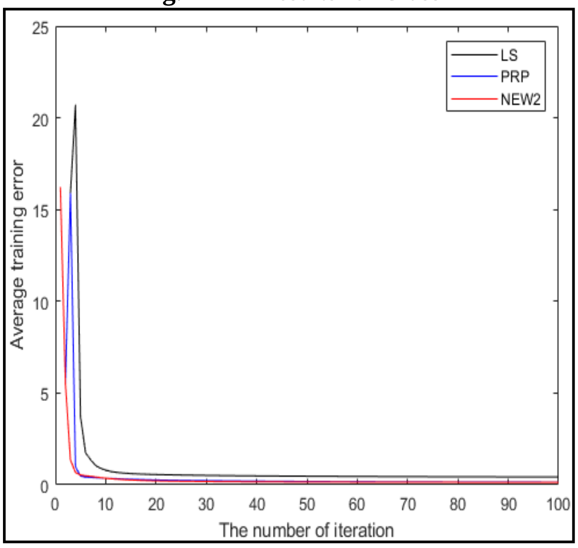

Fig. 9 ATE for Wine

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442

e-ISSN : 2541-2019
p-ISSN : 2541-044X

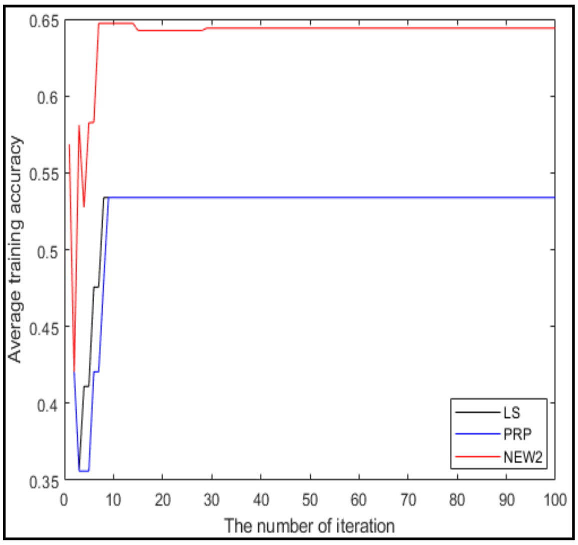

Fig. 10 ATA for Breast Cancer

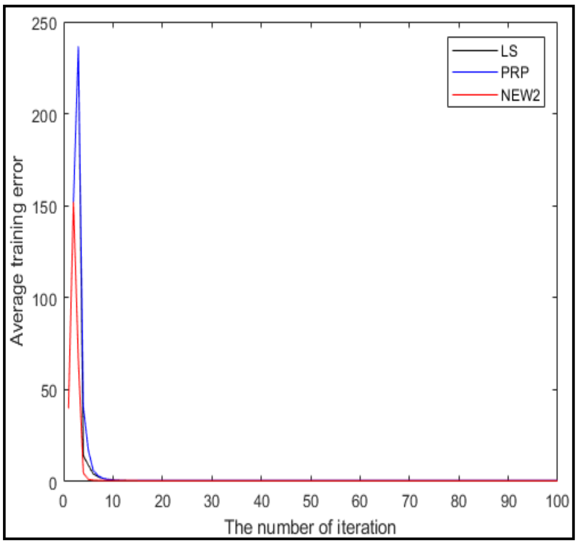

Fig. 11 ATE results for Breast Cancer

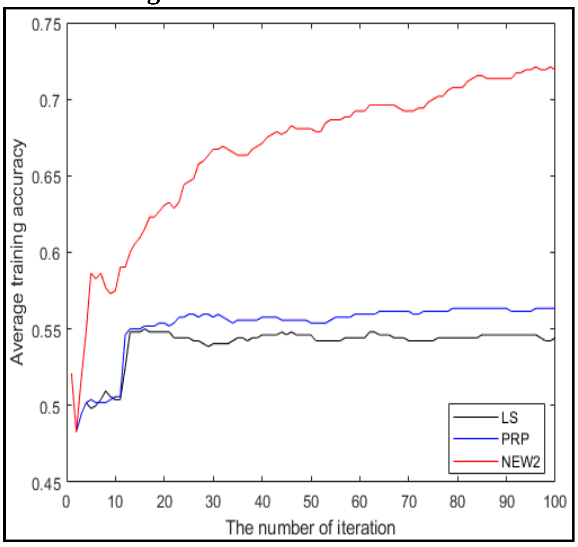

Fig. 12 ATA for Sonar

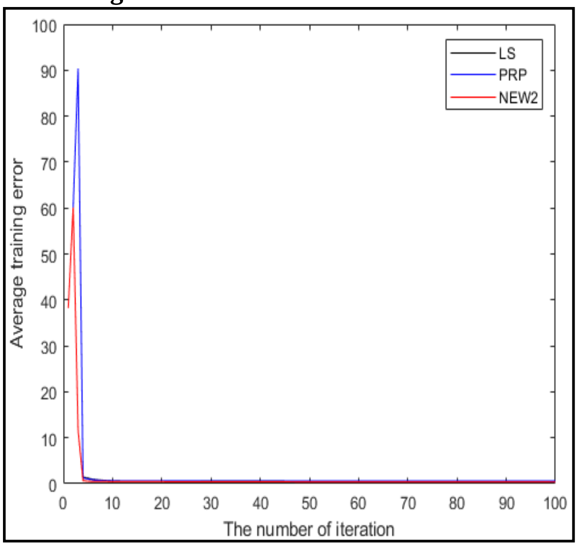

Fig. 13 ATE results for Sonar

CONCLUSION

Our Conjugate gradient technique, which looks for a conjugate descent path with adaptive learning coefficients, is a good alternative to a gradient descent approach because of its faster convergence speed. This paper proposes a modified conjugate gradient method for training the Takagi-Sugeno 0-order fuzzy neural network system (TS). The NEW2 algorithm has a higher generalization performance than its existing predecessors, according to numerical simulations. In addition, the simulations show that the proposed algorithm's converging behavior is excellent. We also conclude that the proposed method can solve optimization functions and can be applied to artificial neural networks training (For example, predicting epidemic diseases such as Corona and SARS).

ACKNOWLEDGMENTS

Thank you to Dr. Tao Gao and Dr. Jian Wang from the Colleges of Information, Control Engineering, and Science at the China University of Petroleum for their great assistance.

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.*, vol. 45, no. 2, pp. 340–353, 2015, doi:

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442
e-ISSN : 2541-2019
p-ISSN : 2541-044X

10.1109/TCYB.2014.2308631.

[2] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets Syst.*, vol. 28, no. 1, pp. 15–33, 1988, doi: 10.1016/0165-0114(88)90113-3.

[3] T. Takagi and M. Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, no. 1, pp. 116–132, 1985, doi: 10.1109/TSMC.1985.6313399.

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

[5] J. C. Bezdek, "Objective Function Clustering," in *Pattern Recognition with Fuzzy Objective Function Algorithms*, Springer, 1981, pp. 43–93.

[6] R. R. Yager and D. P. Filev, "Generation of fuzzy rules by mountain clustering," *J. Intell. Fuzzy Syst.*, vol. 2, no. 3, pp. 209–219, 1994, doi: 10.3233/IFS-1994-2301.

[7] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517–530, 2005, doi: 10.1109/TFUZZ.2004.840099.

[8] H. Shahparast, E. G. Mansoori, and M. Zolghadri Jahromi, "AFCGD: an adaptive fuzzy classifier based on gradient descent," *Soft Comput.*, vol. 23, no. 12, pp. 4557–4571, 2019, doi: 10.1007/s00500-018-3485-2.

[9] 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.*, vol. 47, no. 8, pp. 2005–2020, 2017, doi: 10.1109/TSMC.2016.2598270.

[10] X. Gu, F.-L. Chung, and S. Wang, "Bayesian Takagi–Sugeno–Kang fuzzy classifier," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1655–1671, 2016.

[11] W. Wu, L. Li, J. Yang, and Y. Liu, "A modified gradient-based neuro-fuzzy learning algorithm and its convergence," *Inf. Sci. (Ny).*, vol. 180, no. 9, pp. 1630–1642, 2010, doi: 10.1016/j.ins.2009.12.030.

[12] A. Ghosh, N. R. Pal, and J. Das, "A fuzzy rule based approach to cloud cover estimation," *Remote Sens. Environ.*, vol. 100, no. 4, pp. 531–549, 2006, doi: 10.1016/j.rse.2005.11.005.

[13] N. R. Pal and S. Saha, "Simultaneous structure identification and fuzzy rule generation for Takagi–Sugeno models," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 38, no. 6, pp. 1626–1638, 2008.

[14] J. Wang, W. Wu, and J. M. Zurada, "Deterministic convergence of conjugate gradient method for feedforward neural networks," *Neurocomputing*, vol. 74, no. 14–15, pp. 2368–2376, 2011.

[15] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, vol. 49, no. 1. NBS Washington, DC, 1952.

[16] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *Comput. J.*, vol. 7, no. 2, pp. 149–154, 1964, doi: 10.1093/comjnl/7.2.149.

[17] E. Polak and G. Ribiere, "Note sur la convergence de méthodes de directions conjuguées," *Rev. française d'informatique Rech. opérationnelle. Série rouge*, vol. 3, no. 16, pp. 35–43, 1969, doi: 10.1051/m2an/196903r100351.

[18] 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.*, vol. 27, no. 5, pp. 1175–1189, 2016, doi: 10.1007/s00521-015-1917-9.

[19] 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*, vol. 6, pp. 41551–41565, 2018, doi: 10.1109/ACCESS.2018.2848117.

[20] 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.*, vol. 16, no. 3, pp. 761–778, 2008.

[21] H. M. Khudhur and K. I. Ibraheem, "Metaheuristic optimization algorithm based on the two-step Adams-Bashforth method in training multi-layer perceptrons," *Eastern-European J. Enterp. Technol.*, vol. 2, no. 4 (116), pp. 6–13, Apr. 2022, doi: 10.15587/1729-4061.2022.254023.

[22] K. T. Chaturvedi, M. Pandit, and L. Srivastava, "Modified neo-fuzzy neuron-based approach for economic and environmental optimal power dispatch," *Appl. Soft Comput. J.*, vol. 8, no. 4, pp. 1428–1438, 2008, doi: 10.1016/j.asoc.2007.10.010.

[23] H. O. Hassan, H. M. Khudhur, and K. A. Saad, "Approximate Bessel functions using artificial neural networks," *Al Acad. J. Basic Appl. Sci.*, vol. 3, no. 3, pp. 1–12, 2021.

[24] S. Alla and H. MOHAMMED, "Developing a New Optimization Algorithm to Predict the Risk of Car Accidents Due to Drinking Alcoholic Drinks by Using Feed-Forward Artificial Neural Networks," *J. Multidiscip. Model. Optim.*, vol. 4, no. 1, pp. 11–18, 2021.

[25] C.-F. Juang and J.-S. Chen, "Water bath temperature control by a recurrent fuzzy controller and its FPGA implementation," *IEEE Trans. Ind. Electron.*, vol. 53, no. 3, pp. 941–949, 2006.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022r
DOI : https://doi.org/10.33395/sinkron.v7i3.11442

e-ISSN : 2541-2019
p-ISSN : 2541-044X

[26] H. M. Khudhur and K. K. Abbo, "A New Conjugate Gradient Method for Learning Fuzzy Neural Networks," *J. Multidiscip. Model. Optim.*, vol. 3, no. 2, pp. 57–69, 2020.
[27] Y. Liu and C. Storey, "Efficient generalized conjugate gradient algorithms, part 1: Theory," *J. Optim. Theory Appl.*, vol. 69, no. 1, pp. 129–137, 1991, doi: 10.1007/BF00940464.
[28] 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.*, vol. 21, no. 1, pp. 118–123, 2018.
[29] Y. A. Laylani, K. K. Abbo, and H. M. Khudhur, "Training feed forward neural network with modified Fletcher-Reeves method," *J. Multidiscip. Model. Optim.*, vol. 1, no. 1, pp. 14–22, 2018, [Online]. Available: http://dergipark.gov.tr/jmmo/issue/38716/392124#article_cite.
[30] 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.*, vol. 21, no. 1, pp. 124–129, 2015.
[31] M. M. Abed, U. Öztürk, and H. Khudhur, "Spectral CG Algorithm for Solving Fuzzy Non-linear Equations," *Iraqi J. Comput. Sci. Math.*, vol. 3, no. 1, 2022.
[32] 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.*, vol. 5, no. 7, 2016.
[33] Z. M. Abdullah, M. Hameed, M. K. Hisham, and M. A. Khaleel, "Modified new conjugate gradient method for Unconstrained Optimization," *Tikrit J. Pure Sci.*, vol. 24, no. 5, pp. 86–90, 2019.
[34] H. M. Khudhur, "Numerical and analytical study of some descent algorithms to solve unconstrained Optimization problems," University of Mosul, 2015.
[35] 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.*, vol. 8, pp. 1–9, 2018.
[36] M. Al-Baali, "Descent property and global convergence of the Fletcher—Reeves method with inexact line search," *IMA J. Numer. Anal.*, vol. 5, no. 1, pp. 121–124, 1985.
[37] L. Zhang and W. Zhou, "Two descent hybrid conjugate gradient methods for optimization," *J. Comput. Appl. Math.*, vol. 216, no. 1, pp. 251–264, 2008, doi: 10.1016/j.cam.2007.04.028.
[38] 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.*, vol. 13, no. 2, pp. 101–113, 2018.
[39] 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*, vol. 364, pp. 168–181, 2019, doi: 10.1016/j.neucom.2019.07.035.

*name of corresponding author