



Calibrating wavelet neural networks by distance orientation similarity fuzzy C-means for approximation problems



Pauline Ong^{a,b,*}, Zarita Zainuddin^b

^a Faculty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), 86400 Parit Raja, Batu Pahat, Johor, Malaysia

^b School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

ARTICLE INFO

Article history:

Received 24 June 2015

Received in revised form 6 January 2016

Accepted 24 January 2016

Available online 3 February 2016

Keywords:

Clustering

Distance similarity

Function approximation

Orientation similarity

Time series

Wavelet neural networks

ABSTRACT

Improperly tuned wavelet neural network (WNN) has been shown to exhibit unsatisfactory generalization performance. In this study, the tuning is done by an improved fuzzy C-means algorithm, that utilizes a novel similarity measure. This similarity measure takes the orientation as well as the distance into account. The modified WNN was first applied to a benchmark problem. Performance assessments with other approaches were made subsequently. Next, the feasibility of the proposed WNN in forecasting the chaotic Mackey–Glass time series and a real world application problem, i.e., blood glucose level prediction, were studied. An assessment analysis demonstrated that this presented WNN was superior in terms of prediction accuracy.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Artificial neural networks (ANNs) – computing technology inspired by biological neural networks, have been implemented vigorously in function approximation problems. The popularity of ANNs is due to their attention-grabbing characteristics of self-learning capability, massively parallel architecture, fault tolerance characteristics, and the ability to approximate complex nonlinear mapping directly from the input samples [1]. The combination of the wavelet decomposition with single hidden layer ANNs paved the way for an even more appealing network model, namely, the wavelet neural networks (WNNs). With sufficient training, WNNs can approximate any continuous function to any desired degree of accuracy [2]. Moreover, the constitution of the localized wavelet activation function in the hidden layer facilitates a more compact network topology. The utilization of the wavelet activation function also leads to faster learning ability, as compared to conventional multilayer perceptrons [3]. Even though the WNNs are continuously gaining prevalence [4,5], properly designing a WNN is a critical issue that needs to be addressed cautiously. Improperly modeling a WNN may jeopardize its generalization capability and predicting competence.

Typically, construction of WNNs involves two phases: (i) structure initialization, i.e., specifying the number of hidden nodes, the initial parameters of wavelet activation functions and the initial weight vectors of the WNNs, and (ii) parameter optimization, i.e., solving the connection weight vectors between the hidden layer and the output layer of the WNNs. The focus of this study lies in the determination of the appropriate initial values of the translation parameters. This is crucial in order to achieve a better configuration with fewer hidden nodes, simpler structure and better accuracy. A proper initialization of the translation parameter will do a good job of reflecting the essential attributes of the input samples, which is vital to generate optimal interpolation between the input–output spaces.

Various efforts have been directed to solve this concern, in which the initialization of the translation vectors using an explicit expression is one of the widely used approaches. A recursive mathematical formulation which considered the domain of the function, in conjunction with the stochastic gradient method, was applied by Zhang and Benveniste [2] and Oussar and Dreyfus [6] in WNNs initialization. These developed techniques have been highlighted in several WNNs studies [7–10]. However, using an explicit expression to define translation vectors is merely a rule of thumb. The explicit expression relies on the characteristics of the dataset, which are usually unknown beforehand and vary in different situations. Thus, a tailor-made equation for all cases is rare. Determining the WNNs translation vectors based on the clustering algorithm was another alternative that has been

* Corresponding author at: Faculty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), 86400 Parit Raja, Batu Pahat, Johor, Malaysia. Tel.: +60 74564479.

E-mail address: ongp@uthm.edu.my (P. Ong).

proposed. Wei et al. [11] used hierarchical clustering to determine a set of the best initial translation vectors based on the obtained tree structure. A Euclidean distance-based connectionist clustering method was presented by Lin [12], in which the cluster means of the input data were positioned as the translation vectors. Hwang et al. [13] employed the K-means clustering algorithm to find the optimal locations of the translation vectors. The obtained solutions were updated subsequently by using a dyadic selection scheme in order to eliminate unnecessary hidden nodes. However, unresolved issues remain with the aforementioned clustering algorithms. For instance, the Euclidean distance-based clustering approaches do not consider the non-spherical occupancy in the feature space [14]. For hierarchical clustering, intricacy exists in defining the initial values from the branches of the obtained tree structure [15]. Support vector machines (SVMs) [16] and genetic algorithm (GA) [17] were among the other attempts used in defining the appropriate initial values of the translation vectors. Although such improvements were noteworthy, drawbacks such as algorithmic complexity and time-consuming programming execution occurred during the implementation of the SVMs and GA [18,19].

In this paper, a distance orientation similarity-based fuzzy C-means algorithm (DOS-FCM) – an alternative to existing translation parameter approaches, was developed. The theoretical rationality of using the DOS-FCM in WNNs initialization could be attributed to reasons such as: (i) The DOS-FCM is insensitive to the input data domain, as compared to the commonly used recursive mathematical formulation which takes the input data domain into consideration; (ii) the DOS-FCM does not require prior knowledge of the essential attributes residing in the input space; (iii) the Euclidean-based clustering approaches may fail in reflecting the essential attributes if the data are in non-spherical distribution, but the DOS-FCM works in such cases; (iv) the DOS-FCM involves both coarse-tuning and fine-tuning processes. In this regard, it is insensitive to the initial values of the cluster centers compared to the hierarchical clustering approach; and (v) the control parameters of the mutation rate, crossover rate, mutation percentage in population and the number of maximum generation in the GA need to be assigned judiciously as it may be prone to premature convergence. The DOS-FCM, however, has fewer control parameters to be addressed. The potential benefit of the DOS-FCM has been validated quantitatively and qualitatively for a wide range of artificial and real world data with various geometrical structures and overlapping characteristics. The details are omitted here for conciseness [20].

The performance of the proposed WNNs with DOS-FCM was compared to WNNs with random and fuzzy C-mean (FCM) initialization methods. Simulations were carried out on a few artificial benchmark approximation functions and a real-world application problem.

The paper is organized as follows. Section 2 presents a brief introduction of WNNs, followed by a description of the proposed DOS-FCM algorithm in Section 3. The details on how the DOS-FCM was integrated into the WNNs are described in Section 4. Section 5 presents the experimental simulations and comparisons of the proposed modified WNNs in approximating the nonlinear functions and real-world data followed by a discussion of the results obtained. Finally, Section 6 contains concluding remarks of this study.

2. Wavelet neural networks for function approximation

2.1. Wavelet neural networks architecture

WNNs are three-layered feedforward neural networks with single input layer, hidden layer and output layer. A schematic diagram

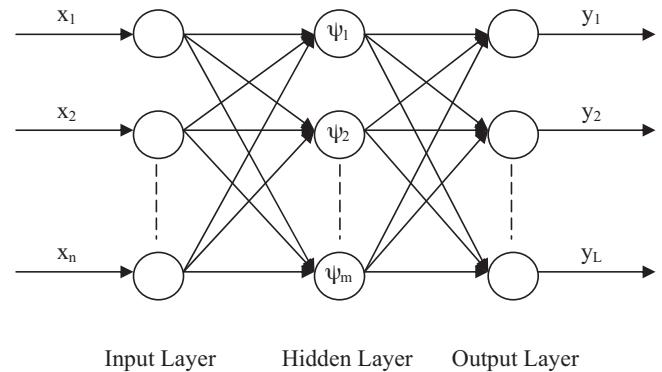


Fig. 1. A schematic diagram for a WNN with n input nodes, m hidden nodes and L output nodes.

of a WNN architecture is illustrated in Fig. 1. The nomenclature and operations involved in each layer are presented as follows:

Input Layer: The input layer receives input data $x = [x_1, x_2, \dots, x_n]$, where n is the number of input dimensions, i.e., the number of input nodes. Subsequently, the accepted input will be transmitted to the hidden layer.

Hidden Layer: In this present work, Mexican Hat wavelet function is integrated as the activation function in the hidden nodes of this layer. The relationship between the wavelet function and the received input data is given as follows:

$$\Phi_{a_{ij}, b_{ij}}(x_i) = \left(n - 2 \left(\frac{x_i - b_{ij}}{a_{ij}} \right)^2 \right) \exp \left(-\frac{(x_i - b_{ij})^2}{a_{ij}^2} \right),$$

for $i = 1, \dots, n$ and $j = 1, \dots, m$ (1)

where a_{ij} and b_{ij} are the dilation and translation parameters of the j th activation function of the i th input variable x_i , respectively. The output of each node in this layer is expressed as the product of the j th multi-dimensional wavelet with n input dimensions of x_i , which is represented as:

$$\begin{aligned} \psi_j(x) &= \prod_{i=1}^n \Phi_{a_{ij}, b_{ij}}(x_i) \\ &= \prod_{i=1}^n \left(n - 2 \left(\frac{x_i - b_{ij}}{a_{ij}} \right)^2 \right) \exp \left(-\frac{(x_i - b_{ij})^2}{a_{ij}^2} \right). \end{aligned} \quad (2)$$

The products of the hidden layer are then propagated to the output layer.

Output Layer: The k th output of the WNN will be the linear combination of the weighted sum of the hidden layer, which is defined in the form of:

$$\hat{y}_k = \sum_{j=1}^m w_{jk} \psi_j = \sum_{j=1}^m w_{jk} \prod_{i=1}^n \left(n - 2 \left(\frac{x_i - b_{ij}}{a_{ij}} \right)^2 \right) \exp \left(-\frac{(x_i - b_{ij})^2}{a_{ij}^2} \right),$$

for $k = 1, 2, \dots, L$ (3)

where w_{jk} represent the connecting weights between the j th hidden node and k th output node.

2.2. Learning algorithm of wavelet neural networks

As mentioned in the introduction, the typical learning process of a WNN involves two stages, i.e., structure initialization and parameter optimization. In this present work, the initial values of b_{ij} will

be chosen systematically with the proposed DOS-FCM approach, discussed in the next section.

The basic idea of parameter optimization in a WNN is to adjust the weight vectors linking the hidden layer and output layer to map the underlying relationship between the input and output space. The network output weights are derived analytically by applying the pseudo-inverse of the hidden layer output matrix. This solves the weakness of the gradient-based learning algorithms where the networks parameters are adjusted step by step, which are usually slow in learning and get trapped in local minima easily [21].

For brevity, Eq. (3) is written as $Y = \Psi W$, where

$$\Psi = \begin{pmatrix} \psi(x_1, a_1, b_1) & \psi(x_1, a_2, b_2) & \dots & \psi(x_1, a_m, b_m) \\ \psi(x_2, a_2, b_2) & \psi(x_2, a_1, b_1) & \dots & \psi(x_2, a_m, b_m) \\ \vdots & \vdots & \dots & \vdots \\ \psi(x_n, a_1, b_1) & \psi(x_n, a_2, b_2) & \dots & \psi(x_n, a_m, b_m) \end{pmatrix}_{n \times m}, \quad (4)$$

denotes the hidden layer output matrix. In most cases, the number of input dimension n is less than the number of training samples m . This indicates that Ψ is a non-square matrix and the least-squares solution of output weight vector W such that $Y = \Psi W$ might not exist. However, assuming that Ψ has full rank, it has been shown that the smallest norm least-squares solution of output weight vector W can be obtained from [21]:

$$W = \Psi^+ Y, \quad (5)$$

where $\Psi^+ = \Psi^T (\Psi \Psi^T)^{-1}$ is the right pseudo-inverse of the hidden layer output matrix Ψ . It is worth mentioning that the right pseudo-inverse is used in this case since $n < m$.

3. The proposed distance orientation similarity-based fuzzy C-means algorithm

Clustering is the act of partitioning a set of observations into dissimilar clusters. The observations belonging to the same cluster will have a stronger degree of similarity than the observations in the others. FCM which integrates the concept of fuzzy partitioning is one of the most widely used clustering algorithms. Compared to crisp clustering, FCM assigns any given observation into more than one cluster. The degree of belongingness of each observation with respect to the particular cluster is graded by the membership function which is quantified by the Euclidean distance between the observation and the cluster centers. Nonetheless, utilization of the Euclidean distance as a measure of similarity in the FCM is unfavorable due to its drawback of not being able to detect the non-spherical shaped clusters.

Instead of assigning each observation to a cluster based on the shortest distance measure, a novel similarity metric that considers both the distance similarity and orientation similarity of the observations with respect to the cluster centers is proposed. For convenience, the proposed novel similarity measure is denoted as DOS-Index, as it measures the distance similarity and orientation similarity of the observations in the feature space.

Given a set of observations, $p_i, i = 1, 2, \dots, N$, in order to determine the most similar observation p_j for observation p_i with respect to a cluster center c_k in terms of distance and orientation aspects, the proposed DOS-Index is defined as:

DOS-Index(p_i, c_k, p_j)

$$= \max_{1 \leq j \leq N, p_j \in C_k} \frac{\text{DSL}(p_i, c_k, p_j) + \text{OSL}(p_i, c_k, p_j)}{2}. \quad (6)$$

The observation p_j is said to be the most similar observation for observation p_i relative to cluster center c_k , if its DOS-Index has the highest value among all considered observations. As observed from this equation, the DOS-Index characterizes the sense of similarity between observation p_i and observation p_j with respect to a cluster center c_k from two aspects, i.e., the distance similarity level (DSL) which takes the distance between p_i to c_k , and distance between p_j to c_k into account, and the orientation similarity level (OSL) which considers the orientation angle between the line segments $\rightarrow p_i c_k$ and $\rightarrow c_k p_j$ based on the scalar projection concept. The DSL and OSL are expressed as:

$$\text{DSL}(p_i, c_k, p_j) = 1.5 - \frac{2}{1 + \exp(-|d_i - d_j|)}, \quad (7)$$

where $d_i = \overline{p_i c_k}$ and $d_j = \overline{p_j c_k}$, and

$$\text{OSL}(p_i, c_k, p_j) = \frac{v_i \cdot v_j}{2||v_i|| ||v_j||} \quad (8)$$

where $v_i = \rightarrow p_i c_k$ and $v_j = \rightarrow c_k p_j$.

3.1. The workflow of the proposed DOS-FCM algorithm

The developed DOS-Index is subsequently integrated as the similarity measure in the conventional FCM algorithm. Presenting a set of observations p_i , where $i = 1, \dots, N$, the proposed DOS-FCM algorithm works as follows:

-
- | | |
|--|---|
| Step 1:
Initialization
Initialize the membership function u_{ij} randomly, where $u_{ij} \in [0, 1]$,
$\sum_{l=1}^K u_{il} = 1, \forall j = 1, \dots, N$ and K is the number of cluster centers. | Step 2:
Coarse-tuning
Use the conventional FCM algorithm to update the membership function. |
| Step 3:
Fine-tuning 1
Find out the set S_{bik} of all possible similar observations p_j for each p_i , such that $\text{OSL}(p_i, c_k, p_j) \geq \beta (= 0.47)$ and for $\text{DSL}(p_i, c_k, p_j)$, the interval $0 \leq d_i - d_j < 2$ is held. The S_{bik} denotes the collection of all the observations p_j which satisfy the conditions $\text{OSL}(p_i, c_k, p_j) \geq \beta$ and $0 \leq d_i - d_j < 2$ for each p_i , while c_k is the k th cluster center. | Step 4:
Fine-tuning 2
Compute Eq. (6) for each observation p_i .
If the value of $\text{DOS-Index}(p_i, c_k^*)$ is the largest and the most similar observation p_j , relative to c_k^* belongs to S_{bik} , assign observation p_i to the cluster center c_k^* .
Otherwise, assign observation p_i to the cluster center c_k^{**} with the shortest distance.
Next, update the membership matrix u_{ij} based on the criterion that if $\text{DOS-Index}(p_i, c_k^*) > \theta \in [0, 1]$, |
| $u_{ij} = \begin{cases} u_{ij} = 1, & \text{if } j = k \\ u_{ij} = 0, & \text{if } j \neq k \end{cases}$ | $u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ji}}{d_{ki}} \right)^{2/(m-1)}}$
where c is the number of cluster centers. |
| Step 5:
Updating
Update the new cluster center for K clusters by
$c_k^{\text{new}} = \frac{\sum_{i=1}^n u_{ji}^m p_i}{\sum_{i=1}^n u_{ji}^m}$,
where n is the number of observations in cluster c_k and $p_i \in c_k$. | Step 6:
Continuation
If there are no pattern changes or when the iterations reach a predefined maximum value, then stop. Otherwise, go to Step 3. |
-

In Step 3, the threshold value $\beta = 0.47$ is imposed on the $\text{OSL}(p_i, c_k, p_j)$ based on the assumption that the tolerance rate of the angle of orientation, θ cannot be greater than 20° , as suggested in [20]. As shown in Fig. 2, if this constraint is violated, the observation p_j is considered as not having a sense of similarity in terms of orientation

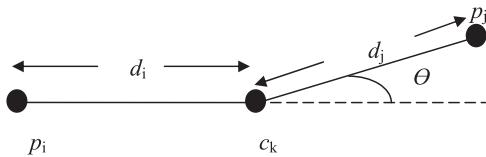


Fig. 2. An example to illustrate the DOS-Index.

angle with the observation p_i , with respect to the cluster center c_k . Thus, from Eq. (8),

$$\beta = \frac{v_i \cdot v_j}{2||v_i|| ||v_j||} = \frac{||v_i|| ||v_j|| \cos \theta}{2||v_i|| ||v_j||} = \frac{\cos 20^\circ}{2} = 0.47$$

Similarly, it is assumed that observation p_j is no longer considered similar to observation p_i if their difference in distance is more than 2. Hence, the value of $|d_i - d_j|$ in DSL operator as shown in Eq. (7) is bounded by 0 and k ($k=2$ in this case) as used in Ref. [20], in order to reduce the total time complexity.

In Step 6, pattern change is defined as the absolute difference between the objective functions of two consecutive iteration steps. In other words, minimizing the objective function is the concern. In the DOS-FCM algorithm, the objective function is specified as:

$$\text{Objective function} = \sum_{i=1}^N \sum_{j=1}^K u_{ij}^m d_{ij}^2 \quad (9)$$

where $u_{ij} \in [0, 1]$, $\sum_{j=1}^K u_{ij} = 1$, $m > 1$ while d_{ij}^2 is the squared Euclidean distance between the i th fuzzy observation and the j th fuzzy cluster center [22].

The efficiency of the DOS-Index as opposed to the Euclidean distance used in the conventional FCM is illustrated in Fig. 3. As shown in this figure, there are three observations $p_1 = (8, 7.2)$, $p_2 = (2, 9.1)$, $p_3 = (12, 9.3)$ and two cluster centers $c_1 = (5, 8)$ and $c_2 = (9, 8)$. Based on the idea of the shortest Euclidean distance, the FCM assigns the

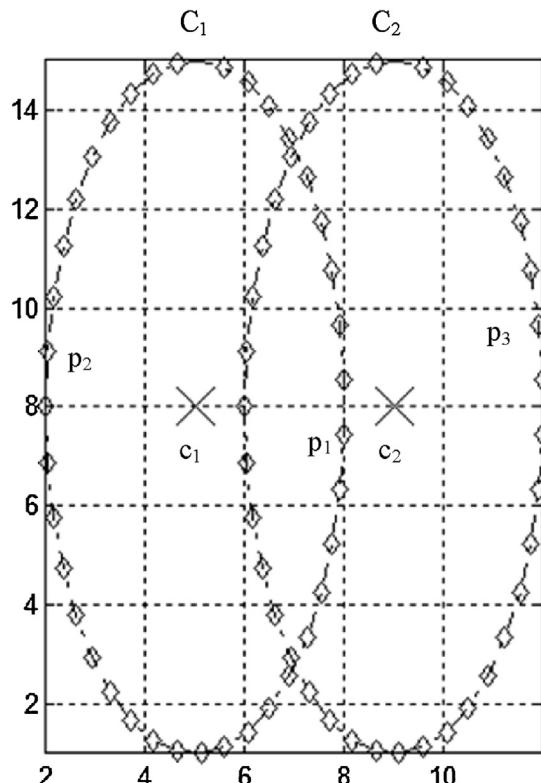


Fig. 3. An example to illustrate the effectiveness of the DOS-Index.

observation p_1 to cluster C_2 , since $\overline{p_1 c_1} = 3.1048$ and $\overline{p_1 c_2} = 1.2806$, which violate the visual inspection. On the other hand, by applying Eqs. (6)–(8):

$$\text{DSL}(p_1, c_1, p_2) = 1.5 - \frac{2}{1 + \exp(-|d_1 - d_2|)} = 0.4548$$

$$\text{OSL}(p_1, c_1, p_2) = \frac{v_1 \cdot v_2}{2||v_1|| ||v_2||} = 0.4979$$

$$\text{DOS-Index}(p_1, c_1, p_2) = \frac{\text{DSL}(p_1, c_1, p_2) + \text{OSL}(p_1, c_1, p_2)}{2} = 0.4764$$

while

$$\text{DSL}(p_1, c_2, p_3) = 1.5 - \frac{2}{1 + \exp(-|d_1 - d_3|)} = -0.2593$$

$$\text{OSL}(p_1, c_2, p_3) = \frac{v_1 \cdot v_3}{2||v_1|| ||v_3||} = 0.4827$$

$$\text{DOS-Index}(p_1, c_2, p_3) = \frac{\text{DSL}(p_1, c_2, p_3) + \text{OSL}(p_1, c_2, p_3)}{2} = 0.1116$$

From the calculation of DOS-Index, $\text{DOS-Index}(p_1, c_1, p_2) > \text{DOS-Index}(p_1, c_2, p_3)$, hence, the DOS-FCM assigns observation p_1 to cluster C_1 , which is in agreement with the visual examination.

3.2. Experimental simulations of the DOS-FCM in clustering artificial dataset

The effectiveness of the DOS-FCM in comparison to the FCM will be evaluated quantitatively and qualitatively in the following two examples. The performances are assessed using two criteria.

In order to assess the effectiveness of the DOS-FCM with respect to accurate partitioning, the Minkowski (MS) score [23] is adopted for quantitative evaluation purpose. The MS score is defined as:

$$\text{MS}(T, C) = \frac{\|T - C\|}{\|T\|} \quad (10)$$

where $\|T\| = \sqrt{\sum_i \sum_j T_{ij}}$. $T_{n \times n}$ defines the reference solution, whereas $C_{n \times n}$ is the resulting solution that we wish to compare to T (n is the number of observations). The element of C_{ij} or T_{ij} is 1, if i and j belong to the same cluster, and 0 if otherwise. A zero MS score indicates that a perfect clustering result is achieved. In addition to the MS score, the computation time taken by the algorithm in order to meet the stopping criteria is recorded to assess the efficiency of the clustering model. For each artificial dataset, the simulations of the FCM and DOS-FCM are executed for 10 trials. The obtained results are averaged and summarized in Table 1.

In the first simulated artificial dataset (Fig. 4(a)), there are six spherical-shaped clusters with 50 observations each. Fig. 4(b) and (c) presents the partitioning results achieved by the FCM and the DOS-FCM, respectively. As shown in Fig. 4(b), it is not surprising that the FCM is able to provide accurate clustering result since the dataset under consideration consists of mainly hyper-spherical shaped clusters. A perfect partition is attained by the DOS-FCM as well, which can be observed in Fig. 4(c) and indicated by the MS score in Table 1. It can be concluded that the proposed DOS-FCM accounts for the existence of spherical-shaped clusters as the standard FCM.

One compact circle and two overlapping stripes with a total number of 340 observations are generated in the second artificial dataset (Fig. 5(a)). As shown in Fig. 5(b), the FCM algorithm is not able to give satisfactory partitioning results, albeit the compact circle is internally hyper-spherical in nature. This suggests the infeasibility of the FCM when the existing clusters are different in shape and intertwined, which are common in practice. Apparently, the DOS-FCM delivers a better partitioning result, which can be observed in Fig. 5(c). Moreover, Table 1 demonstrates that a lower MS score is attained by the proposed DOS-FCM. Despite

Table 1

The MS scores and computation time for the simulated artificial datasets achieved by FCM and the proposed DOS-FCM.

Dataset	FCM		DOS-FCM	
	MS score	Computation time (s)	MS score	Computation time (s)
Data 1	0	0.1966 ± 0.0054	0	103.4980 ± 2.2273
Data 2	0.7140 ± 0.1316	0.2399 ± 0.0158	0.2877 ± 0.1219	852.1351 ± 18.0524

Note: The simulation was carried out for 10 trials. The mean and standard deviations were calculated.

such improvement, however, the DOS-FCM takes longer computation time. The time complexity of the FCM is $O(i_{\text{fcm}}NdK)$, where i is the number of required iterations, N is the number of observations, d is the dimension of the observations and K is the number of cluster centers. However, the time complexity of the DOS-FCM is $O(i_{\text{fcm}}NdK) + O(i_{\text{dos-fcm}}N^2d^2K)$, where the latter is attributed to the calculation of the DOS-Index for each observation p_i with respect to each cluster center c_k at each iteration during the fine-tuning process. An acceleration strategy is would be required.

From the simulation of the artificial dataset, it can be summarized that:

- (i) Both FCM and DOS-FCM provide accurate partition for the first dataset, as indicated from the MS score of 0 in all trials.
- (ii) The DOS-FCM performs better than the FCM, when the existing clusters possess different characteristics.
- (iii) The DOS-FCM remedies the deficiency of the FCM that does not consider the non-spherical occupancy in the feature space.
- (iv) The DOS-FCM has a longer computation time than the FCM.

4. Solutions

Presenting the training samples $\{(x_i, y_i), i=1, \dots, N\}$, the DOS-FCM performs the partitioning, in which the obtained cluster centers are then positioned as the initial translation vectors of the WNNs.

The implementation of the DOS-FCM in selecting the optimal translation vectors is summarized as follows:

Step 1: A set of training samples $\{(x_i, y_i), i=1, \dots, N\}$ is presented to the WNNs.

Step 2: Apply the DOS-FCM as explained in Section 3 to form a set of optimal cluster centers of the training samples, in which the cluster centers are then assigned as the initial translation vectors for WNNs.

Step 3: Construct the WNNs as shown in Equation (3).

Step 4: Update the weight vectors adaptively by using the pseudo-inverse method in order to minimize the criterion function. Stop the learning process when the stopping criteria are met.

Step 5: Validate the forecasting abilities of the obtained WNNs network structure by using a set of testing samples.

The workflow of the integration of the DOS-FCM with the learning process of the WNNs is further summarized in Fig. 6.

5. Experimental simulations

In the previous section, the superiority of the DOS-FCM algorithm has been evaluated qualitatively and quantitatively by clustering the artificial dataset. The competence of the DOS-FCM as an efficient approach in the parameter initialization of WNNs will be explored in this section, where simulations on a few benchmark approximations and a real-world problem were carried out. All the simulations were performed using Matlab R2010a. WNNs were trained using the pseudo-inverse learning algorithm. The number of input nodes was determined by the input dimension, and the number of hidden nodes was determined by the number of training

samples (for random initialization approach) or number of cluster centers (for FCM and DOS-FCM initialization approaches). The weight vectors were initialized randomly, and the error goal was set to 0.01 as an early stopping criterion. In addition, in order to validate the statistical significance of the differences in the performance indexes obtained by the WNNs across various initialization methods, Wilcoxon signed rank test was utilized.

5.1. Benchmark piecewise function

A benchmark piecewise nonlinear function which was referred to by Zhang and Benveniste was described as [2]:

$$f(x) = \begin{cases} -2.186x - 12.864, & -10 \leq x < -2 \\ 4.264x, & -2 \leq x < 0 \\ 10 \exp(-0.05x - 0.5) \sin[(0.03x + 0.7)x], & 0 \leq x \leq 10 \end{cases} \quad (11)$$

was used to study the predictive competence of the WNNs with the proposed DOS-FCM as initialization approach. 300 training data and 150 testing data were randomly generated by a uniform distribution on the definition domain. The performance index, ERR, was utilized to assess all models. It is given as:

$$\text{ERR} = \sqrt{\frac{\sum_{j=1}^n [f(x_j) - y_j]^2}{\sum_{j=1}^n [y_j - \bar{y}]^2}} \quad (12)$$

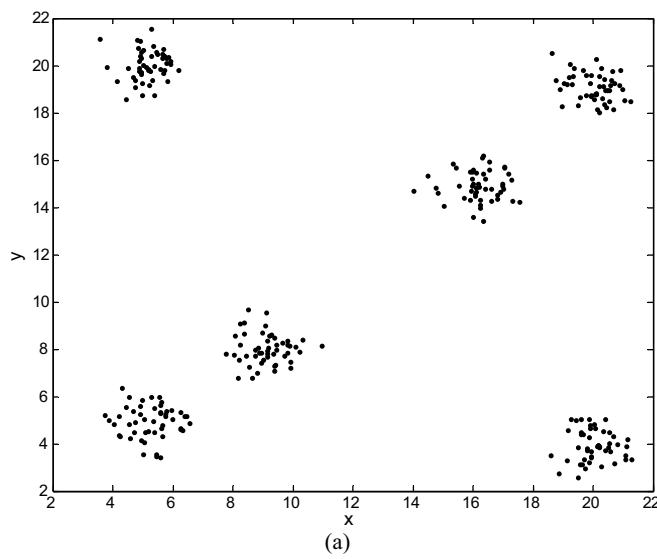
where $f(x_j)$ and y_j were the actual and predicted value at point j , respectively.

\bar{y} was defined as:

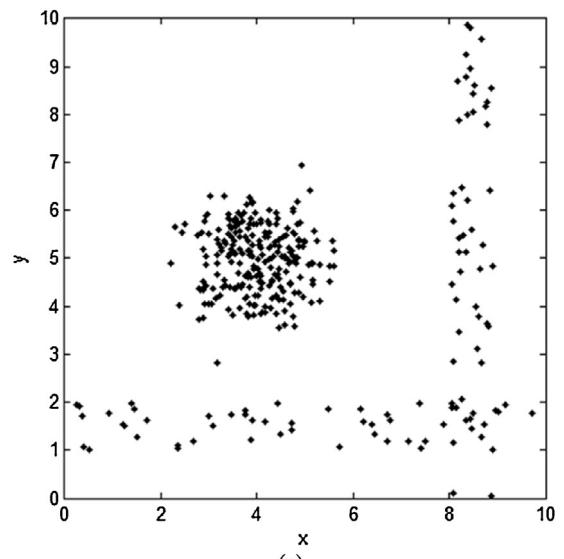
$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j \quad (13)$$

where n was the number of testing data.

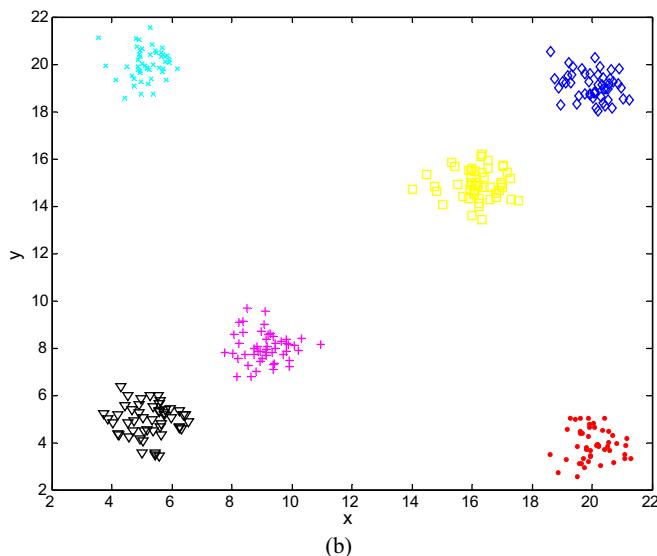
All WNNs were able to approximate the piecewise function almost perfectly since the obtained ERR values were relatively small, i.e., $3.1596e-005$ for WNNs with random initialization approach, $1.3625e-005$ for WNNs with FCM initialization approach, and $3.2005e-007$ for WNNs with DOS-FCM initialization approach. A net improvement was noticeable when clustering algorithms were adopted, as compared to the conventional WNNs that selected the initial parameters randomly. In addition, it was apparent that generalization capability of the WNNs increased the most when the WNNs were trained in conjunction with the proposed DOS-FCM algorithm. Remarkable prediction accuracy improvements of 98.99% and 97.65% were attained, respectively, as opposed to the WNNs with random initialization and FCM approach. It can be concluded that the interpolation capability of the WNNs could benefit from the proposed DOS-FCM algorithm since more effective hidden nodes were used to construct the WNNs model. In addition, statistical analysis based on the Wilcoxon test validated the statistical significance of the differences in the performance scores ($p = 8.3898e-07$ and $p = 3.1518e-22$, for random and FCM initialization approaches, respectively) at the confidence level of 5%. The approximation of the benchmark piecewise function based on the WNNs with DOS-FCM method was depicted in Fig. 7. It is clear that the obtained model was able to capture the manners of the chosen



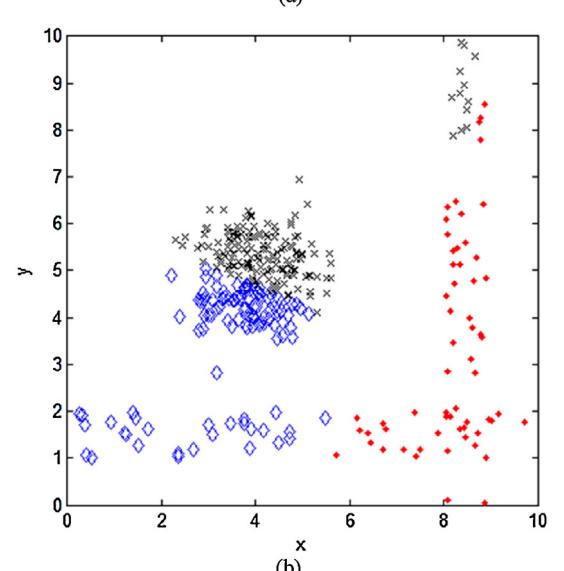
(a)



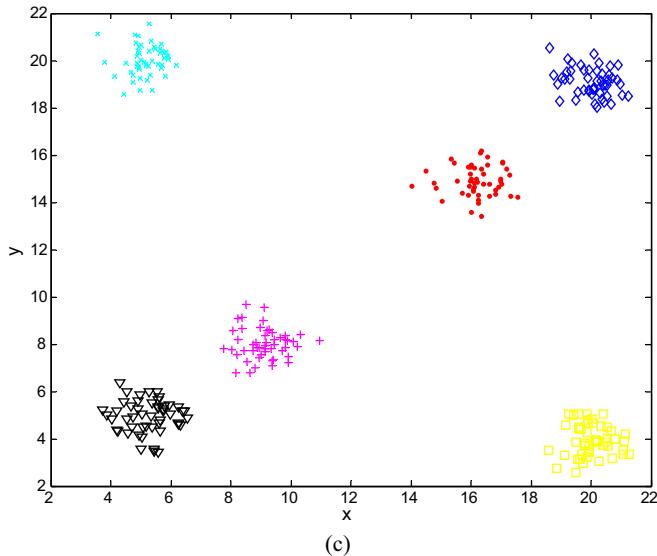
(a)



(b)

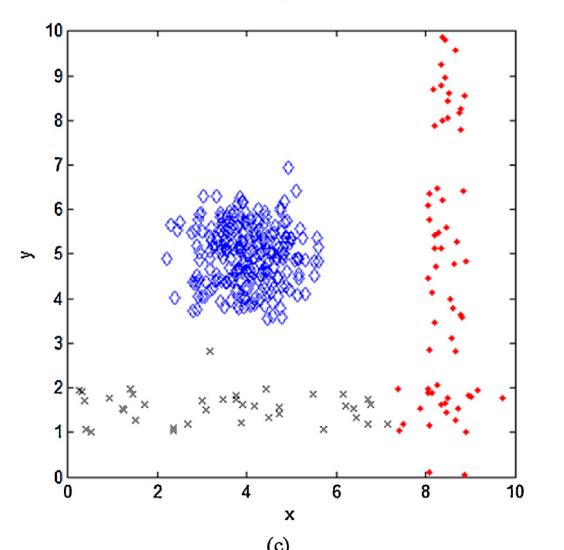


(b)



(c)

Fig. 4. Partitioning of data 1 (a) original dataset; (b) results obtained by FCM; (c) results obtained by the proposed DOS-FCM.



(c)

Fig. 5. Partitioning of data 2 (a) original dataset; (b) results obtained by FCM; (c) results obtained by the proposed DOS-FCM.

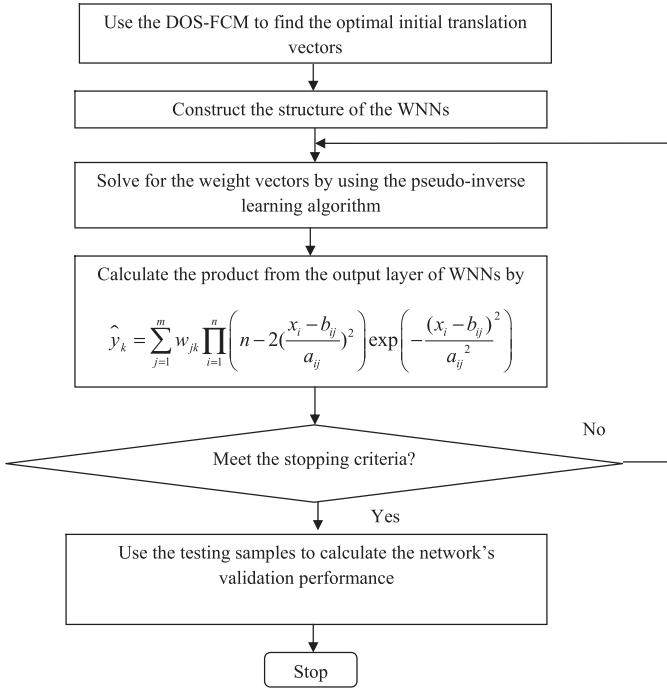


Fig. 6. The workflow of the combination of the DOS-FCM into the learning process of WNNs.

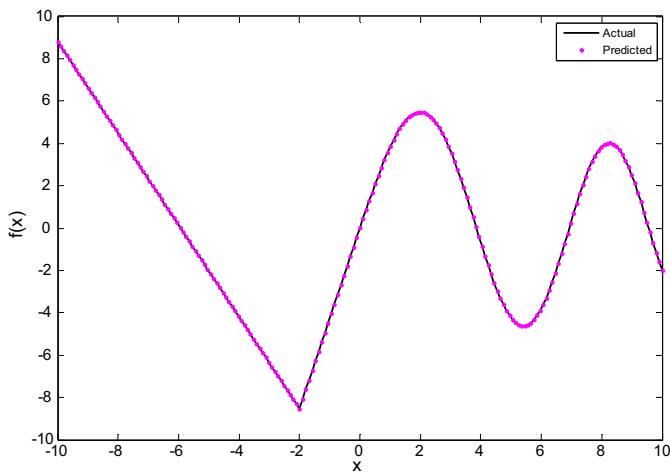


Fig. 7. Approximation of piecewise function using WNNs with DOS-FCM algorithm.

function well, even though the behavior of linearity, nonlinearity and an abrupt sharp spike were coexisting in this function.

Table 2 presents a comparison of performances of the proposed WNNs with DOS-FCM and other wavelet-based models in literature. It can be deduced that the proposed WNNs offered more satisfactory results than the rest. The lowest prediction error achieved by others was 0.0006 (in terms of ERR), where the result of 3.2005e-007 (in terms of ERR) obtained by the proposed WNNs surpassed that respective performance significantly. It is of interest to note that most of the listed models were focused on employing an enhanced learning algorithm or integrating the fuzzy rules with WNNs while using the recursive mathematical formulation [2] or the concept of multiresolution analysis [24–29] for networks initialization. Tzeng assigned the initial translation vectors based on GA [17], whilst Yao et al. integrated a self-tuning process with evolutionary programming to select the translation vectors [30]. It can thus be inferred that the quality of approximation for WNNs

Table 2

Performance comparison of the proposed WNNs with DOS-FCM algorithm with the results in the literature in approximating the benchmark piecewise function.

Model	Approximation accuracy (in terms of ERR)	Reference
WaveARX neural network	0.0480	[31]
Fuzzy WNNs	0.0210	[25]
Fuzzy WNNs	0.00228	[32]
Type-2 Fuzzy WNNs	0.0006	[27]
Fuzzy WNNs	0.0014	[28]
Fuzzy WNNs	0.0303	[33]
Fuzzy WNNs	0.022	[29]
Evolving WNNs	0.0300	[30]
WNNs	0.05057	[2]
WNNs with DOS-FCM	3.2005e-007	

was much improved by integrating the DOS-FCM initialization approach into WNNs learning strategy, though neither enhancement of learning algorithm nor the incorporation of fuzzy rules was taken into account in this present study.

5.2. Mackey–Glass chaotic time series

The Mackey–Glass equation was originally developed by Mackey and Glass [34] to simulate blood cell regulation, and was subsequently used extensively as a benchmark time series to evaluate the performance of ANNs. The Mackey–Glass equation was given by:

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x^{r(t-\tau)}} - bx(t), \quad t > 0. \quad (14)$$

The delay parameter, τ , characterizes the behavior of time series generated by this first-order differential-delay equation. It exhibits chaotic characteristics when $\tau \geq 17$ [35].

In this study, the parameters in Eq. (13) are specified as $a=0.2$, $b=0.1$, $r=10$, $\tau=17$, and $x(0)=1.2$, respectively, according to the published works [36–38]. The fourth-order Runge Kutta method was applied to obtain the numerical solution of this differential equation. The four input variables, $x(t-18)$, $x(t-12)$, $x(t-6)$ and $x(t)$ were used to forecast the single output variable, $x(t+6)$. A set of input–output data pairs with a length of 1000 were extracted, where the first 500 values served as training data and the remaining as testing data. To evaluate the competence of the proposed WNNs in forecasting the future value for the Mackey–Glass time series, mean squared error (MSE), root mean squared error (RMSE) and normalized mean squared error (NMSE) were utilized, which were defined as:

$$MSE = \frac{1}{n} \sum_{j=1}^n (f(x_j) - y_j)^2 \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (f(x_j) - y_j)^2} \quad (16)$$

$$NMSE = \left(\frac{\sum_{j=1}^n (f(x_j) - y_j)^2}{\sum_{j=1}^n (f(x_j) - \bar{y})^2} \right) \quad (17)$$

where $f(x_j)$, y_j and \bar{y} represent the actual value at point j , predicted value at point j , and average of the actual values respectively, and n was the number of testing data.

The approximation results for WNNs with different initialization approaches were summarized in Table 3. It can be observed

Table 3

Prediction accuracy of WNNs with different initialization approaches in approximating the Mackey–Glass time series.

Initialization Approach	MSE	RMSE	NMSE
Random	8.2786e–005	0.0091	0.0016
FCM	3.3161e–007	5.7586e–004	6.5109e–006
DOS-FCM	7.8794e–009	8.8766e–005	1.5473e–007

Table 4

Performance comparison of the proposed WNNs with DOS-FCM algorithm with the results in the literature in approximating the benchmark Mackey–Glass time series (500 testing samples).

Model	Forecasting Accuracy			Reference
	MSE	RMSE	NMSE	
Annealing robust radial basis function networks		0.0094		[39]
ANNs + autoregressive moving average model		0.0025		[40]
ANNs with residual analysis		0.0013		[41]
Ensembles of adaptive network fuzzy inference system		0.0173		[42]
Evolving fuzzy neural networks	0.006			[43]
Finite impulse response neural network + GA			0.1113	[44]
Interval type-2 fuzzy neural networks		0.0020		[45]
Locally liner neural fuzzy		9.61e–04		[46]
Multilayer perceptrons	4.28e–04		4.65e–04	[47]
Multiplicative neuron model	0.001			[48]
Radial basis function network		0.0014		[49]
WNNs with DOS-FCM	7.88e–09	8.88e–05	1.55e–07	

from this table that the network mapping constructed by all WNNs fitted the targeted output adequately, since the approximation errors were within the range of 8.2786e–005 to 7.8794e–009 (in terms of MSE). The conventional WNNs delivered poorer results which were mainly attributed to the initial randomness of the parameters configuration. The initial random selection of translation vectors might overlook the contributable features in the input space that play an important role in refining an optimal architecture for WNNs.

As predicted, the integration of the clustering algorithms in the learning phase of WNNs was accompanied by an improvement in the network generalization capability, which was reflected by a noticeable drop in the performance index. Almost 100% in reduction was attained compared to the conventional WNNs. As presented in Table 3, the employment of the proposed DOS-FCM in WNNs reached an extremely small value of MSE, i.e., a significant improvement of 99.99% from 8.2786e–005 (random initialization) and 97.62% from 3.3161e–007 (FCM) to 7.8794e–009 (DOS-FCM) were measured. The predictive performance of the WNNs with DOS-FCM was statistically different at the confidence level of 5%, as opposed to the random initialization ($p = 1.3431e-83$) and FCM initialization approaches ($p = 3.9841e-77$). Fig. 8(a) displays the forecasted Mackey–Glass time series of WNNs with DOS-FCM. It was evident that the chaotic and the oscillating behavior of the time series under investigation was successfully captured, which was further manifested by the small prediction errors in Fig. 8(b). The goodness-of-fit between the forecasted values and the generated time series values was portrayed in Fig. 8(c), from which it can be concluded that a practically precise forecasting was achieved, since all the data points were located on the diagonal line. The performance assessment of the proposed WNNs and other methods reported in literature related to the Mackey–Glass time series under the same criteria can be found in Table 4, from which the superiority and feasibility of the proposed method can be validated.

5.3. Real-world application: prediction of blood glucose level

Abrupt elevation in blood glucose level for diabetics is life-threatening as it may lead to dehydration, nausea, fatigue and eventually coma. Hence, monitoring of the blood glucose concentration needs to be done judiciously so that patients with diabetes

Table 5

The 19 input variables used for predicting the blood glucose level.

Glucose level	During interval
Short acting insulin	During interval
Food intake	During interval
Exercise	During interval
Glucose level	At the start of interval
Long acting insulin	During the past 24 h
Stress	During interval
Glucose level	At the start of interval on previous day
Short acting insulin	During interval on previous day
Food intake	During interval on previous day
Exercise	During interval on previous day
Resulting glucose	At the end of interval on previous day
Glucose level	At the start of previous interval
Short acting insulin	During previous interval
Food intake	During previous interval
Exercise	During previous interval
Exercise	Average over the past 24 h
Interval length	During interval
Exercise added up squared values	During the past 24 h

can adjust the dose of insulin injection accordingly. In this section, the practicability of the proposed WNNs in extracting the complex interaction of blood glucose metabolism will be explored.

The data used in this study was provided by Kok [50], in which it covered a 77-day study period. Eight fixed measurement points were assigned to night, before breakfast, after breakfast, before lunch, after lunch, before dinner, after dinner, and before sleeping. Based on these measurement points, four different intervals (morning, afternoon, evening and night) with three fixed points in each were formed. For example, the morning interval consisted of before breakfast, after breakfast and before lunch. At each data recording, the patient was required to fill up information such as time of blood glucose test, blood glucose concentration, dosage of short acting insulin injection, dosage of long acting insulin injection, food intake, stress level and exercise. The 19 input variables as listed in Table 5 were selected. Principal component analysis was applied to these variables subsequently in order to extract the characteristic features that signify the predominant trend in the data. By using the condition of percentage of variance, 3 principal components were chosen as the input features for WNNs in predicting the blood glucose level for morning interval. A RMSE criterion was

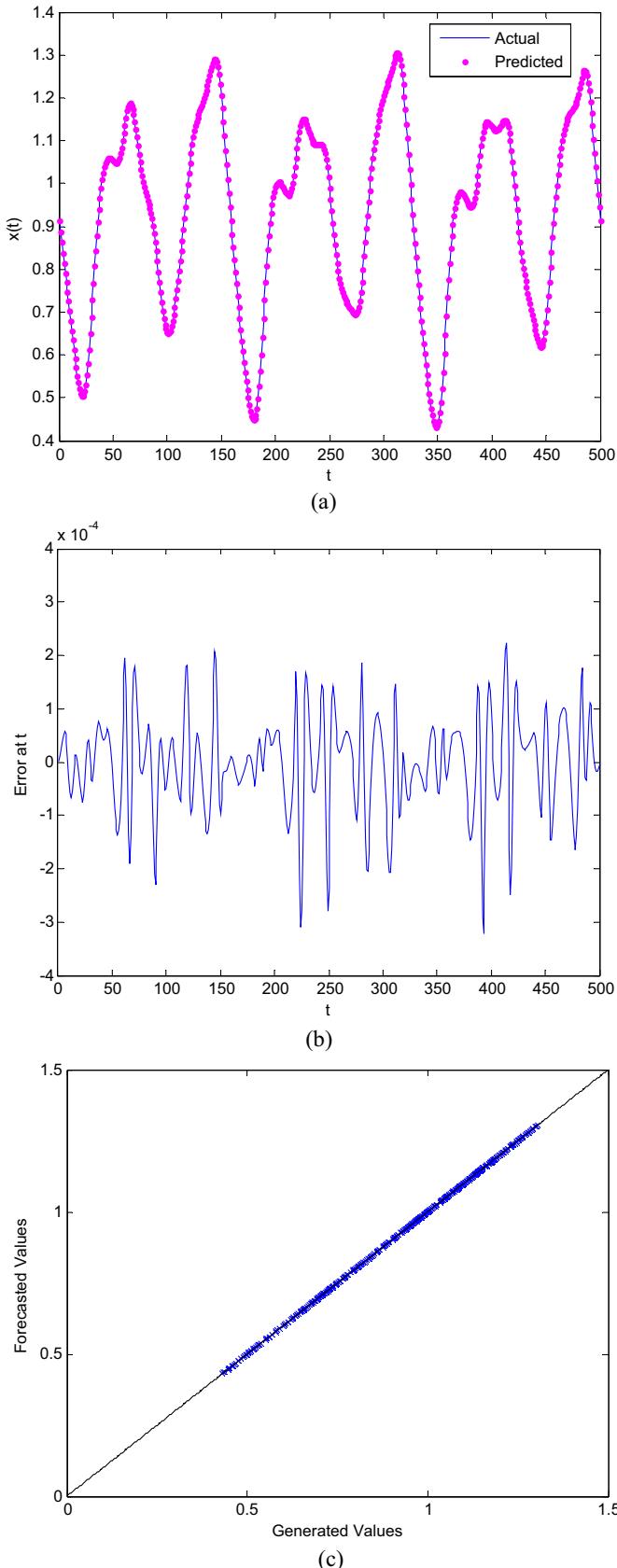


Fig. 8. (a) Approximation of Mackey–Glass time series using WNNs with DOS-FCM algorithm; (b) the prediction errors of Mackey–Glass time series forecasting results using WNNs with DOS-FCM; (c) The goodness-of-fit between the forecasted and the generated Mackey–Glass time series using WNNs with DOS-FCM.

Table 6

Prediction accuracy of WNNs with different initialization approaches in approximating the blood glucose level for diabetics.

Initialization approach	RMSE
Random	0.6465
FCM	0.2890
DOS-FCM	0.0360

used as performance evaluation. Further details of the material and methodology were omitted for brevity [38].

The experimental results for the WNNs in predicting the blood glucose level for diabetics at the end of the morning interval were presented in Table 6. These RMSEs were rather small (within the range of 0.6465–0.0360), which indicated a promising performance of all WNNs. In terms of the initialization approach of WNNs, the obtained results demonstrated that employing clustering algorithms to determine the number and location of translation vectors would lead to a decrease in RMSE. Occurring in a similar trend as previous sections, when classical FCM was compared against the random initialization approach, the former reduced the prediction error considerably, i.e., an improvement of 55.30% was attained. Using the proposed DOS-FCM algorithm in conjunction with WNNs led to the smallest value of RMSE wherein, it surpassed the WNNs with random initialization and FCM by a percentage improvement of 94.43% and 87.54%, respectively. The results were statistically significant at a confidence level of 5% ($p = 3.6061e-14$ and $p = 3.9410e-09$, for random and FCM initialization approaches, respectively). The forecasted blood glucose level at the end of the morning interval by WNNs with DOS-FCM algorithm was plotted in Fig. 9. It was observed that a relatively good prediction was obtained. WNNs with the proposed DOS-FCM algorithm were able to detect the chaotic characteristic in the blood glucose level time series, even though it was highly irregular and had a number of big leaps throughout the time series.

Considering the same input variables as listed in Table 5, the forecasting ability of the proposed WNNs was evaluated and compared with the obtained prediction accuracies by Kok [50], radial basis function networks (RBFNs) [51], probabilistic neural networks (PNNs) and recurrent neural networks (RNNs), as shown in Table 7. Instead of applying the principal component analysis and WNNs, a combination of input variables selection by heuristic approach and predictor based on multilayer perceptrons was developed by Kok [50]. For the blood glucose predictor based on RBFNs used in [51],

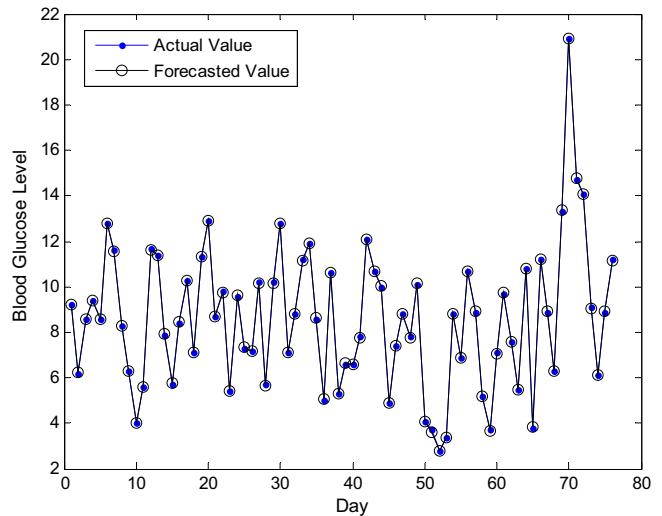


Fig. 9. Approximation of blood glucose level time series using WNNs with DOS-FCM algorithm.

Table 7

Performance comparison of the proposed WNNs with DOS-FCM algorithm with the results in the literature in predicting the blood glucose level for diabetics.

Model	Prediction accuracy (in terms of RMSE)	Reference
Multilayer perceptrons	1.8	[50]
Probabilistic neural networks	1.9305	
Radial basis function networks	0.0826	[51]
Recurrent neural networks	0.0585	
WNNs with DOS-FCM	0.0360	

instead of using random input selection, a pruning method was applied. The input variables of (1, 2, 5, 6, 7, 10, 15, 17, 18) (Table 5) were selected after the elimination process. The blood glucose prediction using the PNNs and RNNs were performed too, in which the same principal components chosen for WNNs were used as the input features.

The dismal performance of Kok's approach can be observed in Table 7, where RMSE of 1.8 was achieved. The unsatisfactory result was probably attributed to the input selection by trial and error, since random input selection might miss the complex interactions between the governing factors for blood glucose metabolism. Moreover, the nature of multilayer perceptrons with gradient-based learning algorithm which tends to get trapped in local minima may jeopardize its performance in a number of ways. On the other hand, WNNs were able to deal with this difficulty by adopting the pseudo-inverse algorithm in its learning strategy. The performance of the proposed WNNs was found to be more superior when its predicting accuracy was compared with the performance of RBFNs, PNNs and RNNs, as shown in Table 7. Its superiority could be due to essential attributes in the input space reflected through the selected optimal translation vectors by the DOS-FCM. The WNNs began the learning process from good initial points hence improving the generalization capability of the WNNs.

6. Conclusions

Initialization of the WNNs parameters needs to be handled precisely as the approximation solution may be drastically affected if improper initialization occurs. The pursuit of this goal was continued in this paper, where a novel algorithm, i.e., the DOS-FCM was proposed and adopted in the learning phase of WNNs. The usefulness of the DOS-FCM in searching for the appropriate features was first shown using two artificial datasets. Simulation demonstrated that the DOS-FCM outperformed the FCM in partitioning the observations into appropriate clusters, regardless of the geometric structures and overlapping characteristics.

Subsequently, its feasibility in the initialization of the WNNs parameters was evaluated and validated in the context of function approximation: (i) approximation of a benchmark piecewise function; (ii) approximation of a benchmark Mackey–Glass chaotic time series; and (iii) prediction of blood glucose concentration of diabetics in a real-world application problem. Simulation studies showed that the DOS-FCM possessed good selective ability. The improvements of 87% to 100% in terms of the performance index were observed in the considered cases compared to the WNNs with random and FCM initialization approaches. The DOS-FCM initialization method was able to select a set of best initial translation vectors from the input space, which eventually enhanced the validity of modeling of the WNNs. The WNNs with the DOS-FCM initialization approach also consistently outperformed the models reported in the literature.

Possible future work will be the utilization of the DOS-FCM in selecting the appropriate weight vectors during the learning process of the WNNs. Apart from that, it could also be interesting to investigate the application of the DOS-FCM in clustering the

stationary and non-stationary time series data [52–55] and anomalous data [56].

Acknowledgements

Financial supports from the Malaysian Government with the cooperation of Universiti Tun Hussein Onn Malaysia in the form of FRGS grant Vot 1490 and Universiti Sains Malaysia in the form of FRGS grant 203/PMATHS/6711323 are gratefully acknowledged.

References

- [1] S. Haykin, *Neural Networks A Comprehensive Foundation*, Prentice Hall PTR, 1998.
- [2] Q. Zhang, A. Benveniste, Wavelet networks, *IEEE Trans. Neural Netw.* 3 (1992) 889–898.
- [3] Z. Zainuddin, O. Pauline, Modified wavelet neural network in function approximation and its application in prediction of time-series pollution data, *Appl. Soft Comput.* 11 (2011) 4866–4874.
- [4] X. He, H. Guan, J. Qin, A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall, *J. Hydrol.* 527 (2015) 88–100.
- [5] Z. Wang, F. Lu, H.-d. He, Q.-C. Lu, D. Wang, Z.-R. Peng, Fine-scale estimation of carbon monoxide and fine particulate matter concentrations in proximity to a road intersection by using wavelet neural network with genetic algorithm, *Atmos. Environ.* 104 (2015) 264–272.
- [6] Y. Oussar, G. Dreyfus, Initialization by selection for wavelet network training, *Neurocomputing* 34 (2000) 131–143.
- [7] J. Cao, X. Lin, Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique, *Eng. Appl. Artif. Intell.* 21 (2008) 1255–1263.
- [8] J. Cao, Z. Lin, G.-b. Huang, Composite function wavelet neural networks with extreme learning machine, *Neurocomputing* 73 (2010) 1405–1416.
- [9] J.M. Gutiérrez, A. Gutés, F. Céspedes, M. del Valle, R. Muñoz, Wavelet neural networks to resolve the overlapping signal in the voltammetric determination of phenolic compounds, *Talanta* 76 (2008) 373–381.
- [10] B. Zhou, A. Shi, F. Cai, Y. Zhang, Wavelet neural networks for nonlinear time series analysis, in: F.-L. Yin, J. Wang, C. Guo (Eds.), *Advances in Neural Networks – ISNN 2004*, Springer, Berlin, Heidelberg, 2004, pp. 430–435.
- [11] M. Wei, S. Jin, L. Wang, Y. Zhou, Defect characteristic prediction of pipeline by means of wavelet neural network based on the hierarchical clustering algorithm, in: 2004 International Pipeline Conference, ASME, Alberta, Canada, 2004, pp. 921–924.
- [12] C.-J. Lin, Wavelet neural networks with a hybrid learning approach, *J. Inf. Sci. Eng.* 22 (2006) 1367–1387.
- [13] K. Hwang, S. Mandayam, S.S. Udpal, L. Udpal, W. Lord, M. Atzal, Characterization of gas pipeline inspection signals using wavelet basis function neural networks, *NDT & E Int.* 33 (2000) 531–545.
- [14] M.-C. Su, C.-H. Chou, A modified version of the K-means algorithm with a distance based on cluster symmetry, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2001) 674–680.
- [15] L. Feng, M.-H. Qiu, Y.-X. Wang, Q.-L. Xiang, Y.-F. Yang, K. Liu, A fast divisive clustering algorithm using an improved discrete particle swarm optimizer, *Pattern Recognit. Lett.* 31 (2010) 1216–1225.
- [16] M. Han, J. Yin, The hidden neurons selection of the wavelet networks using support vector machines and ridge regression, *Neurocomputing* 72 (2008) 471–479.
- [17] S.-T. Tzeng, Design of fuzzy wavelet neural networks using the GA approach for function approximation and system identification, *Fuzzy Sets Syst.* 161 (2010) 2585–2596.
- [18] C.-C. Hsu, T.-C. Lu, Minimum-phase criterion on sampling time for sampled-data interval systems using genetic algorithms, *Appl. Soft Comput.* 8 (2008) 1670–1679.
- [19] E.G. Ortiz-García, S. Salcedo-Sanz, Á.M. Pérez-Bellido, J.A. Portilla-Figueras, Improving the training time of support vector regression algorithms through novel hyper-parameters search space reductions, *Neurocomputing* 72 (2009) 3683–3691.
- [20] Z. Zainuddin, O. Pauline, An effective fuzzy C-means algorithm based on symmetry similarity approach, *Appl. Soft Comput.* 35 (2015) 433–448.
- [21] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, *Neurocomputing* 70 (2006) 489–501.
- [22] R. Coppi, P. D'Urso, Fuzzy K-means clustering models for triangular fuzzy time trajectories, *Stat. Methods Appl.* 11 (2002) 21–40.
- [23] E. Hartuv, A.O. Schmitt, J. Lange, S. Meier-Ewert, H. Lehrach, R. Shamir, An algorithm for clustering cDNA fingerprints, *Genomics* 66 (2000) 249–256.
- [24] J. Chen, D.D. Bruns, WaveARX neural network development for system identification using a systematic design synthesis, *Ind. Eng. Chem. Res.* 34 (1995) 4420–4435.
- [25] D.W.C. Ho, P.A. Zhang, J.H. Xu, Fuzzy wavelet networks for function learning, *IEEE Trans. Fuzzy Syst.* 9 (2001) 200–211.
- [26] E. Karatepe, M. Alci, A new approach to fuzzy wavelet system modeling, *Int. J. Approx. Reason.* 40 (2005) 302–322.

- [27] M. Singh, S. Srivastava, M. Hanmandlu, J.R.P. Gupta, Type-2 fuzzy wavelet networks (T2FWN) for system identification using fuzzy differential and Lyapunov stability algorithm, *Appl. Soft Comput.* 9 (2009) 977–989.
- [28] S. Srivastava, M. Singh, M. Hanmandlu, A.N. Jha, New fuzzy wavelet neural networks for system identification and control, *Appl. Soft Comput.* 6 (2005) 1–17.
- [29] J. Wang, J. Xiao, H. Peng, X. Gao, Constructing fuzzy wavelet network modeling, *Int. J. Inf. Technol.* 11 (2005) 68–74.
- [30] S. Yao, C.J. Wei, Z.Y. He, Evolving wavelet neural networks for function approximation, *Electron. Lett.* 32 (1996) 360.
- [31] J.H. Chen, D.D. Bruns, Waveearx neural-network development for system-identification using a systematic design synthesis, *Ind. Eng. Chem. Res.* 34 (1995) 4420–4435.
- [32] E. Karatepe, M. Alci, A new approach to fuzzy wavelet system modeling, *Int. J. Approx. Reason.* 40 (2005) 302–322.
- [33] S.T. Tseng, Design of fuzzy wavelet neural networks using the GA approach for function approximation and system identification, *Fuzzy Set Syst.* 161 (2010) 2585–2596.
- [34] M. Mackey, L. Glass, Oscillation and chaos in physiological control systems, *Science* 197 (1977) 287–289.
- [35] X. Wu, Y. Wang, Extended and Unscented Kalman filtering based feedforward neural networks for time series prediction, *Appl. Math. Model.* 36 (2012) 1123–1131.
- [36] J.R. Castro, O. Castillo, P. Melin, A. Rodríguez-Díaz, A hybrid learning algorithm for a class of interval type-2 fuzzy neural networks, *Inf. Sci.* 179 (2009) 2175–2193.
- [37] C.-J. Lin, Y.-J. Xu, A self-adaptive neural fuzzy network with group-based symbiotic evolution and its prediction applications, *Fuzzy Sets Syst.* 157 (2006) 1036–1056.
- [38] Z. Zainuddin, P. Ong, C. Ardin, A neural network approach in predicting the blood glucose Level for diabetic patients, *Int. J. Inf. Math. Sci.* 5 (2009) 72–79.
- [39] Y.-Y. Fu, C.-J. Wu, C.-N. Ko, J.-T. Jeng, L.-C. Lai, ARRBFDNs with SVR for prediction of chaotic time series with outliers, *Artif. Life Robot.* 14 (2009) 29–33.
- [40] I. Rojas, O. Valenzuela, F. Rojas, A. Guillen, L.J. Herrera, H. Pomares, L. Marquez, M. Pasadas, Soft-computing techniques and ARMA model for time series prediction, *Neurocomputing* 71 (2008) 519–537.
- [41] M. Ardalan-Farsa, S. Zolfaghari, Residual analysis and combination of embedding theorem and artificial intelligence in chaotic time series forecasting, *Appl. Artif. Intell.* 25 (2011) 45–73.
- [42] P. Melin, J. Soto, O. Castillo, J. Soria, A new approach for time series prediction using ensembles of ANFIS models, *Expert Syst. Appl.* 39 (2012) 3494–3506.
- [43] J.A.M. Hernandez, F.G. Castaneda, J.A.M. Cadena, An evolving fuzzy neural network based on the mapping of similarities, *IEEE Trans. Fuzzy Syst.* 17 (2009) 1379–1396.
- [44] H. Mirzaee, Linear combination rule in genetic algorithm for optimization of finite impulse response neural network to predict natural chaotic time series, *Chaos Soliton Fract.* 41 (2009) 2681–2689.
- [45] J.R. Castro, O. Castillo, P. Melin, A. Rodriguez-Díaz, A hybrid learning algorithm for a class of interval type-2 fuzzy neural networks, *Inf. Sci.* 179 (2009) 2175–2193.
- [46] A. Gholipour, B.N. Araabi, C. Lucas, Predicting chaotic time series using neural and neurofuzzy models: a comparative study, *Neural Process. Lett.* 24 (2006) 217–239.
- [47] H. Mirzaee, Long-term prediction of chaotic time series with multi-step prediction horizons by a neural network with Levenberg–Marquardt learning algorithm, *Chaos Soliton Fract.* 41 (2009) 1975–1979.
- [48] R.N. Yadav, P.K. Kalra, J. John, Time series prediction with single multiplicative neuron model, *Appl. Soft Comput.* 7 (2007) 1157–1163.
- [49] L.C. Chang, F.J. Chang, Y.P. Wang, Auto-configuring radial basis function networks for chaotic time series and flood forecasting, *Hydrol. Process.* 23 (2009) 2450–2459.
- [50] P. Kok, Prediction Blood Glucose Levels of Diabetics Using Artificial Neural Networks, Delft University of Technology, 2004.
- [51] G. Baghdadi, A.M. Nasrabadi, Controlling blood glucose levels in diabetics by neural network predictor, in: Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, 2007, pp. 3216–3219.
- [52] P. D'Urso, E.A. Maharaj, Wavelets-based clustering of multivariate time series, *Fuzzy Sets Syst.* 193 (2012) 33–61.
- [53] P. D'Urso, L. De Giovanni, E.A. Maharaj, R. Massari, Wavelet-based self-organizing maps for classifying multivariate time series, *J. Chemom.* 28 (2014) 28–51.
- [54] R. Coppi, P. D'Urso, Fuzzy unsupervised classification of multivariate time trajectories with the Shannon entropy regularization, *Comput. Stat. Data Anal.* 50 (2006) 1452–1477.
- [55] E. Ann Maharaj, P. D'Urso, D.A. Galagedera, Wavelet-based fuzzy clustering of time series, *J. Classif.* 27 (2010) 231–275.
- [56] P. D'Urso, L. De Giovanni, Robust clustering of imprecise data, *Chemom. Intell. Lab. Syst.* 136 (2014) 58–80.