Logic-based Fuzzy Neural Networks based on Fuzzy Granulation

Keun-Chang Kwak*, and Dong-Hwa Kim**

* Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada
(Tel : +1-780-437-3462; E-mail: kwak@ece.ualberta.ca)
**Department of Instrumentation and Control Engineering, Hanbat National University, Daejeon, Korea
(Tel : +82-42-000-0000; E-mail: kimdh@hanbat.ac.kr)

Abstract: This paper is concerned with a Logic-based Fuzzy Neural Networks (LFNN) with the aid of fuzzy granulation. As the underlying design tool guiding the development of the proposed LFNN, we concentrate on the context-based fuzzy clustering which builds information granules in the form of linguistic contexts as well as OR fuzzy neuron which is logic-driven processing unit realizing the composition operations of T-norm and S-norm. The design process comprises several main phases such as (a) defining context fuzzy sets in the output space, (b) completing context-based fuzzy clustering in each context, (c) aggregating OR fuzzy neuron into linguistic models, and (c) optimizing connections linking information granules and fuzzy neurons in the input and output spaces. The experimental examples are tested through two-dimensional nonlinear function. The obtained results reveal that the proposed model yields better performance in comparison with conventional linguistic model and other approaches.

Keywords: Logic-based fuzzy neural networks, context-based fuzzy clustering, fuzzy granulation, linguistic model, OR fuzzy neurons

1. INTRODUCTION

During the past few years, a considerable number of studies have been conducted on fuzzy model, together with a rapid growth in the variety of applications based on the concepts of fuzzy set theory, fuzzy if-then rules, and fuzzy reasoning. In general, three types of fuzzy models have been widely employed in various applications: linguistic fuzzy model [1], fuzzy relational model [2], and Takagi-Sugeno fuzzy model [3]. The main differences between these three fuzzy models lie in the consequents of their fuzzy rules, and their aggregation and defuzzification procedures differ accordingly. Among three fuzzy models, we concentrate on linguistic fuzzy model, also known as Mamdani fuzzy model [4] that constructs at the level of linguistic granules rather than given input-output data set. The Mamdani fuzzy model was proposed as the first attempt to control a steam engine and boiler combination by a set of linguistic control rules obtained from experienced human operators. The advantages of this model are more intuitive and well-suited to human input in comparison with other models. In this study, we follow the fundamental idea of novel linguistic models introduced by Pedrycz [5]. In contrast to numerically driven identification techniques, this model is based on the building meaningful linguistic granules in the space of experimental data and forming the ensuing model as a web of associations between such granules. The underlying algorithm used in the development of the models utilizes an augmented version of the clustering technique known as context-based or conditional fuzzy clustering that is centered through a notion of linguistic contexts. The effectiveness of this linguistic model has demonstrated on knowledge discovery and data mining [6]. On the basis of the concepts and characteristics of this linguistic model, the intent of this study is to develop the Logic-based Fuzzy Neural Networks (LFNN) based on a generic linguistic granules-oriented modeling technique and logic-driven processing unit realizing the T-S composition with aid of OR fuzzy neuron. In this setting of fuzzy neurons, the synergy of learning and transparency is well articulated [7].

Furthermore, the context-based fuzzy clustering technique helps reveal interesting behavior of the model with regard to the imposed granularity in the input and output space, in contrast to the context-free clustering frequently used for extracting fuzzy rules in the design of fuzzy model [8-10] and fuzzy neural networks [11-12]. Here the modal values in each context and connections linking information granules and fuzzy neurons in the input and output spaces are adjusted by gradient descent learning algorithm guided by the minimization of some performance index. The usefulness of the proposed model is discussed and demonstrated through two-dimensional nonlinear function.

2. GENERAL DESCRIPTION OF AND/OR FUZZY NEURONS

This section introduces the characteristics of two general types of fuzzy neurons (AND/OR) in connection with T-norm and S-norm. As the most commonly used logic processing elements, the AND and OR fuzzy neurons were introduced by Pedrycz [7]. These fuzzy neurons are achieved by fuzzy intersection and fuzzy union operators, which are usually referred to as T-norm and S-norm, respectively. The AND fuzzy neuron is a nonlinear logic processing element with n-input $\mathbf{x}\ [0,1]^n$ producing an output y governed by the following expression

$$y = \mathrm{AND}(\mathbf{x};\mathbf{w}) \tag{1}$$

where $\mathbf{w}$ denotes an n-dimensional vector of adjustable weighs. The AND neuron performs an S-norm operation on the inputs $\mathbf{x}$ and weights $\mathbf{w}$ and then a T-norm operation on the results of the S-norm operation as follows

$$y = \mathop{\mathrm{T}}_{i=1}^{n} \left(w_i \, s \, x_i \right) \tag{2}$$

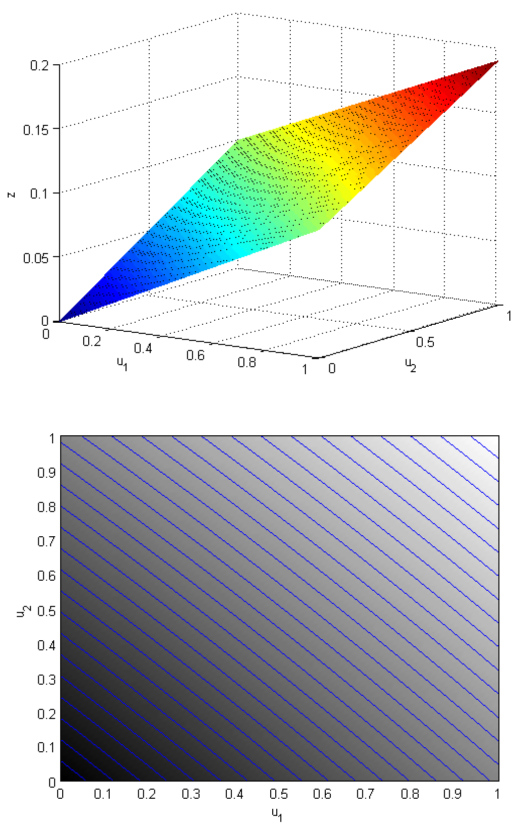
where s denotes S-norm operator. On the other hand, the OR fuzzy neuron is obtained by reverting the order of the T-norm and S-norm as follows

$$y = \mathrm{OR}(\mathbf{x};\mathbf{w}) \tag{3}$$

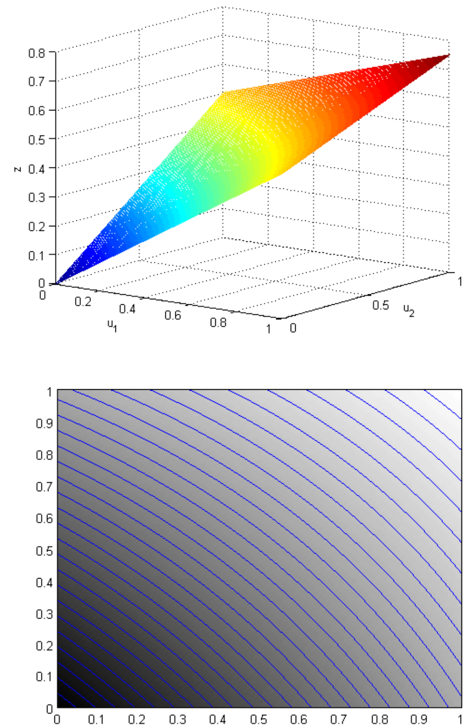
$$y = \underset{i=1}{\overset{n}{S}}\left(w_i\, t\, x_i\right) \tag{4}$$

where t denotes T-norm operator. The OR fuzzy neuron performs a T-norm operation on the inputs $\mathbf{x}$ and the weights $\mathbf{w}$ and then performs an S-norm operation on the results of the S-norm operator.

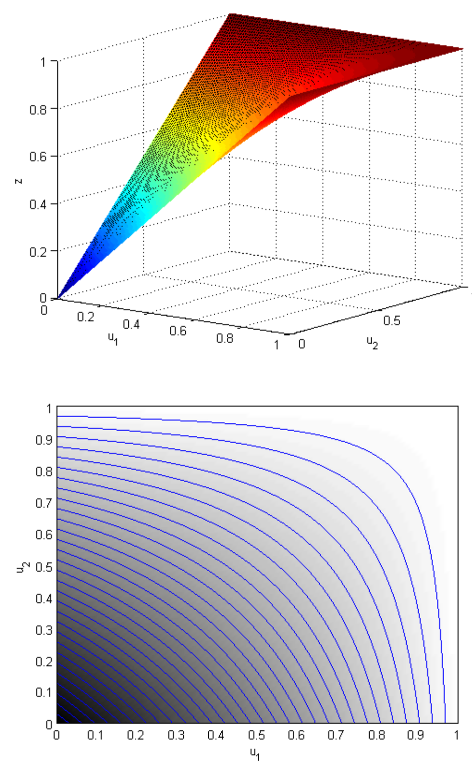
In general, four of the most frequently used T-norm operators are minimum, product, bounded product, and drastic product. Meanwhile, corresponding to the four T-norm operators, four S-norm operators are maximum, probabilistic sum, bounded sum, and drastic sum. In this paper, we shall concentrate on product as a T-norm operator and probabilistic sum as a S-norm. The performance of this composition will be demonstrated in comparison with conventional summation method in Section 5. Fig. 1 shows characteristic of the OR neuron for selected combinations of the connections.



(a) $\mathbf{v}$ = [0.1 0.1]



(b) $\mathbf{v}$ = [0.5 0.5]



(c) $\mathbf{v}$ = [1 1]

Fig. 1 Characteristics of the OR neurons

3. LINGUISTIC MODELS

In this section, we follow the fundamental idea of linguistic models introduced by Pedrycz [5] and describe the generic concepts and architecture of linguistic model with the use of fuzzy granulation realized via context-based fuzzy clustering.

3.1 The context-based fuzzy clustering

The purpose of context-based fuzzy clustering is to generate prototypes preserving homogeneity of the clustered patterns associated with their similarity in the input variables as well as in the output variable. This clustering technique is naturally geared towards direction aware clustering. For this context variable, we define a fuzzy set of context as follows

$$\mathrm{T}:\mathbf{Y}\rightarrow[0,1] \tag{5}$$

where $\mathbf{Y}$ is a universe of discourse of output variable. We assume that the values of context for the given data are available. The $f_k = T(y_k),\ k = 1,2,\ldots,N,$ represents a level of involvement of the k'th data in the assumed context of the output space. Here the value of f_k produced the membership degree between 0 and 1. For this reason, we can modify the requirements of the membership matrix as follows

$$\mathbf{U}(\mathbf{f}) = \left\{ u_{ik} \in [0,1] \,|\, \sum_{i=1}^{c} u_{ik} = f_k \ \forall k \ \text{and} \ 0 < \sum_{k=1}^{N} u_{ik} < N \ \forall i \right\} \tag{6}$$

We now convert the principle of the context-based clustering into the new partition matrix

$$u_{ik} = \frac{f_k}{\sum_{j=1}^{c} \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{\frac{2}{m-1}}} \quad i = 1,2,\ldots,c, \quad k = 1,2,\ldots,N \tag{7}$$

where the linguistic contexts to obtain f_k are generated through a series of triangular membership functions with equally spaced along the domain of an output variable and an 1/2 overlap between successive fuzzy sets. We arrive at the above formula by transforming a standard unconstrained optimization by making use of Lagrange multipliers and determining a critical point of the resulting function. The computations of the prototypes are the same as for the original Fuzzy C-Means (FCM) clustering algorithm. Moreover, the convergence conditions for the method are the same as thoroughly discussed for the original FCM clustering algorithm [13].

3.2 The architecture of linguistic models

The context-based fuzzy clustering has provided us with a backbone of the linguistic model as mentioned before. The development of the linguistic model comprises of two main phases. That is (a) forming fuzzy sets or relations of context, and (b) conditional clustering for the already available collection of contexts. Contexts invoke conditional fuzzy clustering and in the sequel produce fuzzy relations distributed in the input space. For each context, we end up with some number of clusters. Their activation levels are summed up at the first layer of the network; note that the number of summation nodes there is equal to the number of contexts. At this point we obtain activation levels of the individual contexts. The output layer is realized as a single granular neuron whose connections are just fuzzy sets of context as follows

$$\mathrm{Y} = \sum_{\oplus} w_t \otimes u_t \tag{8}$$

where $w_1, w_2, \ldots, w_p$ denote granular weights (connections) and "p" is the number of context. The symbols of generalized addition and multiplication ($\oplus, \otimes$) are used here to emphasize a granular character of the arguments being used in this aggregation.

4. THE PROPOSED LOGIC-BASED FUZZY NEURAL NETWORKS (LFNN)

In what follows, we concentrate on the design of the LFNN including learning scheme and logic-driven processing unit based on fuzzy neuron. Although the linguistic model has a structured knowledge representation in the form of fuzzy if-then rules, it lacks the adaptability to deal with changing external environment. Thus, we incorporate logic-based neural network learning concepts in the design of the linguistic model. From the integration of these two complementary approaches, we can construct a solid computing framework as to the underlying ideas of fuzzy granulation and their role in the construction of the proposed LFNN. Fig. 2 shows the architecture of the proposed model. As shown in Fig. 2, the proposed model is composed of four layers. The characteristics of each layer are as follows

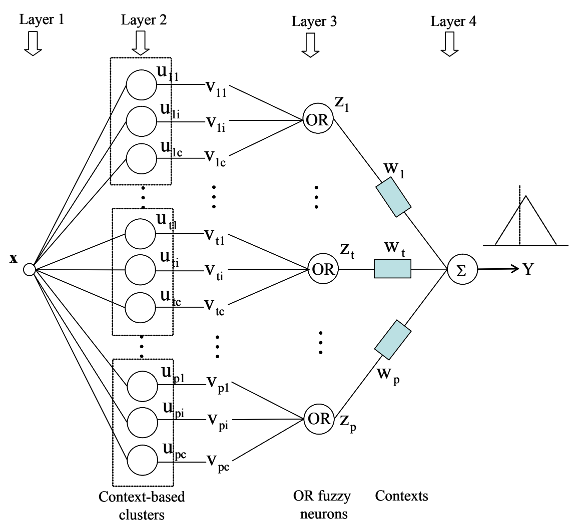


Fig. 2 The architecture of the proposed LFNN

[Layer 1] In this layer, each of the input variables is normalized to within the unit interval [0,1] since the clustering algorithms are sensitive to the ranges of elements in the input vectors.

[Layer 2] Every node in this layer is a cluster (rule) generated through each context. For simplicity, we assume that the number of cluster "c" in each context is equal. Here, the activation level u_{ti} denotes i'th cluster obtained from t'th context.

[Layer 3] Every node in this layer consists of OR fuzzy neurons. We use product T-norm and probabilistic sum S-norm as mentioned in Section 2

$$z_t = \left[S_{i=1}^{c} \left(u_{ti} \, T \, v_{ti} \right) \right] \quad (9)$$

where z_t denotes t'th output of OR fuzzy neuron. We note that this neuron carries on some *and*-wise aggregation of the activation level followed by the global *or*-wise combination of these partial results. Here, the hidden weights v_{ti} between the activation level and those neurons are connected by performing OR fuzzy neurons and their initial values are one. These weights are adjusted by gradient descent method in the learning stage.

[Layer 4] The single node in this layer computes the overall output as linear combination of the output in the layer 3 and modal value in each context as the following form

$$Y = \sum_{t=1}^{p} z_t w_t \quad (10)$$

While the activation levels and fuzzy neuron's output are numerical, the connections w_t are linguistic values and they are just the contexts used for the previous clustering purposes. These connections are updated by gradient descent in the normalized form. Depending on specific realization, these connections can realized as intervals, fuzzy sets, shadowed sets, rough sets, and alike. As a result of processing, the output Y is also a granular whose granular character is inherently associated with the nature of the connections of the neuron. The resulting granular input Y is completely characterized by its lower and upper bounds as follows

- the lower bound $Y_- = \sum_{t=1}^{p} z_t w_{t-}$
- the upper bound $Y_+ = \sum_{t=1}^{p} z_t w_{t+}$

where w_{t-} and w_{t+} are values that determines break points in the t'th context. In general, the range of the modal values in each context depends on the output of target system to be modeled. Thus, we use the normalized modal value of each context to facilitate learning as follows

$$\overline{w}_t = \frac{w_t - \min(\mathbf{w})}{\max(\mathbf{w}) - \min(\mathbf{w})} \quad (11)$$

where $\overline{w}_0 = 0$, $\overline{w}_p = 1$, and $\mathbf{w} = [w_1, w_2, \ldots, w_p]$. Here $\overline{w}_0$ and $\overline{w}_p$ are fixed regardless of learning as shown in Fig. 3. This figure shows the example of linguistic contexts with the normalized value when the number of context "p" is 5. When the model output is computed, the normalized values are transformed into the modal value of each context in the output space to compute the overall output.

On the other hand, the learning scheme is performed with a gradient descent algorithm frequently used in the design of neural networks. Here the error measure is computed as follows

$$E = [M(Y) - 1]^2 \quad (12)$$

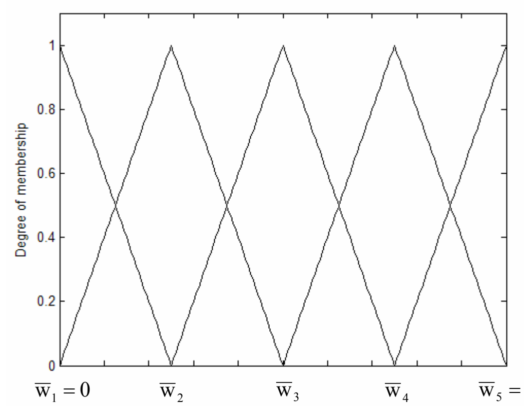


Fig. 3 Normalized modal value in each context (p=5)

where $M(Y)$ is fuzzy number of the model output obtained by the proposed LFNN. This error measure is focused on the maximization of the average agreement of the linguistic model with experimental numeric data. The difference occurring in the above expression, that is $M(Y) - 1$ expresses a departure from an "ideal" matching being equal to 1.0. Naturally, the error measure E should be optimized so that we achieve the highest level of matching possible. The gradient vector is defined as the derivative of the error measure with respect to each parameter, so we have to apply the chain rule again to find the gradient vector as follows

$$\Delta \overline{w}_t = -\eta_{\overline{w}} \frac{\partial E}{\partial \overline{w}_t} = -\eta_{\overline{w}} \frac{\partial E}{\partial Y} \frac{\partial Y}{\partial \overline{w}_t} \quad (13)$$

$$\frac{\partial E}{\partial Y} = 2(M(Y) - 1), \quad \frac{\partial Y}{\partial \overline{w}_t} = z_t \quad (14)$$

where $\eta_{\overline{w}}$ is a learning rate to train $\overline{w}_t$. Accordingly, for gradient descent, the update formula for the normalized modal value of context $\overline{w}_t$ is as follows

$$\Delta \overline{w}_t = -2\eta_{\overline{w}} (M(Y) - 1) z_t \quad (15)$$

In the case of the weight v_{ti}, we use the chain rule to find gradient vector as follows

$$\Delta v_{ti} = -\eta \frac{\partial E}{\partial v_{ti}} = -\eta \frac{\partial E}{\partial Y} \frac{\partial Y}{\partial z_t} \frac{\partial z_t}{\partial v_{ti}} \quad (16)$$

$$\frac{\partial Y}{\partial z_t} = \overline{w}_t \quad (17)$$

where η_v is a learning rate to train the hidden weight v_{ti}. If we substitute the symbol A_t for the norms of the terms not involving a weight $i = q$, then we can solve for the partial derivative of z_t with respect to that weight v_{tq} as follows

$$A_t = S_{i \neq q}^{c} u_{tq} \, T \, v_{tq} \quad (18)$$

$$\frac{\partial z_t}{\partial v_{tq}} = \frac{\partial\left[A + v_{tq}u_{tq} - Av_{tq}u_{tq}\right]}{\partial v_{tq}}$$

$$= (1 - A_t)u_{tq} = (1 - \mathrm{probOR}_{i \neq q}^{c}\left[u_{ti}v_{ti}\right])\ u_{tq} \quad (19)$$

The final update rule for the weight v_{ti} is as follows

$$\Delta v_{tq} = -2\eta_v (M(Y) - 1)\overline{w}_t (1 - A_t)u_{tq} \quad (20)$$

For more details on learning scheme including OR fuzzy neuron, see [14].

Fig. 4 shows flowchart concerning the procedure of the proposed LFNN. In this procedure, we assume that the number of context "p" is fixed. In the step of performance comparison, we use the test error as a true measure of the model's performance [15]. Here the best model we can achieve occurs when the test error is minimal. Although the training error is decreased as the number of iteration "i" increases, it will degrade the performance of the proposed model on unforeseen inputs. Thus, the resultant model is not biased toward the training data set and it is likely to have a better generalization capacity to new data.

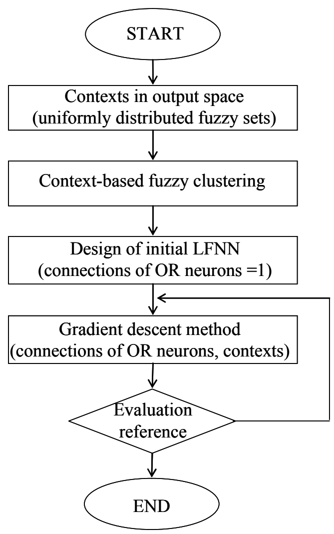


Fig. 4 Flowchart of the proposed LFNN procedure

5. EXPERIMENTS

In this section, the effectiveness of the proposed LFNN is tested through two-dimensional synthetic data. We also report on the comprehensive set of experiments and draws conclusions as to the performance of the proposed method. We consider two-dimensional nonlinear function of two variables given by [5]

$$y = f(x_1, x_2) = 0.6 + 2x_1 + 4x_2 + 0.5x_1x_2 + 25\sin(0.5x_1x_2)$$

This function is defined in the Cartesian product of two input range **X** = [-4,6]×[-2,4] of the preceding equation. The three-dimensional plot of this nonlinear relationship is visualized in Fig. 5.

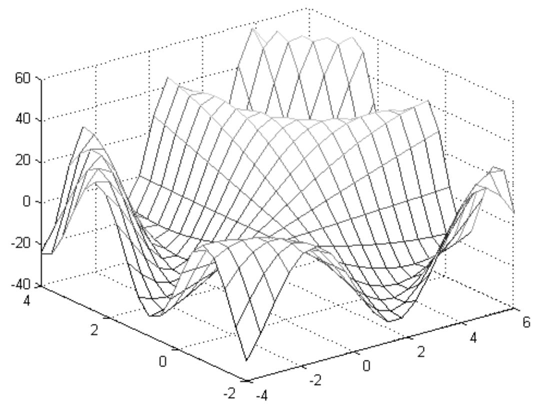


Fig. 5 3-D plot of the nonlinear relationship

Each training (60%) and test data (40%) set were generated randomly from the already defined universe of discourse **X**. The training and test data set consist of 100 input-output data pairs, respectively. The training data set is used for model construction, while the test set is used for model validation. The experiment is repeated for ten times. Fig. 6 and 7 show the approximation and generalization capability of the model for training and test data when the "p" and "c" are six and five, respectively. In these figures, the boundary lines are represented as fuzzy number of the predicted output. The proposed LFNN method outperformed conventional linguistic model although the "p" and "c" increase. Here the best model we can achieve occurred in the case of "p=6" and "c=5" which the test error is minimal.

Here convectional linguistic model is performed by simple summation. Table 1 lists the RMSE of mean and standard deviation for comparative analysis. The experimental results reveal that the proposed LFNN designed by logic-based processing unit and parameter optimization leads to the improved performance in comparison with other approaches.

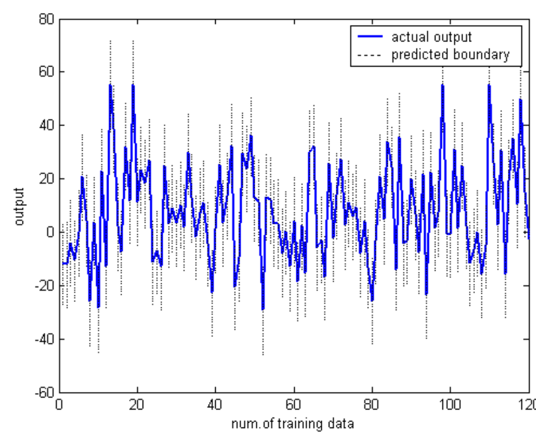


Fig. 6 Approximation capability for training data set (p=6, c=5)

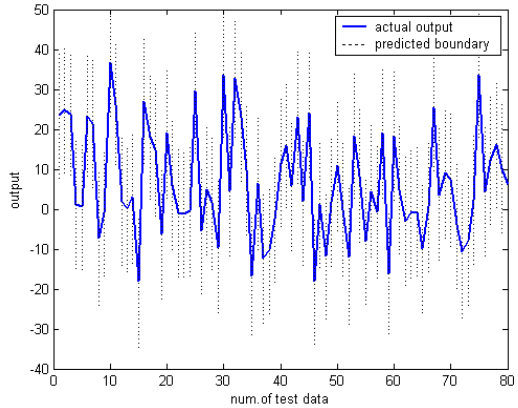


Fig. 7 Generalization capability for test data set (p=6,c=5)

Table 1 RMSE of mean and standard deviation for the proposed model (p=6, c=5)

		No. of hidden nodes	Training data	Test data
RBFNN		30	13.78±0.62	15.13±0.97
Linguistic model [5]		30	10.90±0.64	13.28±1.21
The proposed method	without optimization	30	10.40±0.53	13.03±1.07
	with optimization	30	10.23±0.59	12.78±0.93

6. CONCLUSIONS

We have developed the LFNN based on a generic linguistic granules-oriented modeling technique and logic-driven processing unit with optimization realizing the T-S composition with aid of OR fuzzy neuron. The experimental results revealed that the proposed LFNN yields better approximation and generalization capability in comparison with linguistic models based on fuzzy granulation. Consequently, the proposed model is supposed to possess humanlike expertise within a specific domain, adapt itself and learn to do better in nonlinear environments, Furthermore, this model can be represented the fundamental concept of information granules regarded as semantically meaningful conceptual entities that are crucial to the overall framework of user-centric modeling.

ACKNOWLEDGMENTS

This work was also supported by EESRI (R-2003-0-285).

REFERENCES

[1] L.A. Zadeh, "Fuzzy sets," *Information and Control*, Vol. 8, pp. 338-353, 1965

[2] W. Pedrycz, "An identification algorithm in fuzzy relational system," *Fuzzy Sets and Systems*, Vol. 13, pp.153-167, 1984.

[3] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. on Systems, Mans, and Cybernetics,* Vol. 15, pp. 116-132, 1985.

[4] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. Journal of Man-Machine Studies*, Vol. 7, No. 1, pp.1-13, 1975.

[5] W. Pedrycz and A. V. Vasilakos, "Linguistic models and linguistic modeling," *IEEE Trans. on Systems, Man, and Cybernetics-Part C*, Vol. 29, No. 6, pp.745-757, 1999.

[6] W. Pedrycz, "Fuzzy set technology in knowledge discovery," *Fuzzy Sets and Systems*, Vol. 98, pp. 279-290, 1998.

[7] W. Pedrycz, "Logic-based neurons: extensions, uncertainty representation and development of fuzzy controllers," *Fuzzy Sets and Systems*, Vol. 66, pp. 251-266, 1994.

[8] E. Kim, M. Park, S. Ji, and M. Park, "A new approach to fuzzy modeling," *IEEE Trans. on Fuzzy Systems*, Vol. 5, No. 3, pp.328-337, 1997.

[9] J. Abonyi, R. Babuska, and F. Szeifert, "Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models", *IEEE Trans. on Systems, Man, and Cybernetics-Part B*, Vol. 32, No. 5, pp.612-621, 2002.

[10] S. L. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of Intelligent and Fuzzy Systems*, Vol. 2, No. 3, pp.267-278, 1994.

[11] S. K. Oh, W. Pedrycz, and H. S. Park, "Hybrid identification in fuzzy-neural networks," *Fuzzy Sets and Systems*, Vol. 138, No. 2, pp.399-426, 2003.

[12] K. C. Kwak, M. G. Chun, J. W. Ryu, and T. H. Han, "FCM-based adaptive fuzzy inference system for coagulant dosing process in a water purification plant", *Int. Journal of Knowledge-Based Intelligent Engineering Systems*, Vol. 4, No. 4, pp.230-236, 2000.

[13] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[14] L. H. Tsoukalas, R. E. Uhrig, and J. W. Hines, *Fuzzy and Neural Approaches in Engineering*, Wiley, John and Sons, 1996.

[15] J. S. R. Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference Systems," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 23, No. 3, pp. 665-685, 1993.