Improving the Interpretability of Takagi-Sugeno Fuzzy Model by Using Linguistic Modifiers and a Multiple Objective Learning Scheme

Shang-Ming Zhou
Department of Computer Science
University of Essex
Colchester CO4 3SQ, United Kingdom
Email: szhoup@essex.ac.uk

John Q. Gan
Department of Computer Science
University of Essex
Colchester CO4 3SQ, United Kingdom
Email: jqgan@essex.ac.uk

***Abstract*–In this paper we present a new Tagaki-Sugeno (TS) type model whose membership functions (MFs) are characterized by linguistic modifiers. As a result, during adaptation, the trained local models tend to become the tangents of the global model, leading to good model interpretability. In order to prevent the global approximation ability from being degraded, an index of fuzziness is proposed to evaluate linguistic modification for MFs with adjustable crossover points. A new learning scheme is also developed, which uses the combination of global approximation error and the fuzziness index as its objective function. By minimizing the multiple objective performance measure, a tradeoff between the global approximation and local model interpretation can be achieved. Experimental results show that by the proposed method good interpretation of local models and transparency of input space partitioning can be obtained for the TS model while at the same time the global approximation ability is still preserved.**

I. INTRODUCTION

In data-driven fuzzy modeling, interpretation preservation during adaptation can be regarded as one of the most important issues [1] [2] [3] [4] [5]. The first aspect of interpretability of fuzzy models is about the transparency of partitioning of input space, or generation of interpretable fuzzy sets. Although there exists no unified standard for selecting MFs during adaptation, in the interests of preserving or enhancing the interpretability, some researchers have suggested some semantic criteria or heuristic criteria to guide the generation of MFs. de Oliveira proposed several semantic criteria for designing MFs, such as distinguishability of MFs, normalization of MFs, moderate number of linguistic terms per variable, natural zero positioning, and coverage of the universe of discourse etc. [6], which have been proved to be reasonable [4] [7]. On the other hand, in fuzzy cluster analysis, one of the main schemes to partition input space, as the criterion for "optimal partition" of a data set, the basic heuristic that "good" clusters are actually not very fuzzy could be acceptable [8]. Although fuzzy algorithms are used in data clustering, the aim of the clustering is to generate a "harder" partitioning of the data set [9]. In other words, the produced partitioning should achieve better interpretation. A requirement directly related to this interpretability is that MFs should be less overlapped among adjacent fuzzy sets and have large core regions, *i.e.*, large support sets with membership degree of 1. However classical rule induction algorithms, such as neuro-fuzzy algorithms [10], generate fuzzy sets with "too much" or "absolutely no" overlap due to their accuracy-oriented nature.

Currently, most of the efforts to improve the interpretability of fuzzy model are focused on the interpretation of partitioning of input space. As a matter of fact, for some special fuzzy models, such as Takagi-Sugeno (TS) fuzzy model [11], in addition to the interpretation of partitioning of input space, there exists another type of interpretation that needs to be further decrypted, *i.e.*, the interaction between the global model and local models. These local models often exhibit some eccentric behaviors that are hard to be interpreted as shown in Fig.1, thus such a model could not be interpreted in terms of individual rules in case there is no priori knowledge available.

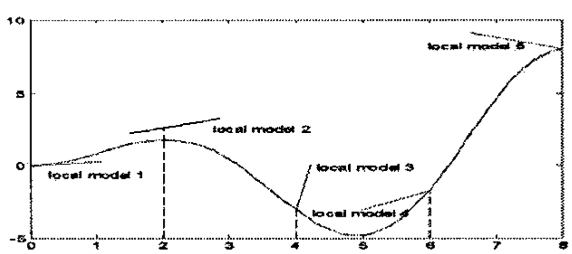

Fig.1 TS model with uninterpretable local models

In case there is no priori knowledge available for local models in a TS system, one may accept the basic heuristic that if the local models match the global model well, then the local models are considered to possess good interpretability, and the best interpretation for the local models could be that they become the tangents of the global model. In such a way, these local linear models possess the abilities to reflect the local properties of the system to be modeled, which is crucial for the success of local linear models' applications to nonlinear state estimation, fusion and predictive control[12][13][14]. Unfortunately, most of the MFs currently used in fuzzy models, such as Gaussian functions, could not achieve this aim, which has received a few researchers' attention [3][15][16].

In this paper, a linguistic modifier proposed in [17] is used to update the shape of MFs during the adaptation. As the MFs become less overlapped and possess larger core regions along with updating, the desired situation would emerge more possibly: there is only one rule applicable or dominate at a time, and the consequents are forced to represent the local behaviors of the system. Thus the eccentric behaviors of local models would be remedied greatly, and the final trained local models could become the tangents of the global model. In this paper, in order to measure the degree of linguistic modification, as an extension of the fuzziness measure proposed by Yager [18], we propose an index of fuzziness to evaluate the performance of linguistic modification for MFs with adjustable crossover points. A tradeoff between global approximation and local model interpretation can be achieved by minimizing a performance measure that combines the global error measure and the proposed index of fuzziness.

II. THE TS MODEL USING LINGUISTIC MODIFIERS AS FUZZY MFS

In this paper, TS model with the following rules will be addressed:

$$R_i: \text{if } x_1 \text{ is } A_{i_1}^{(1)} \text{ and } \dots \text{ and } x_n \text{ is } A_{i_n}^{(n)} \\ \text{then } y_i = a_{i0} + a_{i1}x_1 + \cdots + a_{in}x_n \tag{1}$$

where x_j are input variables, y_i is the output variable of the ith local model, $A_{i_j}^{(j)}$ are fuzzy sets about x_j, a_{ij} are the consequent parameters that have to be identified in terms of given data sets, R_i is the ith rule of the TS system, and $1 \le i_1 \le L_1$, ..., $1 \le i_n \le L_n$, $1 \le i \le L = \prod_{j=1}^{n} L_j$, with L_j being the number of fuzzy sets about x_j. The global output of the system is calculated by

$$y = \sum_{i=1}^{L} w_i y_i \tag{2}$$

where w_i is the normalized firing strength of rule R_i:

$$w_i = \tau_i / \sum_{i=1}^{L} \tau_i \tag{3}$$

and τ_i is called the firing strength of rule R_i, which is computed as $\tau_i = \prod_{j=1}^{n} A_{i_j}^{(j)}(x_j)$.

It can be seen that in this TS model, given the fuzzy sets about every variable on its domain of discourse, the rule base includes all the possible combinations of these fuzzy sets to cover the whole input space. For the sake of representing the rules clearly, we sort the rules as follows: corresponding to a combination of premise fuzzy sets $A_{i_1}^{(1)}$, ... , $A_{i_n}^{(n)}$, the rule is indexed as i in the rule base, where $i = \sum_{j=1}^{n-1} [(i_j - 1) \cdot \prod_{q=j+1}^{n} L_q] + i_n$.

In this paper, the linguistic modifiers proposed in [17] are used as the MFs for the above TS model. The initial fuzzy sets of these linguistic modifiers are chosen to be triangular functions, so the MFs of fuzzy sets $A_{i_j}^{(j)}$ are obtained as follows:

$$A_{i_j}^{(j)}(x_j; \beta_{i_j-1,j}, C_{i_j,j}^{(1)}, \beta_{i_j,j}, C_{i_j,j}^{(2)}, \beta_{i_j+1,j}, p_j)$$
$$= \begin{cases} \dfrac{1}{\mu_{C_{i_j,j}^{(1)}}^{p_j-1}} \left(\dfrac{x_j - \beta_{i_j-1,j}}{\beta_{i_j,j} - \beta_{i_j-1,j}} \right)^{p_j}, & \beta_{i_j-1,j} \le x_j < C_{i_j,j}^{(1)} \\ 1 - \dfrac{1}{\left(1 - \mu_{C_{i_j,j}^{(1)}}\right)^{p_j-1}} \left(\dfrac{\beta_{i_j,j} - x_j}{\beta_{i_j,j} - \beta_{i_j-1,j}} \right)^{p_j}, & C_{i_j,j}^{(1)} \le x_j < \beta_{i_j,j} \\ 1 - \dfrac{1}{\left(1 - \mu_{C_{i_j,j}^{(2)}}\right)^{p_j-1}} \left(\dfrac{x_j - \beta_{i_j,j}}{\beta_{i_j+1,j} - \beta_{i_j,j}} \right)^{p_j}, & \beta_{i_j,j} \le x_j < C_{i_j,j}^{(2)} \\ \dfrac{1}{\mu_{C_{i_j,j}^{(2)}}^{p_j-1}} \left(\dfrac{\beta_{i_j+1,j} - x_j}{\beta_{i_j+1,j} - \beta_{i_j,j}} \right)^{p_j}, & C_{i_j,j}^{(2)} \le x_j < \beta_{i_j+1,j} \end{cases} \tag{4}$$

where $C_{i_j,j}^{(1)}$ and $C_{i_j,j}^{(2)}$ are the left and right crossover points of $A_{i_j}^{(j)}$ respectively, $\mu_{C_{i_j,j}^{(1)}}$ and $\mu_{C_{i_j,j}^{(2)}}$ are evaluated as follows:

$$\mu_{C_{i_j,j}^{(1)}} = \frac{C_{i_j,j}^{(1)} - \beta_{i_j-1,j}}{\beta_{i_j,j} - \beta_{i_j-1,j}}; \quad \mu_{C_{i_j,j}^{(2)}} = \frac{\beta_{i_j+1,j} - C_{i_j,j}^{(2)}}{\beta_{i_j+1,j} - \beta_{i_j,j}} \tag{5}$$

It can be proved [19] that as the linguistic modifier parameters p_j increase, the ε-insensitive cores of fuzzy sets will become bigger and bigger, and at the same time the overlapping among adjacent fuzzy sets will become smaller and smaller. In such a way, the local models will be forced to dominate the local behaviors of the system, and tend to become the tangents of the global model. As a result, improvements could be made in not only the interpretation of local models but also the transparency of partitioning of input space. However, based on MFs with less overlapping and larger core regions, the global approximation ability of the TS model could be degraded. In the following section we propose a scheme to make the linguistic modifiers optimally adjusted so that the accuracy and interpretability of the model can be balanced in terms of a multiple objective performance measure.

III. A LEARNING ALGORITHM BASED ON A MULTIPLE OBJECTIVE FUNCTION

A. Fuzziness measure of a fuzzy set

The proposed fuzziness measure is based on the distance between a fuzzy set A and an ordinary (crisp) set $\underline{A}$ near to A. $\underline{A}$ is defined as follows:

$$\underline{A}(x)=\begin{cases}1 & \text{if } c^{(1)}\le x\le c^{(2)}\\ 0 & \text{otherwise}\end{cases} \tag{6}$$

where $c^{(1)}$ and $c^{(2)}$ are the left and right crossover points of fuzzy set A respectively. Given a data set $\{x(k)\}_{k=1}^{N}$ on domain x, the index of fuzziness, $F(A)$, is defined based on the distance between A and $\underline{A}$ as follows:

$$F(A)=\frac{2}{N^{1/r}}d_r(A,\underline{A}) \tag{7}$$

where N is the length of the data set, and r is the order of the distance d_r between A and $\underline{A}$. Obviously, in case $c^{(1)}=c^{(2)}=c$ and $A(c)$=0.5, $F(A)$ becomes the classic fuzziness measure proposed by Yager [18]. Particularly, in case r=2 (Euclidean distance is used), (8) defines a *quadratic index of fuzziness*,

$$F_q(A)=\frac{2}{\sqrt{N}}\sqrt{\sum_{k=1}^{N}\left(A(x(k))-\underline{A}(x(k))\right)^2} \tag{8}$$

In this paper, the *quadratic index of fuzziness* is used in the proposed learning algorithm.

B. A multiple objective function and the learning algorithm

For a given data set $\{(x(k),d(k))\}_{k=1}^{N}$, a hybrid learning scheme is employed to update the consequent parameters a_{ij} and the premise parameters p_j. In the first pass of the algorithm, the premise parameters p_j are fixed (starting with fixed value 1), and the consequent parameters a_{ij} are identified by least squares estimates in terms of the global accuracy measure. In the second pass, the newly obtained consequent parameters a_{ij} are fixed, and the premise parameters p_j are updated by a gradient descent algorithm in terms of a multiple objective performance measure as defined below:

$$J=\varphi\cdot E+\theta\cdot F \tag{9}$$

where φ and θ are two positive constants satisfying the condition: $\varphi+\theta=1$, E is the global accuracy measure:

$$E=\frac{1}{2}\sum_{k=1}^{N}\|d(k)-y(k)\|^2 \tag{10}$$

and F is the index of fuzziness of the TS model defined as

$$F=\sum_{i_1=1}^{L_1}\cdots\sum_{i_n=1}^{L_n}\sum_{j=1}^{n}F(A_{i_j}^{(j)}) \tag{11}$$

$F(A_{i_j}^{(j)})$ is the *quadratic index of fuzziness* defined by (8).

1) Least-squares estimates for consequent parameters

In order to identify the consequent parameters in the TS model, we reformulate some expressions in (1)-(3). Defining a base matrix M as follows:

$$M=\begin{bmatrix}M_1^T(1) & \cdots & M_L^T(1)\\ \\ M_1^T(N) & & M_L^T(N)\end{bmatrix}_{N\times L(n+1)} \tag{12}$$

where $M_i^T=(w_i\ w_ix_1\ \cdots\ w_ix_n)$, and representing the consequent parameters by a column vector $a=(a_{10}\ a_{11}\cdots a_{1n}\ a_{20}\ a_{21}\cdots a_{2n}\cdots\cdots a_{L0}\ a_{L1}\cdots a_{Ln})^T$, we reformulate the TS model as follows:

$$M\cdot a=d \tag{13}$$

where $d=(d(1)\cdots d(N))^T$. Because the consequent parameters in a do not make any contribution to the index of fuzziness of the TS model, they can be identified practically based on the global approximation accuracy measure E defined in (10). Since the number of training data pairs is usually greater than $L \times (n+1)$, this is a typical ill-posed problem and generally there does not exist exact solution for vector a. The least-squares estimate of a can be obtained by $a^* = M^+ d$, where M^+ is the Moore-Penrose inverse of matrix M [19][20].

2) Gradient descent algorithm for updating the premise parameters

The premise parameters are updated in terms of the multiple objective function defined in (9), which aims at striking a good trade-off between the global approximation ability and the interpretability of local models. The equation for updating the linguistic modifier parameters is as follows:

$$p_j(t+1) = p_j(t) - \rho \frac{\partial J}{\partial p_j} \tag{14}$$

where t is the iteration step, ρ is the learning rate, and

$$\frac{\partial J}{\partial p_j} = \varphi \frac{\partial E}{\partial p_j} + \theta \frac{\partial F}{\partial p_j} \tag{15}$$

$$\frac{\partial E}{\partial p_j} = -\sum_{k=1}^{N} \sum_{i=1}^{L} (d(k) - y(k)) y_i(k) \frac{\partial w_i(k)}{\partial p_j} \tag{16}$$

$$\frac{\partial w_i(k)}{\partial p_j} = \frac{\sum_{l=1}^{L} \tau_l(k) \frac{\partial \tau_i(k)}{\partial p_j} - \tau_i(k) \sum_{l=1}^{L} \frac{\partial \tau_l(k)}{\partial p_j}}{\left(\sum_{l=1}^{L} \tau_l(k) \right)^2} \tag{17}$$

$$\frac{\partial \tau_i(k)}{\partial p_j} = \prod_{q \neq j} A_{i_q}^{(q)}(x_q(k)) \frac{\partial A_{i_j}^{(j)}(x_j(k))}{\partial p_j} \tag{18}$$

$$\frac{\partial F}{\partial p_j} = \sum_{1 \le i_j \le L_j} \frac{\partial F(A_{i_j}^{(j)})}{\partial p_j} \tag{19}$$

$$\frac{\partial F(A_{i_j}^{(j)})}{\partial p_j} = \frac{2}{\sqrt{N}} \frac{\sum_{k=1}^{N} (A_{i_j}^{(j)}(x_j(k)) - \underline{A}_{i_j}^{(j)}(x_j(k))) \frac{\partial A_{i_j}^{(j)}(x_j(k))}{\partial p_j}}{\sqrt{\sum_{k=1}^{N} (A_{i_j}^{(j)}(x_j(k)) - \underline{A}_{i_j}^{(j)}(x_j(k)))^2}} \tag{20}$$

Furthermore, the partial derivatives of $A_{i_j}^{(j)}(x_j)$ with respect to p_j can be calculated in terms of (4) [19].

C. Rule base refinement

Because all possible combinations of fuzzy sets about one-dimensional input variables are considered in the rule base for this TS model, a common problem in fuzzy modeling, *i.e.*, the curse of dimensionality, would emerge. In this subsection we give a simple but efficient method to refine the rule base during adaptation.

For a given training data set $\{(x(k), d(k))\}_{k=1}^{N}$, the total firing strength of the ith rule R_i received from all input samples is obtained by

$$IR_i = \sum_{k=1}^{N} \tau_i(x(k)) \tag{21}$$

If

$$IR_i < firingLimit \tag{22}$$

then rule R_i will be removed, where *firingLimit* is a given lower limit of firing strength.

IV. EXPERIMENTAL RESULTS

In this section two different examples are used to evaluate the proposed method in terms of the local interpretability and global accuracy. Specifically, the first example pays attention to the transparency of local models and the second example the interpretation of partitioning of input space.

For the sake of visualizing experimental results on the transparency of local models well, the first example considers a system with one input variable and one output variable, characterized as follows:

$$y = 50(1 - \cos(\pi x / 50)) \frac{\sin(2x)}{e^{x/5}} \tag{23}$$

200 input-output data pairs were collected for the purpose of parameter identification for this system. The number of local models can be determined by examining the local properties of the data automatically or manually. In this example, there are 14 local areas with distinctive linearizations, so we set the number of fuzzy partitions on input variable x as 14. By clustering the 200 input-output data pairs using a Mercer kernel fuzzy c-means (MKFCM) clustering algorithm [21], 13 data centers on input variable x are obtained, which are then used as the crossover points of the linguistic modifiers. The minimum and maximum values of input variable x are used as the cores of 2 linguistic modifiers, and the cores of the remaining 12 linguistic modifiers are set by the midpoints of the corresponding crossover points generated above. During adaptation, the firing strength lower limit is set to be 0.0001. After 12 iterations of updating the linguistic modifier parameter, the global approximation error of the proposed TS model arrives at a minimum, as shown in Fig.2, which

indicates that if the modifier parameter is updated furthermore, the global approximation performance could be degraded. However, Fig. 3 shows that a trade-off between the global accuracy and the fuzziness measure is found at the 27th iteration by the multiple objective function, and the optimal modifier parameter value is 1.428 in terms of this multiple objective measure. It is noted that the training data fire all the 14 rules very well, and there is no rule being cancelled. The finally generated MFs are depicted in Fig.4. By using these 14 MFs in the TS model, it can be seen from Fig.5 that the corresponding 14 local models exhibit the desired good interpretability: they match the system well and tend to be the tangents of the global model, and at the same time the global prediction accuracy is still preserved.

As a comparison with the proposed method, the well-known ANFIS model [10] is used to model the same system (23), the 14 MFs generated by ANFIS method are shown in Fig.6. Although the ANFIS method can approximate the system very well, as depicted in Fig.7, obviously the interpretability of its 14 local models is poor: they exhibit erratic behaviours, and can not characterize the local activities of the modelled system well.

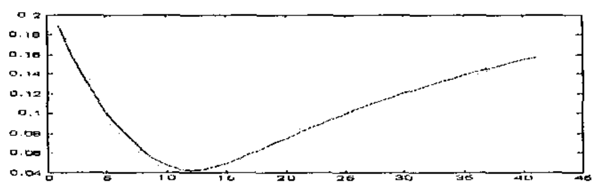

Fig.2 Model approximation error vs linguistic modifier parameter updating

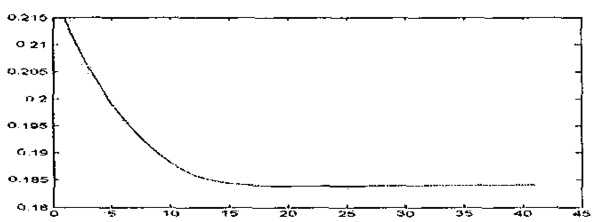

Fig.3 Multiple objective performance measure

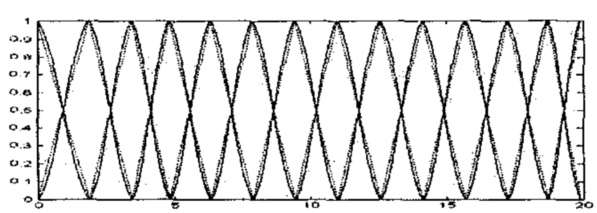

Fig.4 MFs with modifier parameters equal to 1 in dotted line (DOL) and 1.428 in solid line (SL)

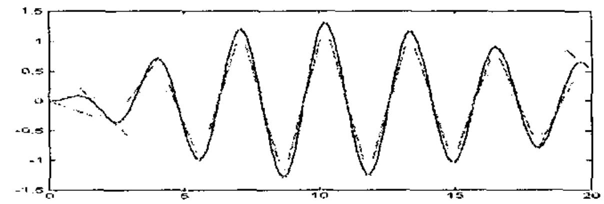

Fig.5 Good interpretability of local models obtained by the proposed method: SL -desired output, DOL -model output, dashed lines (DAL)-local models

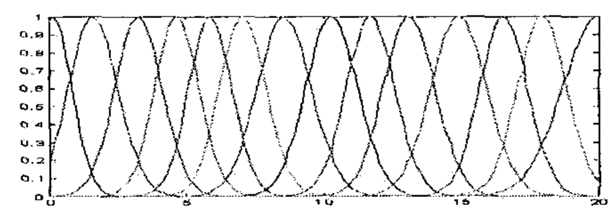

Fig.6 MFs generated by ANFIS method for the TS model

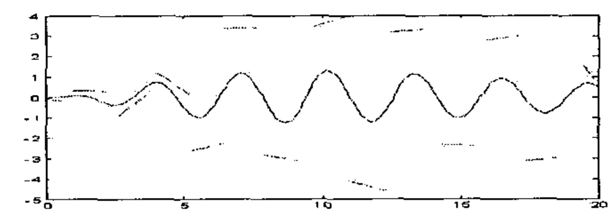

Fig.7 Poor interpretability of local models produced by ANFIS method: SL-desired output, DOL-model output, DAL-local models

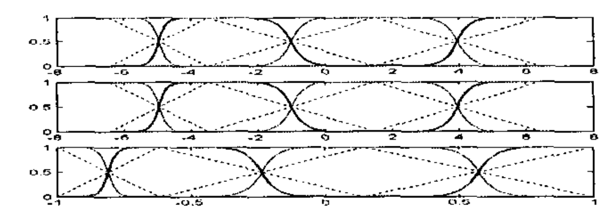

Fig.8 MFs generated for input variables $y(k-1)$ (top row), $y(k-2)$ (middle row), and $u(k-1)$ (lowest row): DOL-initial triangular MFs, SL-final MFs

The second example is to model a dynamic system with two feedback input variables, one external input variable and one output variable, which is described as follows [22]:

$$y(k) = 0.3y(k-1) + 0.6y(k-2) + 0.6\sin(\pi u) + 0.3\sin(0.3\pi u) + 0.1\sin(5\pi u) \quad (24)$$

where $u = \sin(2\pi(k-1)/250)$ is the external input. We select $(y(k-1), y(k-2), u(k-1))$ as the three input variables of the TS model, and $y(k)$ the output variable. Initially, 2000 input-output data pairs were generated to build the proposed TS model, where the first 1500 data pairs were used as training data and the latter 500 pairs as test data. Four initial fuzzy sets were set up for each input variable, whose cores can be obtained by cluster analysis algorithms. In this paper, they were obtained by the MKFCM algorithm. The firing strength lower limit is set as 0.0001. By the proposed multiple objective learning scheme, the optimal modifier parameters for the three input variables are 7.4032, 7.4020, and 7.4026 respectively. The finally generated fuzzy sets for the three input variables are depicted in Fig.8 with large core areas and small overlapping, which shows good distinguishability of partitioning of input space in terms of the criteria introduced in [6] [8][9]. As shown in Fig.9, on both training data and test data, this dynamical system is well approximated by the built TS model. Although 64 rules are constructed in the initial rule base, 31 rules are preserved in the final rule base after the refinement by the proposed method.

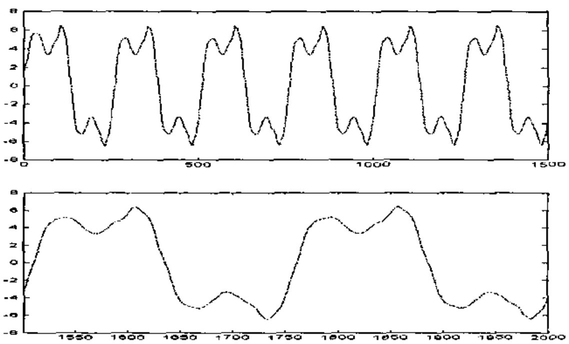

Fig.9 Approximation results of the proposed model for training data (tp) and test data (bottom): SL-desired output and DOL-model output.

V. CONCLUSION

In this paper, a new TS type fuzzy model, whose MFs are characterized by linguistic modifiers, is proposed. In the proposed model, the local models match the global model well and tend to become the tangents of the global model, and the erratic behaviors of local models are remedied greatly. Furthermore, the transparency of partitioning of input space has been improved during parameter adaptation. In order to preserve the global approximation ability, a tradeoff between global approximation and local model interpretation has been achieved by minimizing a multiple objective performance measure. Due to the promising performance exhibited, the proposed method would have potential applications to fuzzy system modeling, particularly, to nonlinear state estimation and control problems.

REFERENCES

[1] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neuro Fuzzy and Soft Computing*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[2] S. Guillaume, "Designing fuzzy inference systems from data: an interpretability-oriented review," *IEEE Trans. on Fuzzy Systems*, vol.9, pp.426-443, 2001.

[3] J. Yen, L. Wang, and C. W. Gillespie, "Improving the interpretability of TSK fuzzy models by combining global learning and local learning," *IEEE Trans. on Fuzzy Systems*, vol. 6, pp. 530-537, 1998.

[4] Y. Jin, "Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement," *IEEE Trans. on Fuzzy Systems*, vol.8, pp.212-221, 2000.

[5] M. Setnes, R. Babuska, and H. B. Verbruggen, "Rule-based modeling: precision and transparency," *IEEE Trans. on Systems, Man and Cybernetics-Part C*, vol.28, pp.165-169, 1998.

[6] J. Valente de Oliveira, "Semantic Constraints for membership function optimization," *IEEE Trans. on Systems, Man and Cybernetics-Part A*, vol.29, pp.128-138, 1999.

[7] C. A. Penna-Reyes and M. Sipper, "Fuzzy CoCo: balancing accuracy and interpretability of fuzzy models by means of coevolution," in J. Casillas, O. Cordon, F. Herrera, L Magdalena, (Eds.), *Trade-off between Accuracy and Interpretability in Fuzzy Rule-Based Modelling*, Studies in Fuzziness and Soft Computing, Physica-Verlag, 2002

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

[9] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. on Pattern Analysis And Machine Intelligence*, vol.11, pp.773-781, 1989.

[10] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 23, 1993, pp. 665~685, 1993

[11] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modelling and control," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 15, pp. 116-132, 1985

[12] S. MaGinnity and G.W. Irwin, "Nonlinear state estimation using fuzzy local linear models", *Int. J. of System Science*, vol.28, pp.643-656, 1996

[13] Q. Gan and C. J. Harris, "Neurofuzzy local linearisation for model identification and state estimation of unknown nonlinear processes", *Int. J. of Intelligent Control and Systems*, vol.3, pp.389-407, 1999

[14] A. Fink, O. nelles and R. Isermann, "Nonlinear internal model control for MISO systems based on local linear neuro-fuzzy models", *The 15th IFAC World Congress*, Barcelona, Spain, 2002

[15] Q. Gan and C. J. Harris, "Fuzzy local linearization and local basis function expansion in nonlinear system modeling," *IEEE Trans. on Systems, Man and Cybernetics-Part B.*, vol. 29, pp.559-565, 1999

[16] T. A. Johansen and R. Babuska, "Multi-objective identification of Takagi-Sugeno fuzzy models," *IEEE Trans. on Fuzzy Systems*, vol. 11, 2003 (accepted)

[17] S.-M. Zhou and Q. Gan, "Fuzzy data clustering based on information entropy measure and relaxation linguistic modifier," *UK Workshop on Computational Intelligence (UKCI03)*, Bristol, UK, Sep. 2003, pp.141-148.

[18] R. R. Yager, "On the measure of fuzziness and negation part I: Membership in the unit interval," *Int. J. Gen. Syst.*, vol. 5, 221–229, 1979.

[19] S.-M. Zhou and J. Q. Gan, "Interpretability improvement of Takagi-Sugeno fuzzy model by multiple objective learning," *Technical Report*, Department of Computer Science, University of Essex, 2003.

[20] A. Ben-Israel, "An iterative method for computing the generalized inverse of an arbitrary matrix", *Math. Comp.*, vol. 19, pp.452-455, 1965

[21] S.-M. Zhou and J. Q. Gan, "Mercer Kernel, Fuzzy C-means Algorithm, and Prototypes of Clusters", Accepted by *Fifth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'04)* August 25-27, 2004, Exeter, UK, in *Lecture Notes in Computer Science series*, Springer-Verlag, 2004

[22] K. S. Narendra and K. Parthsarathy, "Identification and control of dynamical systems using neural networks", *IEEE Trans. on Neural Networks*, vol.1, pp.4-27, 1990.