Incremental Granular Modeling for Predicting the Hydrodynamic Performance of Sailing Yachts

Keun-Chang Kwak

Dept. of Control and Instrumentation Engineering
Chosun University, 375 Seosuk-Dong
Gwangju, Korea

Abstract—This paper is concerned with a design method for modeling Incremental Granular Model (IGM) based on Linguistic Model (LM) and Polynomial Regression (PR) from data set obtained by complex yacht hydrodynamics. For this purpose, we develop a systematic approach to generating automatic fuzzy rules based on Context-based Fuzzy C-Means (CFCM) clustering. This clustering algorithm builds information granules in the form of linguistic contexts and estimates the cluster centers by preserving the homogeneity of the clustered data points associated with the input and output space. Furthermore, IGM deals with localized nonlinearities of the complex system so that the modeling discrepancy can be compensated. After performing the design of 2nd order PR as the first global model, we refined it through a series of local fuzzy if-then rules in order to capture the remaining localized characteristics. The experimental results revealed that the presented IGM showed a better performance in comparison to the previous works for predicting the hydrodynamic performance of sailing yachts.

Keywords—granular networks; particle swarm optimization; linguistic model; two-sided Gaussian contexts

I. INTRODUCTION

During the last decade, it is advantageous to combine several computing techniques synergistically, rather than exclusively, resulting in the construction of complementary hybrid intelligent systems in confronting real-world application domains such as modeling, control, and optimization for complex engineering problems [1]. Among these approaches, the neuro-fuzzy and soft computing as a computational approach to learning and machine intelligence is a widely used computing framework based on the concepts of fuzzy model, neural networks, data clustering, and several stochastic optimization methods for computational intelligence. Here, neural networks provide learning abilities and a connectionist structure to fuzzy systems. Fuzzy systems provide neural networks with a structural framework with human knowledge and reasoning. A considerable number of these studies have been conducted to generate automatic fuzzy if-then rules using clustering techniques from given numerical training data sets [2-4]. On the other hand, Linguistic Model (LM) has been researching to present a nonlinear and complex characteristic based on Context-based Fuzzy C-Means (CFCM) clustering [5]. In contrast to the context-free clustering methods such as FCM clustering [6], subtractive clustering, and mountain clustering, the CFCM clustering is to generate clusters preserving homogeneity of the clustered patterns in connection

with their similarity in the input and output variables [7]. The studies associated with this clustering approach are LM [5], RBFN(Radial Basis Function Networks)-CFCM [8], LM with learning [9], TSK(Takagi-Sugeno-Kang)-LM [10], and Incremental Model (IM) [11]. In this paper, we develop the Incremental Granular Model (IGM) for predicting complex hydrodynamic performance of sailing yacht. The presented IGM is based on LM and Polynomial Regression (PR) to capture the localized nonlinear characteristics. First, we build a PR which could be treated as a preliminary construct. Next, all modeling discrepancies are compensated by a collection of rules that become attached to the regions of the input space where the error is localized. The experiments are achieved by the data set obtained from the complex yacht hydrodynamics. The prediction of resistance of the ship at the initial design stage is of a great value for evaluating the ship's performance and for estimating the required propulsive power [12]. We compared the effectiveness of the presented IGM with the previous works such as RBFN, LM, RBFN-CFCM, and IM. This paper is organized as follows. In Section 2, we describe the architecture of CFCM clustering and LM as a framework of user-centric system modeling. In the Section 3, we present the proposed IGM based on LM and PR. In Section 4, we present the prediction problem of the hydrodynamic performance of sailing yachts [13] and the experimental results. Finally, concluding comments are given in Section 5.

II. CFCM CLUSTERING AND LINGUISTIC MODEL

The CFCM clustering as an interesting variant of the fuzzy c-means is realized via individual contexts as shown in Fig. 1. Each linguistic context has defined semantics that can be interpreted as a *large* negative error, *medium* negative error, etc in the design of IGM.

Let us consider a certain fixed context W_j described by some membership function. The data point in the output space is associated with the corresponding membership value. Let us introduce a family of the partition matrices induced by the *l*-th context and denote it by $\mathbf{U}(W_l)$

$$\mathbf{U} = \left\{ u_{ik} \in [0,1] \mid \sum_{i=1}^{c} u_{ik} = w_{lk} \; \forall k \text{ and } 0 < \sum_{k=1}^{N} u_{ik} < N \right\}$$

(1)

where w_{lk} denotes a membership value of the *k*-th datum implied by the *l*-th context. The underlying objective function is as follows

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \parallel x_k - v_i \parallel^2 \qquad (2)$$

where v_i denotes the *i*-th cluster center. The J is minimized under the constraints imposed by (1) as follows

Min Q subject to $U(W_l)$, *l*=1, 2, …,p $\qquad$ (3)

The minimization of J is realized by iteratively updating the values of the partition matrix and the cluster centers. The successive updates of the partition matrix are completed as follows

$$u_{ik} = \frac{w_{1k}}{\sum_{j=1}^{c} \left(\frac{\parallel \mathbf{x}_k - \mathbf{v}_i \parallel}{\parallel \mathbf{x}_k - \mathbf{v}_j \parallel} \right)^{\frac{2}{m-1}}} \qquad (4)$$

where u_{ik} is the partition matrix induced by the *l*-th context. The cluster centers are as the following expression

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} \mathbf{x}_k}{\sum_{k=1}^{N} u_{ik}^{m}} \qquad (5)$$
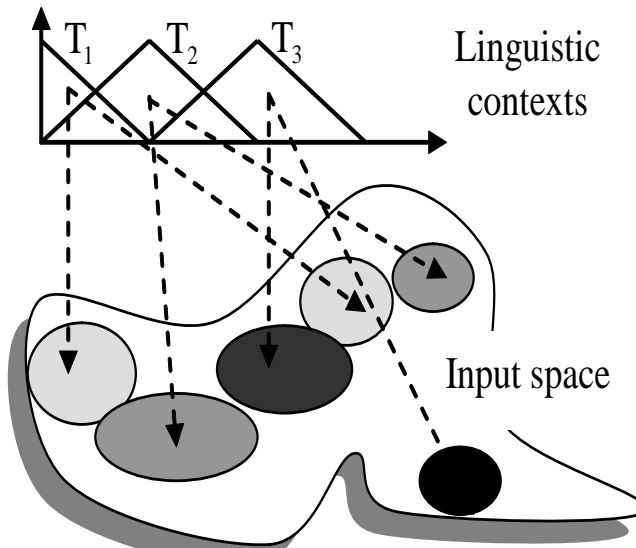

Fig. 1. Concept of context-based fuzzy clustering

In the design of the LM, we consider the contexts to be described by triangular membership functions being equally distributed in the error space **E** with the 1/2 overlap occurring between two successive fuzzy sets. Each context generates a number of induced clusters whose activation levels are afterwards summed up as shown in Fig. 2.
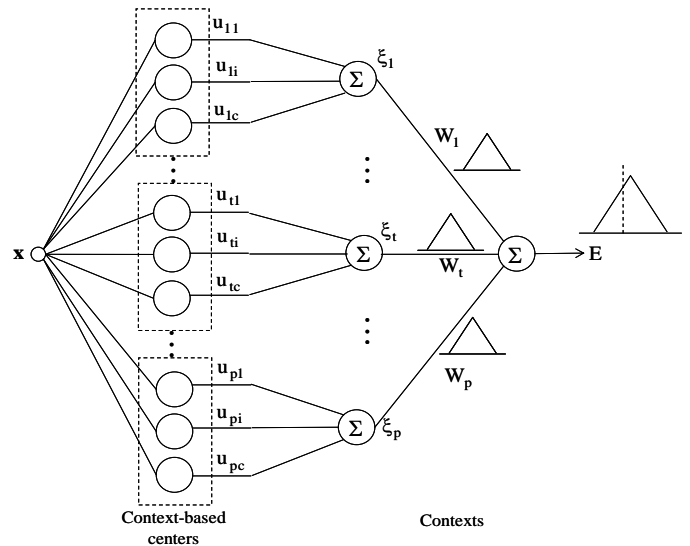

Fig. 2. Architecture of the LM

Assuming the triangular form of the contexts, the result is a triangular fuzzy number E as follows

$$E = W_1 \otimes \xi_1 \oplus W_2 \otimes \xi_2 \oplus .... W_n \otimes \xi_n \qquad (6)$$

We denote the algebraic operations by $\otimes, \oplus$ to emphasize that the underlying computing operates on a collection of fuzzy numbers. As such, E is characterized by its three parameters that are a modal value, the lower bound, and upper bound.

III. INCREMENTAL GRANULAR MODEL (IGM)

The PSO method is one of swarm intelligence methods for solving the optimization problems. The PSO algorithm proposed by Kennedy is performed by social behavior of bird flocking or fish schooling. The character of PSO easily can handle fitness function for solving complex problems. Furthermore, it can control a relationship between global and local search. Here, each particle adjusts information of location with experience of them and their neighborhood. It can form the answer of optimum in short time. As the velocity of particle movement of PSO is only demanded, it is easy to be embodiment and brevity of a theory. The basic element of PSO is simply as follows IGM performs localized nonlinearities of the complex and nonlinear system so that the modeling discrepancy can be compensated. After performing the design of 2nd order Polynomial Regression (PR) as the first global model, we refined it through a series of local fuzzy if-then rules in order to capture the remaining localized characteristics. Fig. 3 shows the main design process of the IGM. Firstly, we decide upon the granularity of information to be used in the development of the model such as the number of contexts and the number of clusters formed for each context. The design procedure of IGM is as follows

[Step 1] Design PR in the input and output space. PR is the extended form of the well-known Linear Regression (LR) in which the relationship between the independent variables $\mathbf{x}_k$ and dependent variable y is modeled as an 2nd order polynomial. PR fits a nonlinear relationship between the value of $\mathbf{x}_k$ and the corresponding conditional mean of y. On the basis of the original data set, a collection of input-error pairs, $(\mathbf{x}_k, e_k)$ is obtained.

[Step 2] Produce linguistic contexts in the error space of the regression model E_1, E_2, …, E_p. The distribution of these fuzzy sets is obtained through the use of statistical distribution or fuzzy equalization while the fuzzy sets are characterized by triangular membership functions with a 0.5 overlap between neighboring fuzzy sets.

[Step 3] Perform CFCM clustering in the input-output space from the linguistic contexts produced in the error space.

[Step 4] Compute the activation levels of the clusters induced by the corresponding contexts and their overall aggregation through weighting by fuzzy sets of the context leading to the triangular fuzzy number of output, $E = F(\mathbf{x}; E_1, E_2, …, E_p)$.

[Step 5] The output of the IGM is then combined with the output of the linear part. The result is a shifted triangular number $Y = z \oplus E$.
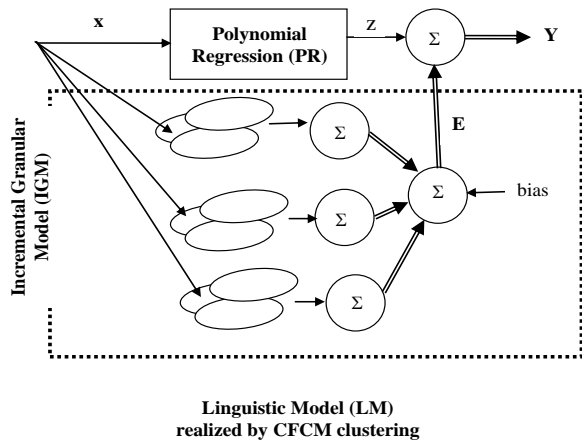

Fig. 3. Main design process of IGM based on PR and LM

IV. EXPERIMENTAL RESULTS

In this experiment, we use to predict hydrodynamic performance of sailing yachts from dimensions and velocity. The well-known Holtrop and Mennen method [14] is widely used as the initial design stage of ships for estimating the resistance of the ships. This method provides a prediction of the total resistance's components. The total resistance of a ship has been subdivided as follows

$$R_{total} = R_F(1+k_1) + R_{APP} + R_W + R_B + R_{TR} + R_A \qquad (7)$$

R_F : Frictional resistance according to the ITTC-1957 friction formula

$(1+k_1)$: Form factor describing the viscous resistance of the full form in relation to R_F

R_{APP} : Resistance of appendages

R_W : Wave-making and wave-breaking resistance

R_B : Additional pressure resistance of bulbous bow near the water surface

R_{TR} : Additional pressure resistance of immersed transform stern

R_A : Model-ship correlation resistance.

The form factor of the hull the prediction formula is as the following equation

$$(1+k_1) = c_{13}\{0.93 + c_{12}(B/L_R)^{0.92497}(0.95 - C_p)^{-0.521448} \cdot$$
$$(1 - C_p + 0.0225lcb)^{0.6096}\}$$

(8)

Here, R_B is the prismatic coefficient based on the waterline length L and lcb is the longitudinal position of the centre of buoyancy forward of 0.5L as a percentage of L. L_R is a parameter reflecting the length of the run. In this example, six input variables are composed of longitudinal position of the center of buoyancy, prismatic coefficient, length-displacement ratio, beam-draught ratio, length-beam ratio, and Froude number as shown in Fig. 4. The output variable to be predicted by the six input variables is residuary resistance per unit weight of displacement. The overall data set consists of 308 full experiments, which were performed at the Delft Ship Hydromechanics laboratory. These experiments include 22 different hull forms.
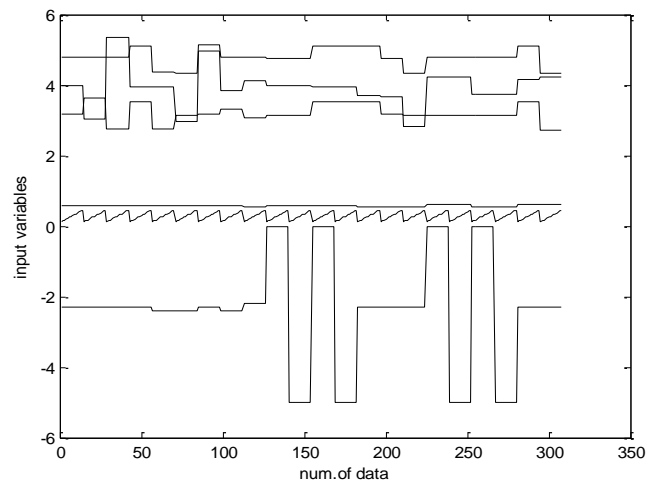

Fig. 4. Six input variables

We randomly divide the data set into training and test data with 60%-40% ratio in the normalized input space between 0 and 1, respectively. The experiments perform 10 runs. The training data set is used for model construction, while the test set is used for model validation. Thus, the resultant model is not biased toward the training data set and it is likely to have a better generalization performance with respect to new data. Firstly, 2nd order PR is performed in the input and output space.

After that, the regression error is obtained as the histogram shown in Fig.5. Fig. 6 visualizes six linguistic contexts to perform CFCM clustering. These contexts are produced by the use of statistical distribution. Here, we assume that the number of cluster per each context is the same.
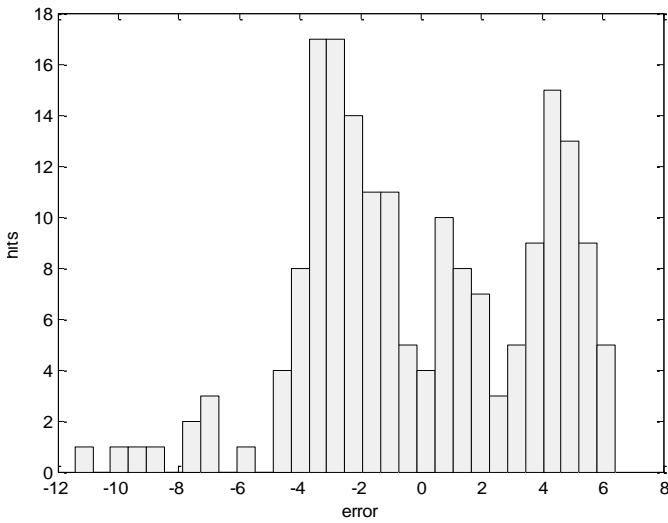

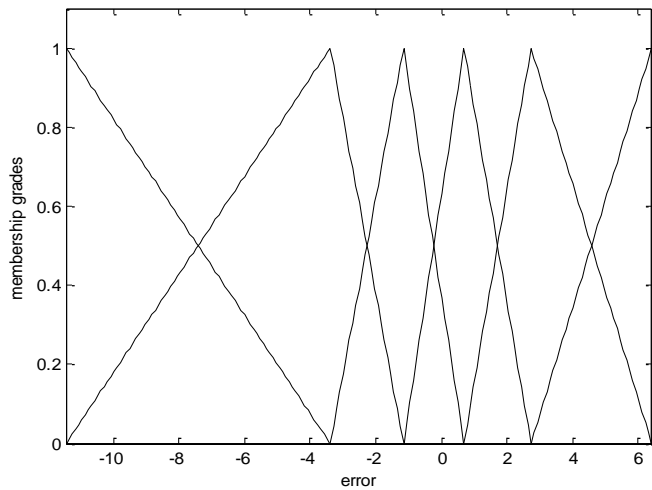
Fig. 5.   Histogram of error obtained by PR

Fig. 6.   Six linguistic contexts obtained in the error space

Fig. 7 shows comparison results between the desired and model output for both the training and test data sets, respectively. As shown in Fig. 7, it is obvious that the proposed IGM has good approximation and generalization performance. Table 1 lists the mean of RMSE and the number of rules for the training and test sets, respectively. In the design of IGM, we obtained the best results in three contexts and two clusters in each context for CFCM clustering as listed in Table 1. Although the conventional LM has a structured knowledge representation in the form of fuzzy if-then rules, it lacks the adaptability to deal with a complex and nonlinear model. Moreover, we obtained each best RMSE result through the construction of RBFN, LM, RBFN-CFCM, and IM in trial and error as listed in Table. 1.
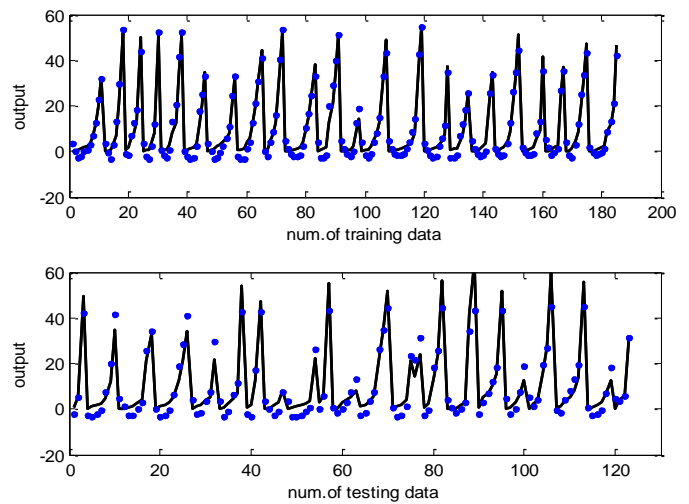

Fig. 7.   Approximation and generalization performance for training and test data set (solid line: actual output, dotted line: model output)

The augmented granular modification of the model was realized by experimenting with the two essential parameters controlling the granularity of the construct in the input and output space, that is "p" and "c". The corresponding results are summarized in Table 2 and 3. Fig. 8 visualizes the uncertain prediction values from upper and lower output obtained from the proposed IGM. As a consequence, the experimental results revealed that the presented IGM yielded a better performance in comparison to RBFN, LM, RBFN-CFCM, and IM for predicting the hydrodynamic performance of sailing yachts as shown in Table 1 and Fig. 8.

TABLE I.   COMPARISON RESULTS OF THE RMSE AND THE NUMBER OF RULES

Methods	no. rule	RMSE (training)	RMSE (test)
Linear regression	-	8.6690	9.3006
RBFN	6*	13.6393	14.9378
LM [5]	6(p=3,c=2)	10.3391	11.7420
RBFN-CFCM [8]	10*(p=5,c=2)	8.1735	9.7952
IM [11]	12(p=6,c=2)	6.4082	7.8182
The proposed IGM	6(p=3,c=2)	3.9333	4.4344
	8(p=4, c=2)	4.0367	4.5022
	10(p=5,c=2)	4.0030	4.5290
	12(p=6,c=2)	3.9298	4.4811

TABLE II.   RMSE (MEAN) - TRAINING DATA

		No. of contexts (p)			
		3	4	5	6
No. of clusters per context (c)	2	3.9333	4.0367	4.0030	3.9298
	3	3.9599	3.9589	3.9546	3.8690
	4	3.9428	3.8716	3.8399	3.7762
	5	3.8243	3.8109	3.7325	3.7190
	6	3.8064	2.7643	3.6775	3.6542

TABLE III.   RMSE (MEAN) – TESTING DATA

		No. of contexts (p)			
		3	4	5	6
No. of clusters per context (c)	2	4.4344	4.5022	4.5290	4.4811
	3	4.5054	4.5114	4.5705	4.5190
	4	4.5435	4.5418	4.5965	4.5835
	5	4.5182	4.5858	4.5875	4.6137
	6	4.5814	4.6392	4.5997	4.5943

V.   CONCLUSIONS

We developed the incremental granular model with uncertainty output based on polynomial regression and linguistic model realized by the context-based fuzzy c-means clustering. Furthermore, we dealt with an incremental model to deal with localized nonlinearities of the system so that all modeling discrepancies can be compensated. This incremental model is quite different from one frequently used in conjunction with fuzzy modeling with the predominant concept of a rule-based architecture. The experimental results on complex yacht hydrodynamics revealed that the presented method showed a good approximation and generalization ability in comparison to conventional other methods. Thus, these results lead to the conclusion that incremental granular model can be represented as the prototypes that exhibits certain characteristics of the complex system to be modeled.

REFERENCES

[1] J. S. R. Jang, C. T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Prentice Hall, 1997.

[2] K. C. Kwak and D. H. Kim "Adaptive Neuro-Fuzzy Networks with the Aid of Fuzzy Granulation", *IEICE Trans. on Information and Systems*, Vol. E88-D, No. 9, pp.2189-2196, 2005.

[3] 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.

[4] 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", *International Journal of Knowledge-Based Intelligent Engineering Systems*, Vol. 4, No. 4, pp.230-236, 2000.

[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.

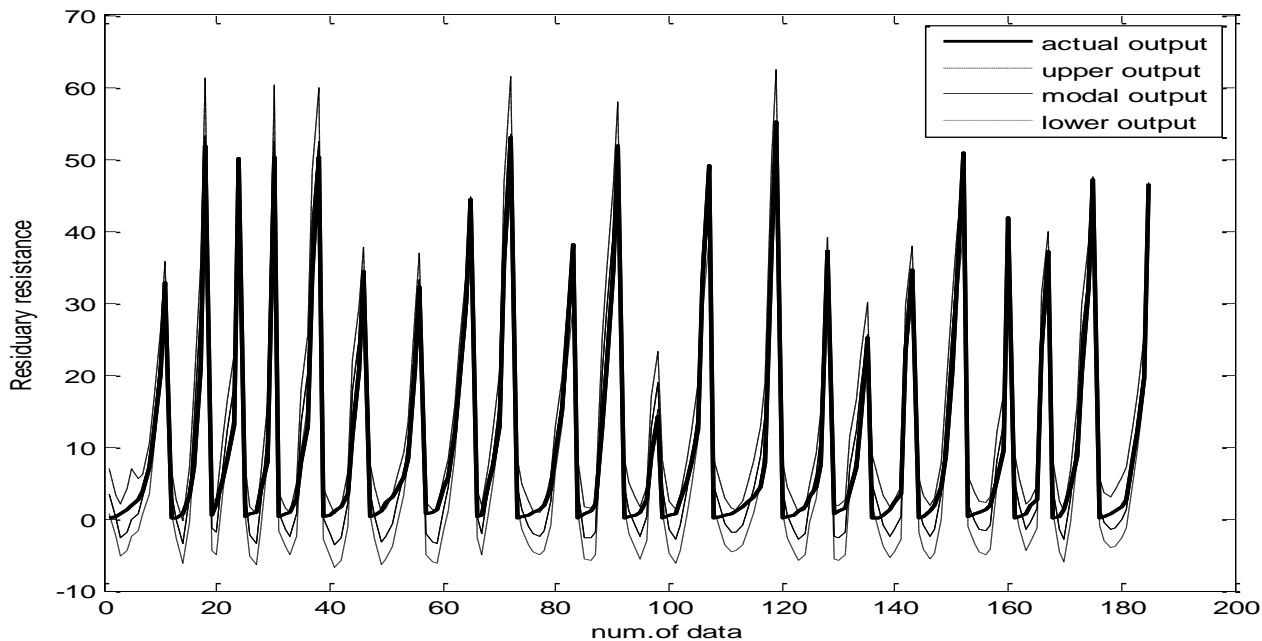[6] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

Fig. 8.   Prediction performance with uncertain output

[7] W. Pedrycz, "Conditional fuzzy C-Means", *Pattern Recognition Letters*, Vol. 17, pp. 625-632, 1996.

[8] W. Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks", *IEEE Trans. on Neural Networks*, Vol. 9, No. 4, pp.745-757, 1999.

[9] W. Pedrycz and K. C. Kwak, "Linguistic models as a framework of user-centric system modeling", *IEEE Trans. on Systems, Man, and Cybernetics-Part A,* Vol. 36, No. 4, pp.727-745, 2006.

[10] K. C. Kwak and D. H. Kim, "TSK-based linguistic fuzzy model with uncertain model output", *IEICE Trans. on Information and Systems*, Vol. E89-D, No. 12, pp.2919-2923, 2006.

[11] W. Pedrycz and K. C. Kwak, "The development of incremental models", *IEEE Trans. on Fuzzy Systems*, Vol. 15, No. 3, pp. 507-518, 2007.

[12] I. Ortigosa, R. Lopez and J. Garcia. "A neural networks approach to residuary resistance of sailing yachts prediction", In Proceedings of the International Conference on Marine Engineering MARINE 2007, 2007.

[13] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/index.html

[14] J. Holtrop and G.G.J. Mennen, "An approximate power prediction method", International Shipbuilding Progress, pp. 166-170, 1982.

AUTHOR PROFILE

Keun-Chang Kwak received the B.Sc., M.Sc., and Ph.D. degrees from Chungbuk National University, Cheongju, Korea, in 1996, 1998, and 2002, respectively. During 2003–2005, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. From 2005 to 2007, he was a Senior Researcher with the Human–Robot Interaction Team, Intelligent Robot Division, Electronics and Telecommunications Research Institute, Daejeon, Korea. He is currently the Assistant Professor with the Department of Control, Instrumentation, and Robot Engineering, Chosun University, Gwangju, Korea. His research interests include human–robot interaction, computational intelligence, biometrics, and pattern recognition. Dr. Kwak is a member of IEEE, IEICE, KFIS, KRS, ICROS, KIPS, and IEEK..