nternational Journal of Fuzzy Logic and Intelligent Systems, vol. 6, no. 2, June 2006 pp. 138-143

Optimal Fuzzy Models with the Aid of SAHN-based Algorithm

Jong-Seok Lee, Kyung-Won Jang* and Tae-Chon Ahn*+**

***Dept. of Electrical Electronic and Information Eng., Wonkwang University, 344-2 Shinyong-dong, Iksan, Chon-Buk, Republic of Korea, 570-749**

****Dept. of Radiotechnology, Wonkwang Health Science College, 344-2 Shinyong-dong, Iksan, Chon-Buk, Republic of Korea, 570-749**

Abstract

In this paper, we have presented a Sequential Agglomerative Hierarchical Nested (SAHN) algorithm-based data clustering method in fuzzy inference system to achieve optimal performance of fuzzy model. SAHN-based algorithm is used to give possible range of number of clusters with cluster centers for the system identification. The axes of membership functions of this fuzzy model are optimized by using cluster centers obtained from clustering method and the consequence parameters of the fuzzy model are identified by standard least square method. Finally, in this paper, we have observed our model's output performance using the Box and Jenkins's gas furnace data and Sugeno's non-linear process data.

Key Words : *SAHN based algorithm, Fuzzy inference system, Membership, Standard least square method.*

1. Introduction

The basic structure of a fuzzy inference system consists of three conceptual components: a rule base, which contains a selection of fuzzy rules; a data base which defines the membership functions used in fuzzy rules; and a reasoning mechanism, which performs the inference procedure. Fuzzy modeling has been a focal point of the technology to use fuzzy inference technique and to deal with complex, ill-defined and uncertain systems.

The main theme of clustering analysis is to classify the objects (data) in accordance with similarities among them. There are different type of clustering techniques available: HCM (Hard C-means), FCM (Fuzzy C-means), Hierarchical and so on. HCM and FCM are representative method to deal with to efficient system identification. However, these algorithms require a priori knowledge about dataset and may fail to give an optimal result.

In this paper, we have introduced the alternative method of SAHN-based data clustering algorithm to detect the underlying data patterns (clusters) and to give possible range of number of clusters with cluster centers which are used in our fuzzy model to get optimal performance[1]. In our model, this algorithm has been applied for system parameter identification.

In this paper, our model have implemented with constant, linear, quadratic and modified quadratic type inference using triangular type membership function for two inputs. The membership functions were defined by cluster centers obtained from clustering method. The consequence parameters of the fuzzy model are identified by standard least square method. In order to evaluate, we have used Box's and Jenkins's gas furnace data set [2] and Sugeno's nonlinear process data set [3] and finally compared our model's performance with other fuzzy models previously developed by researchers.

Manuscript received Apr. 12, 2006; revised Jun. 5, 2006.
+ Corresponding Author
This work was supported by a research fund provided by Wonkwang Health Science College in 2004.

2. Optimization of Fuzzy Inference System

2.1 Clustering

Partition a set of objects (data) with similarity into group is called cluster. In hierarchical agglomerative clustering (bottom-up), process starts with the individual object (data) and grouping the most similar ones to join cluster with maximum similarity [1].

For decide axes of MF, we have applied the sequential agglomerative hierarchical nested (SAHN) model based data clustering algorithm. In this experiment, we have used this algorithm to detect underlying data patterns which gives possible range of the number of clusters. The obtained cluster centers, we have used to our fuzzy model and shown optimal performance index than conventional fuzzy model.

2.2 SAHN based algorithm

The SAHN-based algorithm and pattern estimation parameters are explained shortly with several steps:

Step 1: Similarity measure with nearest pair detection.

The similarity measure between data points is obtained from:

$$d_{ik} = \sum_{j=1}^{m}(x_{ij} - x_{kj})^2 \quad i=1\cdots \;,\; k=i+1\cdots \tag{1}$$

Where j=1⋯ n, n is the number of data, mis the dimension of data and d_{ik} is the component of distance matrix formed by Euclidean distances.

Step 2: Calculate the centroid v_s.

In this step, we calculate the centroid v_s of the selected data pair (or cluster) by equation (2).

$$v_s = \frac{1}{|c_s|}\sum_{k,x_k \in c_s} x_k \tag{2}$$

Step 3: Calculation of the pattern estimation parameter P_s and ΔP_s.

a) Average distance measure of the point-to-centroid of newly clustered data group; in equation (3), d_s denotes sum of the distance of the point-to-centroid and denotes average distance of the sum of distance for the point-to-centroid. Where, vs is the cluster center (centroid) of the s-th clustering process, Ps is the average distance and $|c_s|$ is the cardinality of the selected data pair (or cluster) in the clustering process.

$$d_s = \sum_{k,x_k \in c_s} \|x_k - v_s\|^2, P_s = \frac{d_s}{|c_s|} \tag{3}$$

b) Pattern estimation parameter P_s and ΔP_s; in this phase, proposed method applied weighted average value P_s and its variation ΔP_s to observe under-lying data patterns in data set.

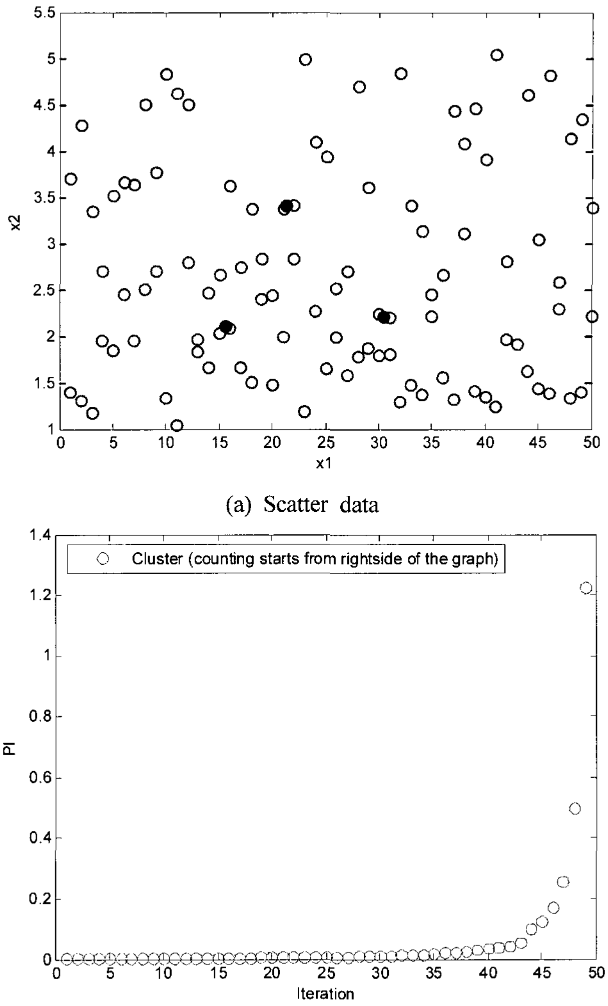
$$p_s = \frac{1}{c_s} p_s, \Delta p_s = p_s - p_{s-1} \tag{4}$$

Where C_s is the number of clusters at the s-th clustering process and total process is n-1 [1].

By this SAHN base algorithm technique, we have extracted cluster centers to detect Performance Index of our proposed fuzzy model.

In Fig.1 (a) ' ': denotes data point and ' ' : denotes center of data and in Fig.1(b) shows number of clusters.

Simulation results of SAHN based algorithm are shown in Fig. 1.



(a) Scatter data

(b) Number of Cluster

Fig.1. Scatter plot of Sugeno's data and selected number of Clusters

3. Identification of the Fuzzy Model

A fuzzy model is a nonlinear model that consists of a set of fuzzy if-then rules. It has a transparent and interpretable model structure and is capable of representing a highly nonlinear functional relation using a reasonable number of fuzzy rules. Fuzzy model consists of two basic elements: i) Premise part, ii) Consequence part [3].

3.1 Premise Identification

The structure of premise part of fuzzy model involves four tasks:

i) Selection of input variable: in our paper we have used two input variables: for gas furnace data, we have used inputs u(t-3), $y(t-1)$ and Sugeno's data inputs x_1 and x_2.

ii) Selection of membership function: in our model we have used 2MF or 3MF triangular type.

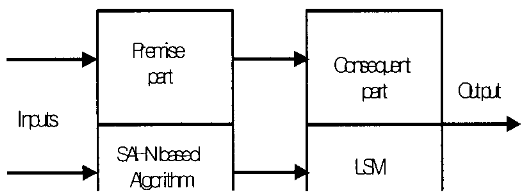


Fig.2. Basic Fuzzy Model Diagram

iii) Determination of Axes of MF: in the paper, Min-Max is used as the conventional method and SAHN based algorithm is used as a new method.

iv) Determine the number of fuzzy rules: we have used 2x2, 2x3, 3x2 and 3x3 rules.

3.2 Consequence Identification

To determine the unknown parameters of this part of model, we have used least-squares method. For example, we have shown the equation for Quadratic type inference and also describe the other equations. The expression of quadratic equation shows below:

The expression of quadratic equation shows below:

R^j : if x_1 is A_{j1} and x_2 is A_{j2} then $y=f(x_1, x_2)$

$y_j=a_{j0}$ for constant

$y_j=a_{j0}+a_{j1}x_1+a_{j2}x_2$ for linear

$y_j=a_{j0}+a_{j1}x_1+a_{j2}x_2+a_{j3}x_1x_2$ for modified quadratic

$y_j=a_{j0}+a_{j1}x_1+a_{j2}x_2+a_{j3}x_1^2+a_{j4}x_2^2+a_{j5}x_1x_2$ for Quadratic (5)

Output of the model:

$$\hat{y}=\frac{\sum_{j=1}^{n} w_{ji} y_j}{\sum_{j=1}^{n} w_{ji}}=\sum_{j=1}^{n}\hat{w}_{ji}(a_{j0}+a_{j1}x_1+a_{j2}x_2+a_{j3}x_1^2+a_{j4}x_2^2+a_{j5}x_1x_2) \quad (6)$$

Where, n is the number of fuzzy rule, $j=1\cdots n$, $i=1\cdots n$, A_{j1} are membership function, a_{j0}, $a_{j1}\cdots$ are constant, w_{ji} is the premise fitness of R^j is the fuzzy rule and $\hat{w}_{ji}$ is the normalized premise fitness of R^j.

Least-square method (LSE): Consequent parameter such as a_{j0}, a_{j1} ······ are found by least-square method is determined by the following equation :

$$X^T XA = X^T Y$$
$$A = (X^T X)^{-1} X^T Y \quad (7)$$

Where, A is the consequent parameter.

So, performance index of our model as a sum of squared errors shown in equation (8).

$$PI = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \quad (8)$$

Where m is the total number of data, y_i is real output and $\hat{y}_i$ is our model output.

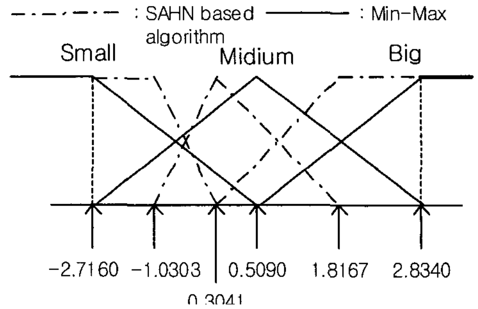
4. Simulation Results

In this experiment, our proposed rule-base fuzzy model has been used time series data of gas furnace used by Box's and Jenkins's dataset [2] and Sugeno's nonlinear process data set [3].

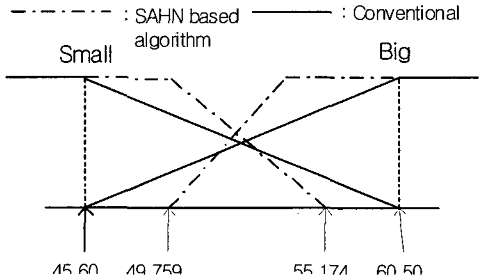
4.1 Gas Furnace Data

For gas furnace data, we have used the data to detect performance index of our model. We have divided the total data set by training data set (1-148) and testing data set (149-296) respectively. In table we have shown PI and E_PI for training and testing data set. For simulation, we have used two inputs $u(t-3)$, $y(t-1)$ and one output $y(t)$.

In Fig. 3, we have shown cluster center obtained from SAHN based algorithm and min-max values from Jenkins's data. Here we have shown membership function points are changed by cluster centers.

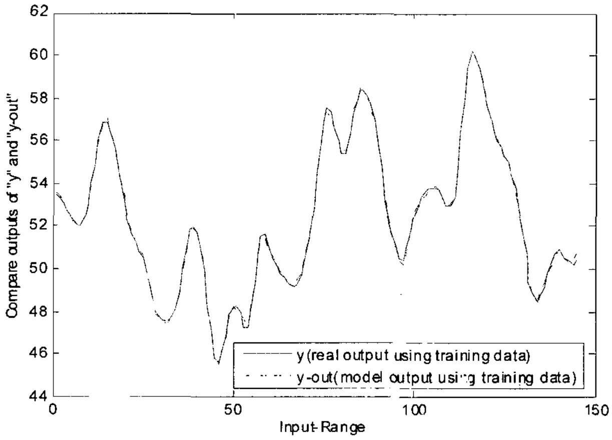


(a) $u(t-3)$

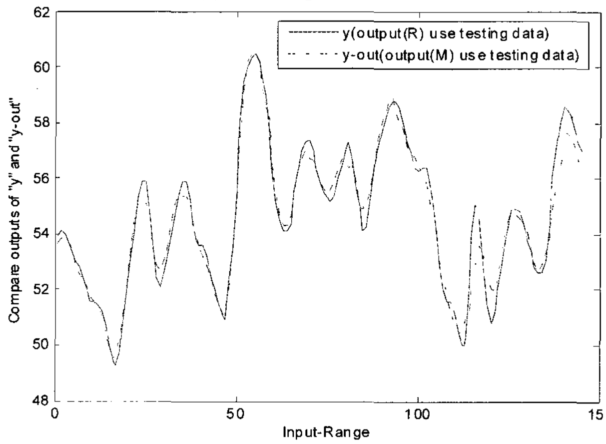


(b) $y(t-1)$

Fig. 3. Partition of Fuzzy model input space (using Jenkins's dataset



(a) Training Data



(b) Testing Data

Fig. 4 Simulation result for model's output using SAHN based algorithm for quadratic type (3X2 rule)

Table 1. Result sheet of Our Model for PI and E_PI using Jenkins's Training and Testing Dataset

No. of Rule	Type	PI	E_PI
2 x 2	Constant	0.9580	0.8538
	Linear	0.0197	0.3149
	Quadratic	0.0179	0.2994
	Modified Q	0.0191	0.3072
2 x 3	Constant	0.4748	0.8980
	Linear	0.0180	0.2961
	Quadratic	0.0149	0.2896
	Modified Q	0.0167	0.2931
3 x 2	Constant	0.6729	0.9411
	Linear	0.0179	0.2853
	Quadratic	0.0154	0.2740
	Modified Q	0.0171	0.2872
3 x 3	Constant	0.3702	0.9426
	Linear	0.0161	0.2838
	Quadratic	0.0135	0.3372
	Modified Q	0.0151	0.2995

Table 2. Comparison of Performance Index (PI) for several Fuzzy Models using Jenkins's data

Model	Type	No. Input	No. Rule	PI	E_PI
Pedrycz & Oh	Linear (Trianguar)	2	6	0.021	0.364
Min-Max [4]	Linear (Trianguar)	2	6	0.022	0.336
HCM [6]	Linear (Trianguar)	2	6	0.018	0.286
HCM+GA [6]	Linear (Trianguar)	2	6	0.020	0.264
Our model	Linear (Trianguar)	2	6	0.0179	0.2853

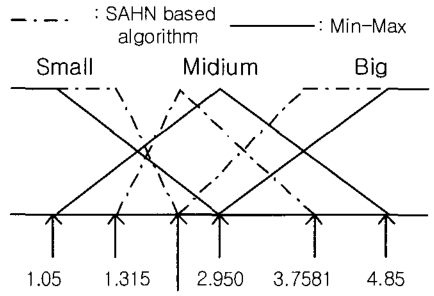
4.2 Sugeno's Nonlinear Data

We have also used Sugeno's nonlinear input-output data set for Performance Index (PI) of our proposed model.

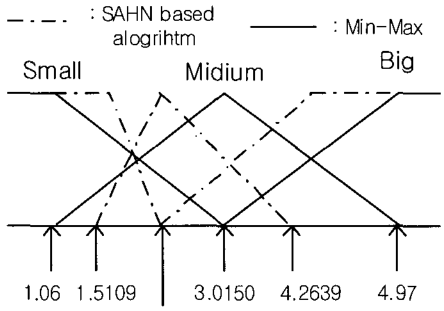
Sugeno's input(x_1, x_2) and output(y) equation for nonlinear process shown below:

$$y = (1 + x_1^{-2} + x_2^{-1.5})^2, \text{ where } 1<=x_1,\ x_2<=5 \qquad (9)$$

In Fig.5, we have shown cluster center obtained from SAHN based algorithm and min-max values from Sugeno's nonlinear process data. Here we have shown membership function points are changed by using cluster centers.

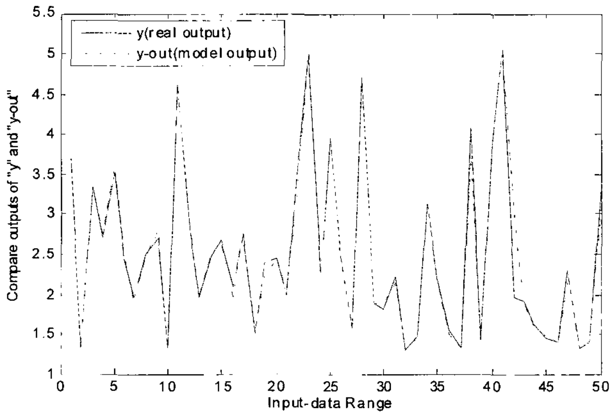


(a) Input x1

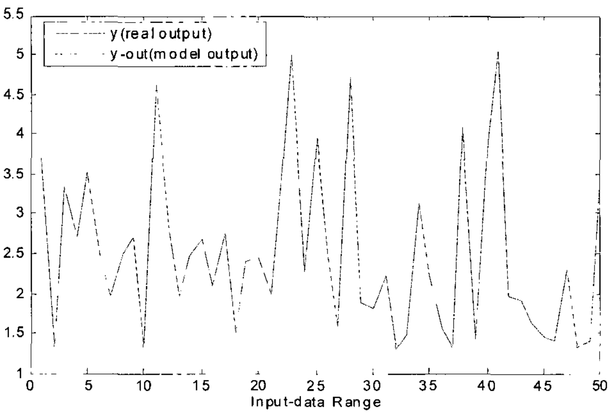


(b) Input x2

Fig. 5. Partition of Fuzzy model input space (using Sugeno's nonlinear dataset



(a) Sugeno' data(6 rule)



(b) Sugeno' data (9 rule)

Fig. 6 Simulation result for model's output using SAHN based algorithm for quadratic type (3X2 rule)

Table 3 Result sheet of Our Model for PI using Sugeno's nonlinear process Dataset

No. of Rule	Type	PI
2 x 2	Constant	0.3508
	Linear	0.0658
	Quadratic	0.0347
	Modified Q	0.0539
2 x 3	Constant	0.2331
	Linear	0.0499
	Quadratic	0.0107
	Modified Q	0.0335
3 x 2	Constant	0.2111
	Linear	0.0447
	Quadratic	0.0264
	Modified Q	0.0387
3 x 3	Constant	0.1049
	Linear	0.0268
	Quadratic	1.585e-16
	Modified Q	0.0183

In Table 3, we have shown, good performance index (PI=1.5850e-016) to use SAHN algorithm in our model for quadratic type (3X3 rule).

Table 4 Comparison of Performance Index (PI) for several Fuzzy Models using Sugeno's nonlinear data

Model	No. of Rule		PI	
Sugeno and Yasukawa [3]	6		0.079	
Tong[7]	19		0.569	
Pedrycz[8]	81		0.320	
Xu [9]	25		0.328	
Our model	6(C)	9(C)	0.211	0.1049
	6(L)	9(L)	0.044	0.0268
	6(Q)	9(Q)	0.026	1.5e-16

5. Discussion and concluding remarks

In this paper, we have shown to implement a new approach rule-based fuzzy model using SAHN-based data clustering algorithm and fuzzy inference method. In the premise part of our proposed model, we have used cluster centers extracted by using that clustering algorithm and in consequence part, we have used least-squares method (LSM). In our model, we have used constant, linear, quadratic, modified quadratic type for our model optimal performance and we have shown that quadratic type performed better than others. In order to evaluate the performance index of our model, we have used Box's and Jenkins's gas furnace dataset and Sugeno's nonlinear dataset. Finally, we have observed that using cluster centers, our model's performed well than conventional method. We have compared our model to other several fuzzy models and shown that our model's performance is well. We know that the fuzzy model is a well known systematic, effective and efficient model for nonlinear process. So, we hope that our proposed model is an improved fuzzy model than other fuzzy model developed. In future research, we will modify our model design for further good performance using constant, linear, quadratic and modified quadratic type inference mechanism.

References

[1] K.W. Jang, T. C. Ahn, "An Empirical study on Data Pattern Estimation with Center of Gravity", *Proceedings of IEEK Summer Conference*, vol.26, no. 1, pp.1541-1544, 2003.

[2] G. E. P. Box, G.M. Jenkins, "*Time Series Analysis-Forcasting and Control*", Holden Day, Inc., 1976.

[3] M. Sugeno, T. Yasukawa, "A Fuzzy Logic Based Approach to Qualitative Modeling", *IEEE Trans. on Fuzzy Systems*, Vol.1, No.1, PP. 7-31, 1993.

[4] S. K. Oh and W. Pedrycz, "Identification of fuzzy systems by means of an auto-tuning algorithm and its application to nonlinear systems", *Fuzzy Sets and Syst.* Vol. 115, pp. 205-230, 2000.

[5] B. J. Park, C. S. Park, S. K. Oh and H. K. Kim, "Optimization of fuzzy system by means of GA", *Proc. KFIS*, pp. 112-115,1998.

[6] B. J. Park, W. Pedrycz, S. K. Oh, "Identification of fuzzy model with the aid of evolutionary data granulation", *IEE Proc.-CTA*, vol. 148, no. 5, pp. 406-418, 2001.

[7] R. M. Tong, "The evolution of fuzzy models derived from experimental data", *Fuzzy Sets Syst.*, vol. 4, pp. 1-12, 1980.

[8] W. Pedrycz, "An identification algorithm in fuzzy relational systems", *Fuzzy Sets Syst.*, vol. 13, pp. 153-167, 1984.

[9] C. W. Xu and Y. Z. Lu, "Fuzzy model identification and self-learning for dynamic systems", *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-17, pp. 683-689, July/Aug. 1987.

Tae-Chon Ahn
He was born in Inchon, Korea on October 11, 1955. He received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from the Yonsei University, Korea in 1978, 1980 and 1986, respectively. From 1980 to 1981, he was employed as researcher at Korea Power Co., Korea. From 1987 to 1988, he was a visiting scholar in the department of Control System

Engineering at Upscale University, Sweden. From 1996 to 1997, he was a visiting professor in the department of electrical and computer engineering at Georgia Institute of Technology, USA. Since 1981, he has been with the School of Electrical and Electronics Engineering at Wonkwang University, Korea, where he is currently a Professor. He is engaged in research on digital granular control, real-time data measuring and processing, FPGA design, intelligent systems, stability and control of synchronous machines.

Phone : +82-63-850-5319
Fax : +82-63-853-2196
E-mail : tcahn@wonkwang.ac.kr

Jong-Seok Lee

He was born in Iksan, Korea, in 1957. He received the B.S. and M.S. degrees in Electric Engineering of graduate school Wonkwang University, Iksan, Korea, in 1984 and 1988, respectively, and Ph.D. degree in Electric Engineering of graduate school, Wonkwang University, Iksan, Korea, in 2002. Since 1996, he has been with the Department of Radio-technology, College of the Wonkwang Health Science, Iksan, Chonbuk, Korea, where he is currently a Professor.

Phone : +82-63-840-1234
Fax : +82-63-840-1239
E-mail : jslee@wkhc.ac.kr

Kyung-Won Jang

He was born in Korea in 1974. He received the B.S. and M.S. degrees in control and instrument engineering from Wonkwang University, Iksan, Korea, in 2001 and 2003, respectively. He is currently working toward the Ph. D. degree in control and instrumentation engineering at the Wonkwang University, Iksan, Korea. His current research deals with the model based controller design, pattern recognition, granular computing and its application.

Phone : +82-63-850-5319
Fax : +82-63-853-2196
E-mail : jaang@wonkwang.ac.kr