

Hard 분산 분할 기반 추론 시스템을 이용한 비선형 공정 모델링

박건준*, 김용갑*

Nonlinear Process Modeling Using Hard Partition-based Inference System

Keon-Jun Park*, Yong-Kab Kim*

요약 본 논문에서는 Hard 분산 분할 방법을 이용하는 추론 시스템을 소개하고 비선형 공정을 모델링한다. 이를 위해 입력 공간을 분산 형태로 분할하고 소속 정도가 0 또는 1을 갖는 Hard 분할 방법을 이용한다. 제안한 방법은 C-Means 클러스터링 알고리즘에 의해 구현되며, 초기 중심값에 민감한 단점을 보완하기 위해 LBG 알고리즘을 적용하여 이진 분할에 의한 초기 중심값을 이용한다. Hard 분산 분할된 입력 공간은 규칙 기반의 시스템 모델링에서 규칙을 형성한다. 규칙의 전반부 파라미터는 C-Means 클러스터링 알고리즘에 의한 소속행렬로 결정된다. 규칙의 후반부는 다항식 함수의 형태로 표현되며, 각 규칙의 후반부 파라미터들은 표준 최소자승법에 의해 동정된다. 비선형 공정으로는 널리 이용되는 데이터를 이용하여 비선형 공정을 모델링한 후 특성을 평가한다.

Abstract In this paper, we introduce an inference system using hard scatter partition method and model the nonlinear process. To do this, we use the hard scatter partition method that partition the input space in the scatter form with the value of the membership degree of 0 or 1. The proposed method is implemented by C-Means clustering algorithm. and is used for the initial center values by means of binary split. by applying the LBG algorithm to compensate for shortcomings in the sensitive initial center value. Hard-scatter-partitioned input space forms the rules in the rule-based system modeling. The premise parameters of the rules are determined by membership matrix by means of C-Means clustering algorithm. The consequence part of the rules is expressed in the form of polynomial functions and the coefficient parameters of each rule are determined by the standard least-squares method. The data widely used in nonlinear process is used to model the nonlinear process and evaluate the characteristics of nonlinear process.

Key Words : Inference Systems, Hard Scatter Partition, C-Means Clustering Algorithm, LBG Algorithm, Nonlinear Process Modeling

1. 서론

복잡하고 비선형적인 특성을 갖는 시스템을 동정하고 해석하기 위해 많은 노력을 기울여 왔다. 특히, 1965년 Zadeh에 의해 처음으로 소개된

퍼지 집합 이론을 비선형적인 실 시스템에 적용함으로써 좋은 결실을 맺어왔으며 꾸준히 연구되고 있다[1-3]. 비선형적이고 다변수인 시스템을 대상으로 한 퍼지 모델링 기법의 유용성은 잘 알려져 있으며, 이들은 퍼지 추론 시스템에

* Corresponding Author : Department of Information Communication Professor of Wonkwang University (ykim@wku.ac.kr)
Received : November 03, 2014 Revised : November 20, 2014 Accepted : December 2, 2014

기초하고 있다. 초기 퍼지 모델의 동정연구로는 언어적 접근 방식[4, 5]과 퍼지 관계 방정식에 기초한 접근방식[6, 7]이 제안되었다. 언어적 접근방식에서, Tong은 논리적 조사 방법에 의해 가스로 공정을 동정하였고[8], Xu와 Zailu는 이 방법의 수정으로 더 좋은 결과를 얻는 방법과 결정 테이블에 기초한 자기 학습 알고리즘을 제안하였다. 이 알고리즘은 필요한 컴퓨터 용량 및 계산시간 때문에 고계다변수 시스템의 적용에 문제점을 발생시켰다[9, 10].

퍼지 모델링에서 퍼지 규칙의 생성은 일반적으로 입력 공간의 분할하는 방법과 소속 함수 및 소속 함수의 수에 의해 결정되며, 고차원의 비선형 공정을 모델링하는 것은 상당히 많은 규칙 수를 갖는 한계를 갖고 있다. 특히, 격자 분할 기반의 규칙 생성은 시스템을 표현하는 데 많은 어려움이 있으며 데이터 특성을 이용한 접근방식이 연구되고 있다.

본 논문에서는 C-Means 클러스터링 알고리즘[11]을 이용하여 입력 공간을 분산 형태로 Hard 분할하고 추론 방법에 의한 추론 시스템의 제안하여 비선형 공정에 대해 입출력 특성을 분석한다. 시스템의 규칙은 C-Means 클러스터링 알고리즘에 의해 입력 공간을 분산 형태로 분할하여 각각의 분할된 지역 공간이 하나의 규칙을 갖도록 형성한다. 전반부 파라미터의 동정에는 C-Means 클러스터링 알고리즘에 의한 소속 행렬에 의해 결정된다. 또한, C-Means 클러스터링 알고리즘의 초기 중심들에 민감한 특성을 보완하기 위하여 LBG (Linde-Buzo-Gray) 알고리즘[12]을 이용한다. 후반부 동정에서 퍼지 추론 방법은 간략추론, 선형추론, 2차식 추론 및 변형된 2차식 추론을 사용하며, 표준 최소자승법을 사용하여 후반부 파라미터를 동정한다. 또한, Box와 Jenkins가 사용한 가스로 공정 데이터[13]를 이용하여 비선형 공정을 모델링함으로써 특성 및 성능을 분석한다.

본 연구에서는 서론에 이어 제 2장에서는 제안한 추론 시스템의 전반부 및 후반부 동정에

대해 다루며, 제 3장에서는 비선형 공정을 모델링하여 제안한 시스템의 특성을 분석하고, 제 4장에서는 결론을 맺는다.

2. Hard 분할 기반 추론 시스템

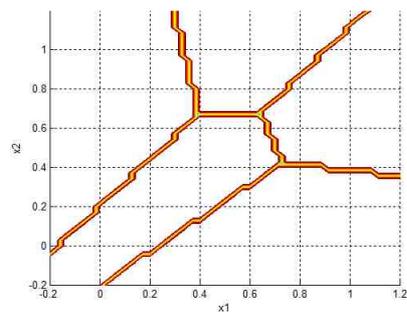
제안한 추론 시스템은 if-then 형식의 규칙을 이용하여 비선형 공정을 기술한다. 제안한 추론 시스템은 전반부 동정과 후반부 동정으로 나눠서 설명한다.

2.1 전반부 동정

모델 동정에서 규칙의 전반부 동정은 입력 공간의 분할 및 소속 함수의 결정 그리고 소속 함수의 파라미터 결정이 필요하다.

전반부 동정을 위해 일반적으로 입력 공간의 분할을 격자 형태로 분할하는 것을 생각할 수 있다. 입력 공간의 분할 방식은 상호 관계에 의한 분할 방식을 많이 사용하고 있으며, 분할된 지역 공간은 소속 함수로 정의되고, 규칙을 형성한다. 이러한 상호 공간에 의한 입력 공간의 분할 방식은 차원이 증가할수록 규칙 수가 지수적으로 증가하는 단점이 있다.

이러한 문제점을 극복하기 위해 입력 공간을 분산 형태로 분할하는 방식을 사용한다. 입력 공간을 분산 형태로 분할하는 방식은 C-Means 클러스터링 알고리즘[11]에 의해 수행된다. 이러한



[그림 1] 2차원 입력 공간의 분산 분할
[Fig. 1] Figure Scatter partition of 2-D input space

분산 분할 방식은 클러스터의 수만큼 입력 공간이 분할되고, 각 분할된 지역 공간은 규칙 수가 된다. 따라서 분산 형태의 입력 공간 분할 방식에서 규칙의 수는 클러스터의 수가 된다. 그림 1은 C-Means 클러스터링 알고리즘에 의해 2차원의 입력 공간이 분산 형태로 분할된 공간을 보여준다.

C-Means 클러스터링 알고리즘의 수행과정은 다음과 같다.

[단계 1] 클러스터의 개수 c 를 결정하고, 소속 행렬 U 를 초기화한다.

$$M_c = \{U \mid u_{ik} \in \{0,1\}, \sum_{i=1}^c u_{ik} = 1, 0 < \sum_{k=1}^n u_{ik} < n\} \quad (1)$$

여기서, u_{ik} ($i=1,2,\dots,c; k=1,2,\dots,n$)는 소속 행렬의 파라미터이다.

[단계 2] 각각의 클러스터에 대한 중심벡터 v_i 를 구한다.

$$v_i^{(r)} = \{v_{i1}, v_{i2}, \dots, v_{id}\}, v_{ij} = \frac{\sum_{k=1}^n u_{ik} \cdot x_{kj}}{\sum_{k=1}^n u_{ik}} \quad (2)$$

[단계 3] 각각의 클러스터 중심과 데이터와의 거리를 계산하여 새로운 소속행렬 $U(r)$ 을 생성한다.

$$d_{ik} = d(\mathbf{x}_k - \mathbf{v}_i) = \|\mathbf{x}_k - \mathbf{v}_i\| = \left[\sum_{j=1}^m (x_{kj} - v_{ij})^2 \right]^{1/2} \quad (3)$$

$$u_{ik}^{(r+1)} = \begin{cases} 1 & d_{ik}^{(r)} = \min\{d_{jk}^{(r)}\} \text{ for each } j \neq i \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

여기서, d_{ik} 는 k 번째 데이터 표본 x_k 와 i 번째 클러스터 중심 v_i 의 기하학적 거리이다.

[단계 4] 만일 식 (5)을 만족한다면 종료하고, 그렇지 않으면 $r = r+1$ 로 놓고 [단계 2]로 간다.

$$\|U^{(r+1)} - U^{(r)}\| \leq \epsilon (\text{tolerance level}) \quad (5)$$

C-Means 클러스터링 알고리즘은 초기 중심들의 선택에 민감한 특징을 나타내며, 이러한 문제점을 해결하기 위해 이진 분할과 C-Means 클러스터링 알고리즘을 결합한 LBG 알고리즘[12]을 이용한다. 이 방법은 C-Means 클러스터의 초기값을 랜덤으로 정하는 대신에 이진 분할로 구한 중심을 이용한다.

2.2 후반부 동정

후반부 구조로는 네 가지 추론 구조에 대해 다룬다. 또한, 최대 피벗팅 알고리즘을 가지는 가우스 소거법에 의한 표준 최소자승법을 이용하여 후반부 파라미터를 동정한다.

제안된 모델은 아래와 같은 네 가지 형태를 가지는 구현 규칙들로 구성된다.

(1) 구조 1 : 간략 추론법

$$R^j: \text{If } x_1 \text{ and } \dots \text{ and } x_d \text{ is } H_j \text{ then } y_j = a_{j0} \quad (6)$$

(2) 구조 2 : 선형 추론법

$$R^j: \text{If } x_1 \text{ and } \dots \text{ and } x_d \text{ is } H_j \text{ then } y_j = a_{j0} + \sum_{k=1}^d a_{jk} x_k \quad (7)$$

(3) 구조 3 : 2차식 추론법

$$R^j: \text{If } x_1 \text{ and } \dots \text{ and } x_d \text{ is } H_j \text{ then } y_j = a_{j0} + \sum_{k=1}^d a_{jk} x_k + \sum_{k=1}^d a_{j,d+k} x_k^2 + \sum_{k=1}^d \sum_{l=k+1}^d a_{j2} x_k x_l \quad (8)$$

(4) 구조 4 : 변형된 2차식 추론법

$$R^j: \text{If } x_1 \text{ and } \dots \text{ and } x_d \text{ is } H_j \quad (9)$$

$$\text{then } y_j = a_{j0} + \sum_{k=1}^d a_{jk}x_k + \sum_{k=1}^d \sum_{l=k+1}^d a_{jl}x_kx_l$$

간략 추론법은 후반부가 단일 상수항만을 가지며, 선형 추론법은 후반부가 일차 선형식으로 표현된다. 2차식 추론법 후반부 구조가 2차식 함수의 다항식 형태로 표현되며, 변형된 2차식 추론법은 입력 변수의 2차항이 생략된 2차식 추론법의 변형된 형태이다.

각 규칙에서 R_j 는 $j(j=1,2,\dots,n)$ 번째 규칙, $x_k(k=1,2,\dots,d)$ 는 입력 변수, H_j 는 C-Means 클러스터링 알고리즘에 의한 j 번째 규칙의 소속 정도이다. a_{jk} 는 후반부 파라미터이다.

각 규칙의 전반부 적합도 w_{jp} 는 C-Means 클러스터링 알고리즘에 의한 소속 행렬에 의해 얻어지며 다음과 같다.

$$w_{jp} = u_{ip} \tag{10}$$

따라서, 간략 추론에 의한 추론된 값 \hat{y} 는 가중 평균에 의해 다음과 같다.

$$\hat{y} = \frac{\sum_{j=1}^n w_{jp} y_j}{\sum_{j=1}^n w_{jp}} = \sum_{j=1}^n w_{jp} a_{j0} \tag{11}$$

여기서, 각 규칙의 적합도의 합은 1이다.

후반부 파라미터 동정에서 전반부 입력변수 및 파라미터가 주어지면, 성능지수를 최소화하는 최적 후반부 파라미터를 결정할 수 있다.

후반부의 파라미터는 a_j 로써 입출력 데이터가 주어졌을 때 최소자승법에 의해 구해진다. 최소자승법에 의한 후반부 파라미터의 동정은 식 (12)에 의해 구해진다.

$$\hat{a} = (X^T X)^{-1} X^T Y \tag{12}$$

선형 추론법, 2차식 추론법 및 변형된 2차식

추론법은 앞서 설명한 간략 추론법과 같은 방법으로 최종 추론값을 구할 수 있으며, 후반부 파라미터 또한 간략 추론법과 같은 방법으로 최소자승법에 의해 구해진다.

3. 비선형 공정 모델링 및 분석

본 논문에서는 널리 사용되고 있는 비선형 공정인 가스로 공정[13]을 사용하여 제안한 방법을 이용하여 비선형 공정을 모델링한다. 모델의 평가 기준인 성능지수는 가스로 공정에 대해서 MSE (Mean Squared Error)를 이용한다.

$$PI = \frac{1}{m} \sum_{p=1}^m (y_p - \hat{y}_p)^2 \tag{13}$$

Box와 Jenkins가 사용한 가스로 시계열 데이터를 이용하여, 입출력 데이터인 가스 흐름률과 연소된 이산화탄소 농도의 가스로 공정을 퍼지 모델링한다. 입력이 가스 흐름률이고 출력이 이산화탄소 농도인 1입력 1출력의 가스로 시계열 입출력 데이터 296쌍을 모의실험을 위해 대표적으로 사용되는 $u(t-3)$ 과 $y(t-1)$ 을 입력으로 사용하고, $y(t)$ 를 출력으로 사용하여 2입력 1출력 시스템을 구성하여 2차원 시스템을 모델링한다. 또한 데이터 집합은 학습과 테스트 데이터로 나누어 수행한다.

[표 1] 2입력 시스템에 대한 성능지수
[Table 1] Performance index for 2 input system
(a) 간략 추론법

Rule	PI	E.PI
2	3.459	4.104
3	1.698	2.165
4	1.015	1.973
5	0.943	2.111
6	0.904	2.077
7	0.867	1.720
8	0.838	1.875
9	0.715	1.310
10	0.641	1.232

(b) 선형 추론법

Rule	PI	E_PI
2	0.022	0.335
3	0.021	0.348
4	0.021	0.342
5	0.019	0.293
6	0.019	0.294
7	0.019	0.303
8	0.017	0.293
9	0.015	0.302
10	0.015	0.305

(c) 2차식 추론법

Rule	PI	E_PI
2	0.022	0.337
3	0.020	0.336
4	0.019	0.350
5	0.016	0.284
6	0.015	0.284
7	0.014	0.293
8	0.014	0.286
9	0.013	0.297
10	0.012	0.302

(d) 변형된 2차식 추론법

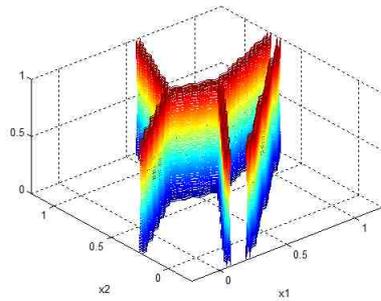
Rule	PI	E_PI
2	0.022	0.347
3	0.021	0.353
4	0.021	0.335
5	0.019	0.300
6	0.019	0.301
7	0.017	0.306
8	0.015	0.300
9	0.014	0.300
10	0.014	0.302

표 1은 규칙의 수 및 추론 방법에 의한 학습 데이터와 테스트 데이터에 대한 성능 지수를 보여준다. 여기서, PI는 학습 데이터에 대한 성능 지수를, E_PI는 테스트 데이터에 대한 성능 지수를 보여준다.

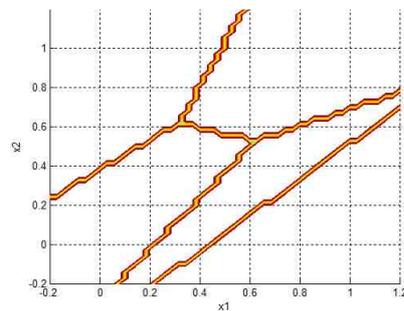
표 1로부터 간략 추론법 보다는 선형 추론법, 2차식 추론법, 변형된 2차식 추론법이 더 좋은 성능을 보여준다. 또한, 규칙 수가 많을수록 일반적으로 성능이 개선되는 것을 알 수 있다. 제안한 추론 모델은 입력 공간을 분산 형태로 5개의 규칙을 가지고 2차식 추론을 이용한 경우가 가장 좋은 성능을 보여준다. 이때의 성능지수

는 PI는 0.016이고 E_PI는 0.284이다.

그림 2는 2차원 입력 공간에서 Hard 분산 분할 추론 방법을 이용하여 클러스터의 수가 5개 일때의 소속 행렬 및 분할된 입력 공간을 보여 준다.



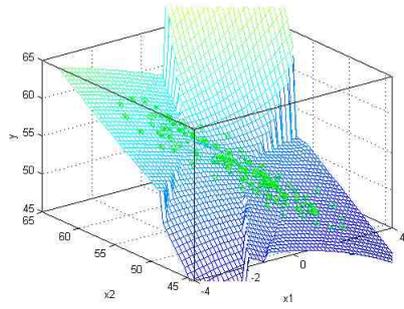
(a) 소속 행렬



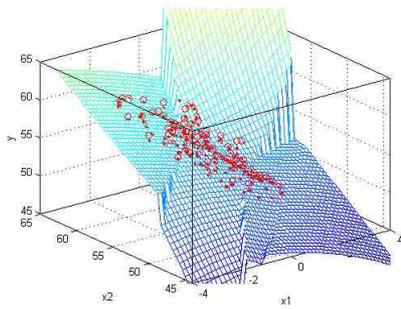
(b) 분산 분할

[그림 2] 소속 행렬 및 분산 분할 (클러스터 5개)
[Fig. 2] Membership matrix and scatter partition (5 clusters)

그림 3은 는 표 1에서 가장 좋은 성능을 보이는 클러스터의 수가 5개이고 2차식 추론 방법에 의한 입출력 특성 평면을 보여준다. 여기서, 원은 원 출력을, 점은 모델 출력을 보여준다. 각 분할된 지역 공간은 2차식 입출력 특성 보여준다.



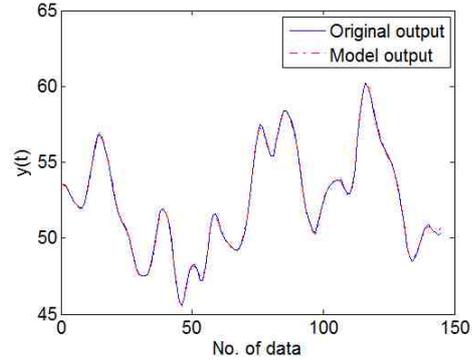
(a) training data



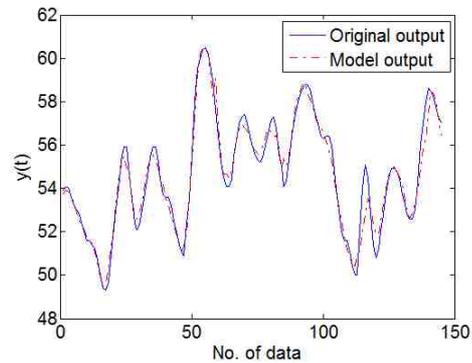
(b) testing data

[그림 3] 입출력 공간 평면 (클러스터 5개, 2차식추론)
 [Fig. 3] Input-output space plane (5 clusters, Quadratic reasoning)

그림 4는 표 1에서 선택된 모델인 5개의 규칙 수를 가지고 후반부 구조가 2차식 추론인 모델에 대하여 학습 데이터와 테스트 데이터에 대한 원 출력과 모델출력을 보여준다. 그림에서 보는 바와 같이 원 출력과 유사한 출력을 갖는 것을 알 수 있다.



(a) training data



(b) testing data

[그림 4] 모델 출력 (클러스터 5개, 2차식 추론)
 [Fig. 4] Model output (5 clusters, Quadratic reasoning)

표 2는 초기 중심값을 다르게 설정하여 20번 반복 실험한 결과를 보여준다. 후반부 구조는 2차식 추론법을 사용하였다. 표 2(a)는 C-Means 클러스터링 알고리즘만을 이용하여 20번 반복 실험하였으며, 표 2(b) 이진 분할을 이용하여 초기 중심값을 설정한 후 C-Means 클러스터링 알고리즘을 20번 반복 실험한 결과를 보여준다. 표 2로부터 이진 분할을 이용하여 초기 중심값을 설정한 방법이 더욱 견고한 결과를 나타내는 것을 알 수 있다.

[표 1] 성능지수 평균과 표준편과 (20번 반복)
 [Table 1] Average and standard deviation of Performance index (20 times)

(a) C-Means clustering algorithm

Rule	Average		STD	
	PI	E_PI	PI	E_PI
2	0.021	0.334	0.00011	0.00267
3	0.020	0.337	0.00028	0.00128
4	0.018	0.332	0.00080	0.01307
5	0.017	0.342	0.00101	0.02714
6	0.016	0.330	0.00094	0.02455
7	0.014	0.312	0.00114	0.01695
8	0.014	0.311	0.00100	0.01699
9	0.013	0.376	0.00073	0.11528
10	0.012	0.336	0.00061	0.05218

(b) LBG algorithm

Rule	Average		STD	
	PI	E_PI	PI	E_PI
2	0.021	0.335	0.00008	0.00305
3	0.020	0.337	0.00010	0.00161
4	0.018	0.357	0.00026	0.00889
5	0.016	0.313	0.00056	0.03164
6	0.015	0.309	0.00073	0.02173
7	0.014	0.317	0.00108	0.01636
8	0.013	0.315	0.00022	0.00773
9	0.012	0.301	0.00062	0.01067
10	0.012	0.310	0.00073	0.02170

4. 결론

본 논문에서는 입력 공간을 Hard 분산 분할 방법을 이용하여 추론하는 시스템을 소개하였으며 비선형 공정을 모델링하였다. 제안된 추론 시스템은 입력 공간을 Hard 분산 형태로 분할하고 분할된 지역 공간은 규칙을 형성하였으며, 각각의 지역 공간은 간략 추론, 선형 추론, 2차식 추론 및 변형된 2차식 추론을 이용하여 표현하였다.

비선형 공정을 모델링하였으며 입력 공간을 분산 형태로 5개의 규칙을 가지고 2차식 추론을 이용한 경우가 가장 좋은 성능을 보여주었다. 또한, 초기 중심값을 다르게 설정하여 20번 반복 실험하여 비교 분석하였다. 이진 분할을 이용하

여 초기 중심값을 설정한 방법이 더욱 견고한 결과를 나타내는 것을 알 수 있다.

Hard 분할에 의해 소속 정도는 0 또는 1의 값을 갖지만 실험을 통해 만족할 만한 성능을 얻을 수 있었으며, 제안한 방법으로 고차원의 시스템 동정에 적용가능할 것으로 사료된다.

감사의 글

본 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2013R1A1A2011835)

References

- [1] J.S.R. Jang, E. Mizutani, C.T. Sun, Neuro-Fuzzy and Soft Computing, A Computational Approach to Learning and Machine Intelligence. Prentice Hall, NJ, 1997.
- [2] K.-J. Park, J.-K. Kang, Y.-K. Kim, "Fuzzy Inference Systems Based on FCM Clustering Algorithm for Nonlinear Process," KIIECT, 5(4), pp. 224-231, 2012.
- [3] C.-S. Bae, H.-Y. Kim, T.-W. Kim, Y.-S. Kang, S.-K. Hwang and S.-K. Lee, "Implementation of Smart car using Fuzzy Rules," KIIECT, 5(2), pp. 81-88, 2012
- [4] R.M. Tong, "Synthesis of fuzzy models for industrial processes," Int. J. Gen. Syst., 4, pp. 143-162, 1978.
- [5] W. Pedrycz, "An identification algorithm in fuzzy relational system," Fuzzy Sets Syst., 13, pp. 153-167, 1984.
- [6] W. Pedrycz, "Numerical and application aspects of fuzzy relational equations," Fuzzy Sets Syst., 11, pp. 1-18, 1983.
- [7] E. Czogola and W. Pedrycz, "On identification in fuzzy systems and its applications in control problems," Fuzzy

Sets Syst., 6, pp. 73-83, 1981.

[8] R. M. Tong, "The evaluation of fuzzy models derived from experimental data," Fuzzy Sets Syst., 13, pp. 1-12, 1980.

[9] C. W. Xu, "Fuzzy system identification," IEEE Proceeding, 126(4), pp. 146-150, 1989.

[10] C. W. Xu and Y. Zailu, "Fuzzy model identification self-learning for dynamic system," IEEE Trans. on Syst. Man, Cybern., SMC-17(4), pp. 683-689, 1987.

[11] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, PlenumPress, NewYork, 1981.

[12] Y. Linde, A. Buzo, R. M. Gray, An Algorithm for Vector Quantizer Design. IEEE Transactions on Communications, 28, 702-710, 1980.

[13] Box and Jenkins, Time Series Analysis, Forecasting and Control, Holden Day, SanFrancisco, CA.

저자약력

박 건 준(Keon-Jun Park) [정회원]



- 2003년 2월 : 원광대학교 전기전자공학부 졸업 (공학사)
- 2005년 2월 : 원광대학교 제어계측공학과 졸업 (공학석사)
- 2010년 8월 : 수원대학교 전기공학과 (공학박사)
- 2010년 ~ 2012년 : 원광대학교 공과대학 POST-BK2 1 Post-Doc
- 2012년 ~ 현재 : 원광대학교 리서치펠로우 연구교수

<관심분야> 컴퓨터 및 인공지능, 지능시스템 및 제어, 상황인식 컴퓨팅

김 용 갑(Yong-Kab Kim) [중신회원]



- 1988년 2월 : 아주대학교 전자공학과 (공학사)
- 1993년 2월 : 엘라마 주립대 (공학석사)
- 2000년 2월 : 노스캐롤라이나 주립대 전기·컴퓨터공학과 (공학박사)
- 2003년 ~ 현재 : 원광대학교 정보통신공학과 정교수
- 2012년 ~ 현재 : LED 인력양성사업단장(진북)
- 2012년 ~ 현재 : 원광대학교 IT·BT창업보육센터장

<관심분야> 가시광통신시스템, 전력선통신