

# 강화된 유전알고리즘을 이용한 이중 동조 기반 퍼지 예측시스템 설계 및 응용

논문

59-1-32

## Design of Fuzzy Prediction System based on Dual Tuning using Enhanced Genetic Algorithms

방영근\* · 이철희†

(Young-Keun Bang · Chul-Heui Lee)

**Abstract** - Many researchers have been considering genetic algorithms to system optimization problems. Especially, real-coded genetic algorithms are very effective techniques because they are simpler in coding procedures than binary-coded genetic algorithms and can reduce extra works that increase the length of chromosome for wide search space. Thus, this paper presents a fuzzy system design technique to improve the performance of the fuzzy system. The proposed system consists of two procedures. The primary tuning procedure coarsely tunes fuzzy sets of the system using the k-means clustering algorithm of which the structure is very simple, and then the secondary tuning procedure finely tunes the fuzzy sets using enhanced real-coded genetic algorithms based on the primary procedure. In addition, this paper constructs multiple fuzzy systems using a data preprocessing procedure which is contrived for reflecting various characteristics of nonlinear data. Finally, the proposed fuzzy system is applied to the field of time series prediction and the effectiveness of the proposed techniques are verified by simulations of typical time series examples.

**Key Words** : Fuzzy system, Optimization, real-coded genetic algorithms, Data preprocessing, Time series prediction

### 1. 서론

유전알고리즘은 좋은 정보와 나쁜 정보를 판단하고 이러한 정보를 개량 및 도태 등, 다양한 진화원리를 토대로 최적의 해를 탐색하는 방법으로 다양한 분야에서 시스템 최적화 기법으로 사용되어 왔다. 특히, 비선형적 특성의 시스템에서 그들은 시스템 최적화의 좋은 수단이 되었다. 이러한 유전알고리즘은 soft computing 기법들의 성능향상에도 많은 도움을 주고 있으며, 퍼지이론과 결합하여 다양한 분야에서 그들의 성능을 입증하고 있다. 하지만 기존에 유전알고리즘의 코딩기법으로 많이 사용되는 이진코딩(binary coding)은 넓은 공간과 정밀한 해의 탐색을 위해선 염색체(chromosome)의 길이를 증가시켜야 하는 문제와 더불어 부호화와 복호화를 거치는 절차의 복잡성을 초래했다 [1-3]. 이러한 이진코딩의 제약을 극복하고자 제안된 실수코딩(real coding)은 염색체와 탐색하고자 하는 해가 정확히 일대일 대응관계에 있고, 부호화와 복호화 과정이 필요 없으므로 넓은 공간의 탐색과 정밀한 해의 탐색을 가능하게 하였다 [4]. 또한, 다양한 유전연산자들의 개발로 수렴시간의 문제나 국부해로의 수렴 등과 같은 문제에 있어 많은 부분 그 성능이 강화되고 있다. 이러한 다양한 유전알고리즘은 퍼지시스템의 동조를 위해 많이 사용되고 있으며, 본 논문 또한 유전알고리즘을 이용하여 설계된 퍼지시스템의 최적화를 유도하였다.

먼저, 본 논문에서는 비선형 데이터의 원형을 사용하지 않고 그들로부터 생성되는 차분 데이터들 중 원형데이터의 특성을 잘 반영할 수 있는 일부를 추출하여 이를 통해 다중 퍼지시스템을 구현함으로써, 데이터 이면에 내재된 다양한 특성들이 충분히 시스템에 적용될 수 있도록 하였다 [5-6]. 다중 퍼지 시스템의 구현에서는 1차적으로 k-means 클러스터링 알고리즘을 이용하여 퍼지시스템의 규칙기반을 동조한다. 그 후, 동조된 규칙기반들은 2차 동조 과정에서 개선된 성능의 유전알고리즘에 의해 세부동조 과정을 거치면서 그들의 적합성이 더욱 개선될 수 있도록 하였다. 제안된 논문의 동조 과정은 1차적으로 구조가 간단한 k-means 클러스터링을 이용하여 시스템 구조의 복잡성을 줄였고, 이를 기반으로 2차 동조 과정에선 실수코딩 유전알고리즘(real-coded genetic algorithms)을 이용함으로써, 넓은 탐색공간에 대한 부담과 수렴시간에 대한 문제를 해결하였다. 또한, 개선된 다양한 유전연산자들을 이용하여 보다 정밀하게 최적 해를 탐색할 수 있도록 함으로써 제안된 시스템의 성능을 강화하였다. 마지막으로 제안된 기법들을 비선형 시계열 예측분야에 적용하고, 다양한 시계열의 예측을 통해 효용성을 검증하였다.

### 2. 제안된 퍼지 시스템의 구조

일반적으로 퍼지 예측 시스템은 퍼지화기, 규칙기반, 추론 엔진, 본 논문의 구조는 크게 3개의 부분으로 구현될 수 있다. 먼저 원형 데이터의 특성을 잘 들어낼 수 있는 일부의 차분 데이터들을 생성하는 데이터의 전처리 과정과 생성된 차분 데이터를 이용하여 다중 퍼지 시스템을 구현하는 1차

† 교신저자, 정회원 : 강원대 전기전자공학부 교수 · 공박

E-mail : chlee@kangwon.ac.kr

\* 정 회원 : 강원대 전기전자공학과 박사과정

접수일자 : 2009년 8월 5일

완료일자 : 2009년 11월 18일

동조과정, 그리고 1차 동조된 퍼지시스템을 유전알고리즘을 통하여 세부 동조하는 2차 동조과정으로 분류할 수 있다. 아래의 그림 1은 제안된 퍼지시스템의 전체 순서도를 보여준다.

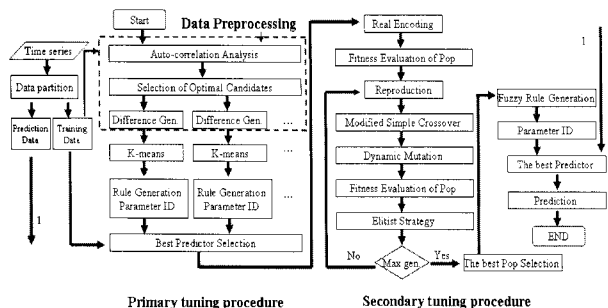


그림 1 제안된 시스템의 전체 순서도  
Fig. 1 The total flow chart of the proposed system

먼저, 제안된 시스템은 시스템의 훈련을 위해 사전 정의된 길이의 훈련데이터를 이용하여, 데이터의 전처리과정을 통해 원형데이터의 다양한 특성들을 잘 반영하는 최적의 차분데이터를 생성하게 된다. 그 다음 생성된 차분데이터들의 패턴이나 규칙성 등을 이용하여 그들 각각에 상응하는 퍼지시스템을 구현함으로써, 시스템이 충분히 비선형 데이터의 유동적인 특성을 고려할 수 있도록 하였다. 퍼지시스템의 구현에는 TSK 퍼지모델을 사용하였고, 각 시스템의 규칙을 동조하기 위해, 그 구조가 간단하면서도 분류능력이 좋은 k-means 클러스터링 알고리즘을 1차적으로 사용하였다. 이렇게 구현된 다중 퍼지시스템은 정의된 훈련데이터로 성능을 검증한 후, 2차 세부 동조를 위한 하나의 시스템만이 선택되게 되며, 선택된 퍼지시스템의 규칙기반은 개선된 유전알고리즘에 의해 세부 동조됨으로써 적합성을 높일 수 있도록 하였다. 2차 동조 과정에 사용된 코딩기법에는 1차 동조과정에서의 사전 지식을 기반으로 한 실수코딩 기법을 사용하여, 탐색공간에 대한 부담을 줄이면서도 정밀한 해의 탐색을 가능하게 하였으며, 재생산, 교배, 돌연변이와 같은 유전연산자들에는 시스템의 성능을 보다 효과적으로 개선할 수 있는 강화된 기법들을 사용하여 시스템이 비선형 데이터의 특성에 더욱 적합하게 동작할 수 있도록 하였다. 따라서 제안된 퍼지시스템은 하나의 데이터 전처리 과정과 2개의 시스템 동조과정으로 구현될 수 있다.

### 3. 비선형 데이터의 전처리

데이터의 전처리 과정은 사전 정의된 훈련데이터를 이용하여 원형데이터의 특성을 잘 내포하고 있는 차분 데이터를 생성하는 과정이다. 길이  $N$ 의 훈련 데이터로부터 생성 가능한 차분데이터들의 수는  $N-1$ 개가 될 것이며, 이들 모두가 원형데이터의 유동적 특성을 잘 내포하고 있다고 보기는 힘들 것이며, 또한 이들 모두를 위한 다중 퍼지시스템을 구현할 경우 시간적으로나 구조적으로 많은 문제점이 야기될 것이다. 따라서 본 논문에서는 다음과 같이 데이터의 전처리 방법을 이용하여 소수이지만, 원형 데이터의 패턴이나 규칙

성들을 잘 반영할 수 있는 차분데이터들만을 생성하여 퍼지시스템 설계에 사용한다.

#### 1 단계)

자기상관 계수에 따른 원형데이터와 차분데이터 사이의 상관성을 판별한다.

$$r_j = \frac{1}{N-j} \sum_{i=1}^{N-j} (y(i) - \bar{y})(y(i+j) - \bar{y})}{\frac{1}{N} \sum_{i=1}^N (y(i) - \bar{y})^2} \quad (1)$$

여기서,  $N$ 은 훈련데이터의 길이이고,  $j$ 는 차분 간격 값이다. 또한,  $y(i)$ 는  $i$ 번째 훈련데이터이며,  $\bar{y}$ 는 훈련데이터의 평균이다.

#### 2단계)

자기 상관계수를 순위에 따라 정렬한다.

#### 3단계)

0이하의 상관계수에 해당되는 차분간격 값은 제거된다.

#### 4단계)

차분데이터의 생성을 위한 차분간격 값은 1차적으로 상위 5개의 상관 값에 상응하는 차분간격 값을 선택한다. 이는 최소 그들에 상응하는 5개 이상의 퍼지 시스템을 설계하여 비선형 데이터의 유동적 특성을 충분히 시스템에 반영하기 위함이다.

#### 5단계)

선택된 차분간격 값을 제외한 나머지 차분간격 값들의 상관계수 값들에 대하여 인접한 두 상관계수 값의 차를 계산한다.

$$CH_s = (cof_s - cof_{s+1}), \text{ where } s = [1, N-7] \quad (2)$$

여기서,  $cof$ 는 (1)에 의해 계산된 상관계수 값들이고  $s$ 는 차 연산을 위한 총 수행 길이를 의미한다.

#### 6단계)

차 연산 값이 가장 크게 나타는 구간의 상관계수 값을 임계치로 정의하고, 이 보다 큰 상관계수 값에 상응하는 차분간격 값을 2차적으로 차분데이터를 생성하기 위한 차분간격 값으로 선택한다.

#### 7단계)

이렇게 선별된 차분간격 값에 따라 각각의 차분데이터들을 다음에 의해 생성한다.

$$\begin{aligned} \Delta_{m(i)}t_1 &= y(N) - y(N-m(i)) \\ \Delta_{m(i)}t_2 &= y(N-1) - y(N-m(i)-1) \\ &\vdots \\ \Delta_{m(i)}t_n &= y(N-n-1) - y(N-m(i)-n-1) \\ &\vdots \\ \Delta_{m(i)}t_{N-m(i)} &= y(m(i)+1) - y(1) \end{aligned} \quad (3)$$

여기서,  $m(i)$ 는 각각의 선택된 차분간격 값들의  $i$ 번째 값이며,  $\Delta_{m(i)}t_n$ 은 생성된  $n$ 번째 차분데이터이다. (3)에 의해 생성된 차분데이터들은 다중 퍼지시스템의 구현을 위해 각각 사용된다. 위의 방법을 통해 생성된 차분데이터들의 수

는 최소 6개 이상이 된다. 이는 상관계수 값이 크다고 해서 생성된 차분데이터가 원형데이터의 특성을 전부 반영할 수 있는 것은 아니기 때문이며, 원형데이터의 특성을 충분히 고려할 수 있는 퍼지시스템을 구현하기 위함이다.

4. 다중 퍼지시스템 설계 및 1차 동조

본 논문에서는 비교적 적은 입력과 퍼지분할로도 적합한 언어적 규칙기반을 구현할 수 있으며, 후반부 선형 식을 통해 고전 선형 회귀모델의 이점도 취할 수 있는 TSK 퍼지모델을 이용하여 다중 퍼지시스템을 설계 하였다. TSK 퍼지모델의 언어적 규칙의 일반식은 다음과 같이 정의된다.

$$R: \text{If } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ and } \dots \text{ and } x_n \text{ is } A_n \quad (4)$$

$$\text{Then } y = p_0 + p_1 x_1 + p_2 x_2 + \dots + p_n x_n$$

TSK 퍼지모델은 언어적 규칙을 표현하는 전반부와 규칙에 따른 동작을 제어하는 후반부로 구성되어 있다. 전반부의 규칙을 생성하기 위해서는 입력공간의 퍼지분할이 필요하며, 규칙의 출력을 위해서는 파라미터 식별이 필요하다. 본 논문은 입력공간의 퍼지분할 과정에서 1차 동조를 위해 k-means 클러스터링 기법을 적용하였으며, 파라미터 추정에는 최소자승법 (least square method)을 사용하였다. 또한, 전반부의 언어적 규칙의 표현과 후반부의 파라미터 추정을 위해 4개의 연속된 차분데이터를 하나의 입력데이터 쌍으로 사용하였다. 따라서 각각의 차분데이터들에 의해 생성될 수 있는 입력데이터 쌍은 아래와 같다.

$$Sets = \{ \Delta_{m(i)} t_{k+1}, \Delta_{m(i)} t_{k+2}, \Delta_{m(i)} t_{k+3}, \Delta_{m(i)} t_k \} \quad (5)$$

where,  $k = [1, N - m(i) - 3]$

여기서,  $N - m(i) - 3$ 은 총 생성될 수 있는 입력데이터 쌍의 개수이다. 또한, 수식(5)의  $\Delta_{m(i)} t_k$ 는 파라미터 추정을 위해 사용되는 출력 값을 의미한다. 따라서 차분 간격  $m(i)$ 에 대한 TSK 퍼지모델의  $j$ 번째 규칙  $R_j$ 는 수식 (6)과 같이 다시 정의된다.

$$\text{Rule } (R_j) \quad (6)$$

$$\text{If } \Delta_{m(i)}^j t_{k+1} \text{ is } A_j \text{ and } \Delta_{m(i)}^j t_{k+2} \text{ is } B_j \text{ and } \Delta_{m(i)}^j t_{k+3} \text{ is } C_j$$

$$\text{Then } \Delta_{m(i)}^j t_k = p_0^j + p_1^j \Delta_{m(i)}^j t_{k+1} + p_2^j \Delta_{m(i)}^j t_{k+2} + p_3^j \Delta_{m(i)}^j t_{k+3}$$

4.1 k-means 클러스터링을 이용한 규칙 기반의 동조

다중 퍼지시스템의 규칙기반은 차분간격  $m(i)$ 에 대하여 생성된 차분데이터들 중 최소값과 최대값 사이를 퍼지분할의 전체영역(universe of discourse)으로 하고, 입력데이터들에 대하여 k-means 클러스터링 알고리즘을 이용하여 그림 2와 같이 삼각형 소속함수와 사다리꼴 소속함수를 이용하여 NB, NS, ZE, PS, PB의 5개의 퍼지집합으로 분할하였다. 그림 2와 같이, 동조된 퍼지집합에 대한 각각의 입력데이터들이 갖는 소속정도는 입력데이터와 그들이 만족하는 퍼지 집합의 중심값을 이용하여 수식 (7)과 같이 구할 수 있다.

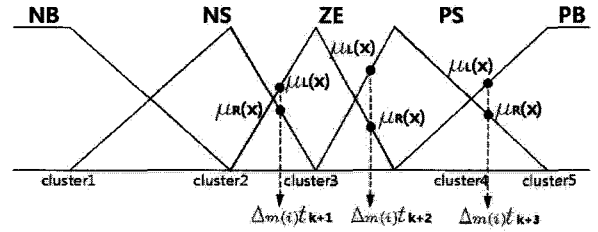


그림 2 퍼지분할 및 소속함수  
Fig. 2 Fuzzy partitioning and membership function

$$\text{If } \Delta_{m(i)} t \in \text{cluster}(1) \quad (7)$$

$$\mu_L(\Delta_{m(i)} t) = 1$$

$$\mu_R(\Delta_{m(i)} t) = \frac{\text{cluster}(z+1) - \Delta_{m(i)} t}{\text{cluster}(z+1) - \text{cluster}(z)}$$

$$\text{Elseif } \Delta_{m(i)} t \in \text{cluster}(z)$$

$$\mu_R(\Delta_{m(i)} t) = \frac{\text{cluster}(z+1) - \Delta_{m(i)} t}{\text{cluster}(z+1) - \text{cluster}(z)}$$

$$\mu_L(\Delta_{m(i)} t) = \frac{\Delta_{m(i)} t - \text{cluster}(z-1)}{\text{cluster}(z) - \text{cluster}(z-1)}$$

$$\text{Elseif } \Delta_{m(i)} t \in \text{cluster}(5)$$

$$\mu_R(\Delta_{m(i)} t) = 1$$

$$\mu_L(\Delta_{m(i)} t) = \frac{\Delta_{m(i)} t - \text{cluster}(z-1)}{\text{cluster}(z) - \text{cluster}(z-1)}$$

여기서,  $\text{cluster}(z)$ 는 입력 값  $\Delta_{m(i)} t$ 가 만족하는 클러스터를 의미하며,  $\mu_L(\Delta_{m(i)} t)$ 는 입력 값  $\Delta_{m(i)} t$ 가 만족하는 클러스터 중심 값으로부터 왼쪽부분의 소속 값이며,  $\mu_R(\Delta_{m(i)} t)$ 는 오른쪽의 소속함수 값을 의미한다. 또한 위 그림 2와 같이 입력데이터 쌍이 주어진다면, 그들에 의해 아래와 같이 8개의 퍼지규칙이 생성 가능할 것이다.

$$R_1: \Delta_{m(i)} t_{k+1} \text{ is } NS \text{ and } \Delta_{m(i)} t_{k+2} \text{ is } ZE \text{ and } \Delta_{m(i)} t_{k+3} \text{ is } PS \text{ then } \dots$$

$$R_2: \Delta_{m(i)} t_{k+1} \text{ is } NS \text{ and } \Delta_{m(i)} t_{k+2} \text{ is } ZE \text{ and } \Delta_{m(i)} t_{k+3} \text{ is } PB \text{ then } \dots$$

$$\vdots$$

$$R_7: \Delta_{m(i)} t_{k+1} \text{ is } ZE \text{ and } \Delta_{m(i)} t_{k+2} \text{ is } PS \text{ and } \Delta_{m(i)} t_{k+3} \text{ is } PS \text{ then } \dots$$

$$R_8: \Delta_{m(i)} t_{k+1} \text{ is } ZE \text{ and } \Delta_{m(i)} t_{k+2} \text{ is } PS \text{ and } \Delta_{m(i)} t_{k+3} \text{ is } PB \text{ then } \dots \quad (8)$$

각각의 퍼지시스템의 규칙기반에는 입력데이터 쌍이 생성하는 퍼지규칙만을 이용한다. 이는 TSK 퍼지모델의 경우, 3 입력 5퍼지집합에 따라 생성 가능한 125개의 규칙들 중 퍼지시스템의 학습을 위해 사용된 훈련데이터가 퍼지 규칙의 전반부를 만족하지 못한다면, 그 규칙의 후반부 파라미터는 최소자승법을 통해 식별될 수 없기 때문이다. 또한, 규칙의 생성과정에서 중복되는 규칙은 삭제되면서 각각의 퍼지 시스템의 규칙기반을 구현하게 된다.

4.2 TSK 퍼지 규칙의 파라미터 식별

퍼지규칙의 후반부 선형 수식의 파라미터 식은 전반부의 규칙을 만족하는 모든 입력데이터 쌍들을 이용하여 최소자승법을 적용하여 추정된다. 따라서 차분간격  $m(i)$ 에 대한 예측기의  $j$ 번째 규칙  $R_j$ 를 만족하는 입력 쌍이  $n$ 개라면, 이들에 의한 후반부 선형식은 수식(9)와 같이  $n$ 개의 연립방정식으로 표현될 수 있다.

$$\begin{aligned} \Delta_{m(i)}^j t_k^1 &= p_0^j + p_1^j \Delta_{m(i)}^j t_{k+1}^1 + p_2^j \Delta_{m(i)}^j t_{k+2}^1 + p_3^j \Delta_{m(i)}^j t_{k+3}^1 \\ \Delta_{m(i)}^j t_k^2 &= p_0^j + p_1^j \Delta_{m(i)}^j t_{k+1}^2 + p_2^j \Delta_{m(i)}^j t_{k+2}^2 + p_3^j \Delta_{m(i)}^j t_{k+3}^2 \\ &\vdots \\ \Delta_{m(i)}^j t_k^n &= p_0^j + p_1^j \Delta_{m(i)}^j t_{k+1}^n + p_2^j \Delta_{m(i)}^j t_{k+2}^n + p_3^j \Delta_{m(i)}^j t_{k+3}^n \end{aligned} \quad (9)$$

위의 연립 방정식은 아래와 같이 벡터 형으로 다시 표현될 수 있다.

$$\begin{bmatrix} \Delta_{m(i)}^j t_k^1 \\ \Delta_{m(i)}^j t_k^2 \\ \vdots \\ \Delta_{m(i)}^j t_k^n \end{bmatrix} = \begin{bmatrix} 1 & \Delta_{m(i)}^j t_{k+1}^1 & \Delta_{m(i)}^j t_{k+2}^1 & \Delta_{m(i)}^j t_{k+3}^1 \\ 1 & \Delta_{m(i)}^j t_{k+1}^2 & \Delta_{m(i)}^j t_{k+2}^2 & \Delta_{m(i)}^j t_{k+3}^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta_{m(i)}^j t_{k+1}^n & \Delta_{m(i)}^j t_{k+2}^n & \Delta_{m(i)}^j t_{k+3}^n \end{bmatrix} \begin{bmatrix} p_0^j \\ p_1^j \\ p_2^j \\ p_3^j \end{bmatrix} \quad (10.a)$$

$$\Delta_j = T_j P_j \quad (10.b)$$

여기서,  $\Delta_j$ 는 출력 벡터,  $T_j$ 는 입력 벡터,  $P_j$ 는 파라미터 벡터이다. 식 (10.b)의 파라미터 벡터는 최소자승법을 이용하여 다음과 같이 구할 수 있다.

$$\hat{P}_j = (T_j^T T_j)^{-1} T_j^T \Delta_j \quad (11)$$

식(11)에 의하여 추정된 파라미터 값들은 다음과 같이 정의되는 오차 파워의 합을 최소화하는 최적해가 된다.

$$E_j = (\Delta_j - T_j \hat{P}_j)^T (\Delta_j - T_j \hat{P}_j) \quad (12)$$

따라서 예측을 위한 새로운 입력데이터 쌍이 규칙  $R_j$ 를 만족한다면 예측 출력은 다음과 같이 얻을 수 있다.

$$\text{input} : X_k = [\Delta_{m(i)} x_{k+1}, \Delta_{m(i)} x_{k+2}, \Delta_{m(i)} x_{k+3}] \quad (13)$$

$$\text{output} : \hat{\Delta}_{m(i)}^j x_k = \hat{p}_0^j + \hat{p}_1^j \Delta_{m(i)} x_{k+1} + \hat{p}_2^j \Delta_{m(i)} x_{k+2} + \hat{p}_3^j \Delta_{m(i)} x_{k+3}$$

또한, 수식(13)의 새로운 입력데이터 쌍이 총  $q$ 개의 퍼지 규칙을 만족한다면, 입력데이터에 대한 전체 출력은 규칙  $R_j$ 의 전반부에서 결정되는 적합도  $\mu_j$ 와 수식(13)에 의해 결정된  $\hat{\Delta}_{m(i)} x_k$ 로부터 다음과 같은 가중합으로 구할 수 있다.

$$\hat{\Delta}_{m(i)} x_k = \frac{\sum_{j=1}^q \mu_j \hat{\Delta}_{m(i)}^j x_k}{\sum_{j=1}^q \mu_j} \quad (14)$$

퍼지 예측기의 출력  $\hat{\Delta}_{m(i)} x_k$ 는 시간  $t$ 에 대하여 예측된 미래 값의 증가분을 의미하므로 원하는 시계열의 최종 예측 값  $\hat{y}(t+p)$ 는 다음과 같이 구할 수 있다.

$$\hat{y}(t+p) = x(t) + \hat{\Delta}_{m(i)} x_k \quad (15)$$

여기서,  $p$ 는 1로써 한 스텝 전방 예측 값을 의미하며,  $x(t)$ 는  $t$ 시간에서의 현재 입력 값을 의미한다.

### 4.3 2차 동조를 위한 퍼지시스템 선택

구현된 다중 모델 퍼지 시스템들을 모두 2차 동조과정에서 유전알고리즘을 통해 세부 동조를 하게 된다면, 이는 시스템 구현의 시간과 구조에 대한 상당한 부담으로 작용할 수 있다. 따라서 본 논문에서는 적절한 성능평가 지수를 이용해 훈련데이터를 이용하여 다중 퍼지시스템 각각에 대한 성능을 검증한다. 그 후, 가장 만족스러운 결과를 나타내는 하나의 시스템만을 선택하여 2차 동조과정에서 세부 동조를 하고, 동조된 하나의 시스템은 최종 예측을 수행하게 된다. 따라서 이러한 시스템 선택 방법은 시스템 운영의 구조적, 시간적 부담을 충분히 완화할 수 있으며, 성능평가 지수로는 다음과 같이 주어지는 자승 오차 평균(MSE: squared error)을 이용하였고, 이를 최소화하는 시스템이 최종 예측 시스템으로 선택된다.

$$MSE = \frac{1}{N - m(i) - 3} \sum_{n=m(i)+3}^N (y(n) - \hat{y}(n))^2 \quad (16)$$

여기서  $\hat{y}(n)$ 은 퍼지시스템에서 출력한  $y(n)$ 에 대한 예측 값이며, 차분간격  $m(i)$ 인 퍼지 시스템에서 예측을 위해 사용되는 입·출력데이터 쌍의 개수가  $N - m(i) - 3$ 이므로 평균을 구할 수 있다.

## 5. 퍼지시스템의 2차 동조

2차 동조 과정은 1차 동조과정에서 선택된 최적의 퍼지시스템의 규칙기반을 더욱 세밀하게 동조하는 과정이다. 선택된 시스템의 1차 동조 과정에서 k-means 클러스터링에 의해 동조된 퍼지 규칙기반을 기반으로 하여 유전알고리즘의 적자생존의 법칙에 따라 더욱 세밀히 동조함으로써 시스템의 성능이 개선될 수 있도록 하였다. 또한 유전알고리즘의 성능을 더욱 향상시키기 위하여, 본 논문에서는 성능이 강화된 다양한 유전연산자를 적용하여 더욱 정밀한 동조가 이루어 질 수 있도록 하였다.

### 5.1 실수코딩에 의한 집단생성

유전알고리즘의 실행을 위해, 먼저, 1차 동조과정에서 탐색된 퍼지집합의 중심 값들로부터 유도되는 초기 집단을 생성하게 된다. 초기 집단의 생성을 위해 본 논문에서는 실수 코딩 기법을 사용하였으며 다음과 같다.  $g$ 세대( $g$ 번째 반복 횟수)에서의 집단  $Pop(g)$ 는 수식(17)과 같이 염색체로 특정지어지는  $L$ 개 개체들의 집합으로 정의될 수 있다.

$$Pop(g) = \{s_1(g), s_2(g), \dots, s_L(g)\} \quad (17)$$

여기서,  $s_i(g)$ 는  $i$ 번째의 염색체로 탐색 공간상의 한 점을 나타내며,  $L(>1)$ 은 집단의 크기를 의미한다. 또한, 수식(18)은  $i$ 번째 염색체에 대한 실수코딩의 한 예이며, 본 논문에서는 퍼지집합의 중심 값을 의미하므로  $z$ 는 5로 정의 된다.

$$s_i(g) = (v_{i1}(g) \ v_{i2}(g) \ \dots \ v_{iz}(g)) \quad (18)$$

본 논문에 사용된 실수코딩 기법은 기존에 많이 사용되는 이진코딩(binary encoding) 기법의 문제점을 개선한 것으로, 정밀도를 높이거나 탐색구간을 확대하기 위해 염색체의 길이를 증가할 필요가 없으며, 또한 부호화 과정과 복호화 과정이 없으므로 유전자와 퍼지집합의 중심 값들이 정확히 일대일의 실수 값으로 표현될 수 있으며, 상당히 넓은 영역을 표현할 수 있다.

5.2 목적함수와 재생산 연산자

본 논문에서는 유전 알고리즘의 진화과정의 지표가 되는 목적함수로 클러스터 내에 분류된 데이터들과 클러스터 중심과의 분산이 작을수록 데이터와 클러스터의 적합성이 높은 것으로 간주할 수 있기 때문에 (19)와 같이 전체 클러스터 내 분산(TWCV:total within-cluster validation)을 최소화 하는 방향으로 정의한다.

$$OF(s) = \sum_{c=1}^z \left( \frac{1}{h} \sum_{u=1}^h (x_u - v_c)^2 \right) \quad (19)$$

여기서,  $x_u$ 는  $v_c$ 를 중심 값으로 하는 클러스터에 포함된 데이터이며,  $h$ 는 포함된 데이터들의 개수이다. 또한 자손을 생성하기 위해 필요한 적합도 함수는 수식(20)과 같은 선형 스케일링 기법을 적용한다.

$$Fit(s) = OF(s) + C_b \quad (20)$$

재생산 연산자는 적합한 자손을 생성하기 위해 구하여진 적합도 함수를 이용하게 되며, 자손을 생성하기 위한 탐색점은 각 개체들의 적합도와 가장 우수한 개체의 적합도를 이용하여 수식(21)에 의하여 계산된다.

$$\bar{s}_{ic}(g+1) = s_{ic}(g) + \eta_h \frac{[Fit_{best}(g) - Fit_i(g)]}{Fit_{best}(g)} [s_{best(c)}(g) - s_{ic}(g)]$$

where,  $(1 \leq i \leq G, 1 \leq c \leq z)$  (21)

여기서,  $s_{ic}(g)$ 는  $g$ 세대의  $i$ 번째 개체의  $c$ 번째 요소를 의미하며,  $Fit_{best}(g)$ 는  $g$ 세대의 가장 우수한 적합도이며,  $Fit_i(g)$ 는  $g$ 세대 개체들의 각각의 적합도를 의미한다. 식(21)의 재생산 연산자의 목적은 약한 개체에는 최적 개체 쪽으로 더 많은 가중치를 부여하고 강한 개체에는 작은 가중치를 부여함으로써 최적 개체 쪽으로 유도되도록 하기 위한 것이다.

5.3 교배, 돌연변이, 엘리트전략

본 논문에서는 교배 연산자로 수정단순교배(modified simple crossover)를 사용하였다. 이 연산자는 이진코딩의 일점교배를 묘사한 연산자로 두 염색체에 대한 교배점이  $c \in [1, z-1]$ 인 구간에서 무작위로 선택되면, 식(22)에 의하여 일차 결합되고, 그 후 교배점을 기준으로 서로 교환되어 자

손을 생성하게 된다. 그림 3은 이 연산자의 연산 결과를 보여주는 한 예이다.

$$\begin{aligned} \tilde{v}_c^a &= \lambda \tilde{v}_c^b + (1-\lambda) \bar{v}_c^a \\ \tilde{v}_c^b &= \lambda \bar{v}_c^a + (1-\lambda) \tilde{v}_c^b \quad (1 \leq c \leq z) \end{aligned} \quad (22)$$

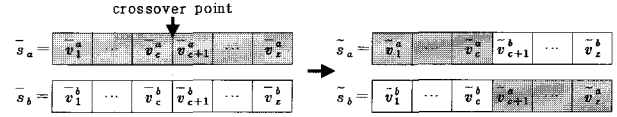


그림 3 수정단순교배의 예  
Fig. 3 Example of modified simple crossover

여기서,  $\lambda \in [0,1]$ 로 제한되는 값을 가진다. 이는 항상 적합한 형태의 자손을 생성하기 위한 것이다. 또한, 본 논문에서는 동적 돌연변이(dynamic mutation), 일명 불균등 돌연변이(non-uniform mutation)를 사용한다. 동적 돌연변이 연산자는 정밀도를 높이기 위하여 미세 조정이 가능하도록 고안된 것이며, 만약 그림 3의 자손 b의  $c$ 번째 유전자에서 돌연변이가 일어나면 그 돌연변이 값은 식(23)에 의해 정의된다.

$$\tilde{v}_c^b(mt) = \begin{cases} \tilde{v}_c^b + \Delta(w, \tilde{v}_c^b(up) - \tilde{v}_c^b), \tau = 1 \\ \tilde{v}_c^b - \Delta(w, \tilde{v}_c^b - \tilde{v}_c^b(low)), \tau = 0 \end{cases} \quad (23)$$

여기서,  $\tau$ 는 0과 1 둘 중 하나를 취하는 난수이며,  $\Delta(w, \beta)$ 로는 식(24)가 적용 된다.

$$\Delta(w, \beta) = \beta \psi \left( 1 - \frac{g}{G} \right)^\alpha \quad (24)$$

여기서,  $\psi$ 는 0과 1사이의 실수 난수이고,  $G$ 는 최대 세대 수,  $\alpha$ 는 불균등 정도를 나타내는 매개변수이다. 이 돌연변이 함수는 초기에는 전 공간을 균등한 확률로 탐색하다가 세대 수의 증가에 따라 지역적으로 탐색하는 것을 가능하게 하므로, 최적개체로의 수렴시간을 개선할 수 있다. 마지막으로 각 세대의 최적 개체가 재생산 연산자에 의해 다음세대의 자손의 생성에만 기여하고 소멸되는 것을 방지하기 위하여, 본 논문은 엘리트 전략(elitist strategy)을 적용한다. 이는 진화의 과정에서 이전 세대의 최적 개체를 기억하였다가 현세대에서 소멸된 것이 확인될 경우 현세대의 가장 약한 적합도의 개체와 대체함으로써 유전알고리즘의 성능을 더욱 개선할 수 있도록 고안된 것이다.

6. 시뮬레이션

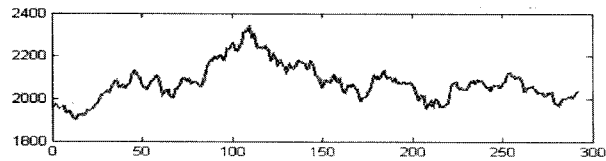
제안된 퍼지시스템의 성능을 검증하기 위하여 비정상 및 혼돈 비선형 특성을 갖는 대표적인 시계열 데이터를 선택하고, 이들을 이용하여 퍼지시스템을 모델링한 후, 예측을 통해 검증하였다. 아래의 표 1은 제안된 시스템에 사용된 파라미터 값들이다.

표 1 제안된 시스템의 변수

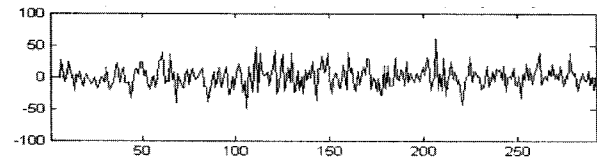
Table 1 Parameters of the proposed system

변수 명	변수 값
예측 스텝 ( $p$ )	1
최대 세대 반복수 ( $G$ )	300
초기 집단 수 ( $Pop$ )	30
교배 확률 ( $P_c$ )	0.9
돌연변이 확률 ( $P_m$ )	0.1
스케일링 상수 ( $C_f$ )	$\min(OF)$
세대의 가중치 ( $\eta$ )	1.7
불균등 정도 ( $\alpha$ )	5

첫 번째 비선형 데이터로는 호주의 다우지수 데이터[7]를 이용하였다. 총 292개의 데이터 중 200개의 데이터를 훈련 데이터로 사용하여 시스템을 설계한 후 나머지 데이터를 이용하여 성능을 검증하였다. 그림 4는 1차 동조 과정 후의 예측 결과를 보여준다.



(a) 1차 동조 후의 예측 결과  
(a) Prediction result after primary tuning

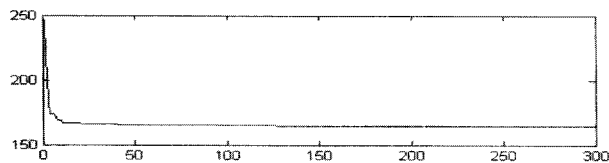


(b) 1차 동조 후의 예측 오차  
(b) Prediction error after primary tuning

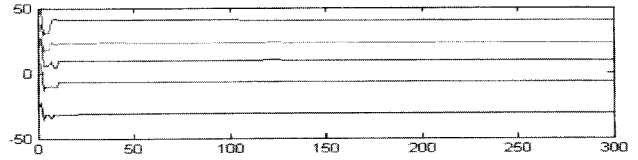
그림 4 1차 동조과정 후의 예측 결과

Fig. 4 Prediction results after primary tuning

그림 4에서, 검은 색은 원형데이터 값이고 파란색은 제안된 시스템의 예측 출력 값이다. 그림 4(b)의 오차 그림을 살펴보면, 제안된 시스템이 1차 동조과정 만으로도 좋은 동작 특성을 보임을 알 수 있다. 아래의 그림 5는 2차 동조 과정을 통해 목적함수와 퍼지 집합의 중심들이 최적화 되는 것을 보여 주며, 그림 6은 2차 동조 과정 후의 시스템의 최종 예측 결과를 보여 준다.



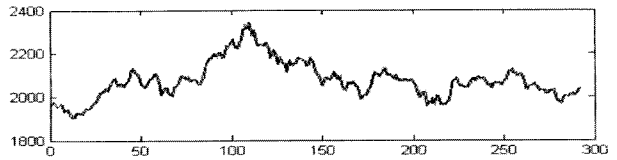
(a) 목적함수의 최적화  
(a) Optimization of objective function



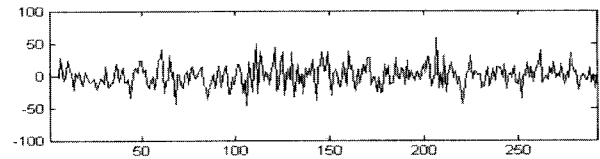
(b) 퍼지집합의 최적화  
(b) Optimization of fuzzy sets

그림 5 파라미터 최적화 결과

Fig. 5 Results of parameter optimization



(a) 2차 동조 후의 예측 결과  
(a) Prediction result after secondary tuning



(b) 2차 동조 후의 예측 오차  
(b) Prediction error after secondary tuning

그림 6 2차 동조과정 후의 예측 결과

Fig. 6 Prediction results after secondary tuning

1차 동조 후와 2차 동조 후의 성능을 좀 더 상세히 비교하기 위하여 다음과 같이 정의되는 MRE(mean relative error)를 이용하였으며, 표 2는 1,2차 동조 후의 성능을 비교한 것이다.

$$MRE = \frac{1}{N} \sum_{i=1}^N \left( \frac{|y(i) - \hat{y}(i)|}{|y(i)|} \times 100 \right) \% \quad (25)$$

표 2 1,2 차 동조 후의 성능 비교

Table 2 MRE comparison between primary and secondary tuning

MRE		최적 차분 간격 값
1차 동조 후	2차 동조 후	
0.5973	0.5868	1

표 2는 데이터 전처리 과정에서 생성된 차분 데이터들 중 차분간격 1에 상응하는 차분데이터를 이용한 퍼지시스템이 훈련구간에서 성능이 가장 좋았으며, 이렇게 선택된 시스템이 1차 동조 후 예측을 수행한 결과와 2차 동조 후 예측을 수행한 결과를 보여주는 것이다. 또한, 그림 5를 살펴보면 1차 동조 과정 후의 퍼지집합의 중심 값들은 2차 동조 과정에서의 유전알고리즘에 의해 보다 최적화 되어가는 모습을 볼 수 있으며, 이를 통해 시스템의 성능이 향상됨을 알 수 있다.

두 번째 시뮬레이션 데이터로는 호주의 전력생산량 데이터[7]를 이용하였다. 호주의 전력생산량 데이터는 총 155개의 데이터로 구성되며, 70개의 데이터를 훈련데이터로 하여 시스템을 모델링하고, 나머지 데이터들은 시스템의 성능을 검증하기 위하여 사용하였다. 아래의 그림 7은 1차 동조 과정을 기반으로 한 2차 동조 과정에서, 목적함수와 퍼지집합이 최적화되는 것을 보여주며, 그림 8은 제안된 시스템의 최종 예측결과를 보여 준다.

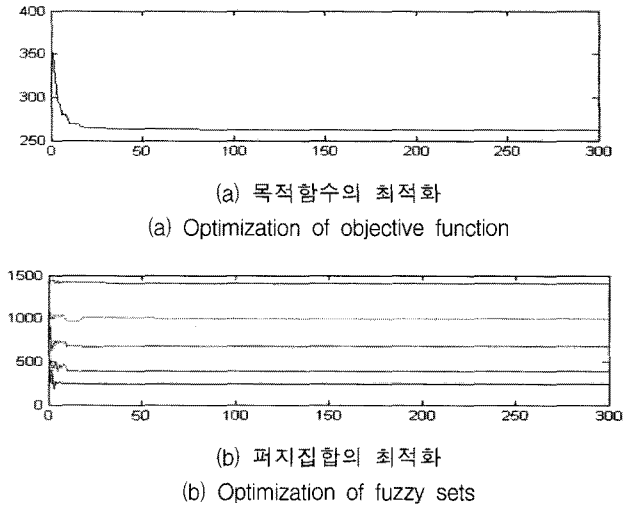


그림 7 파라미터 최적화 결과  
Fig. 7 Results of parameter optimization

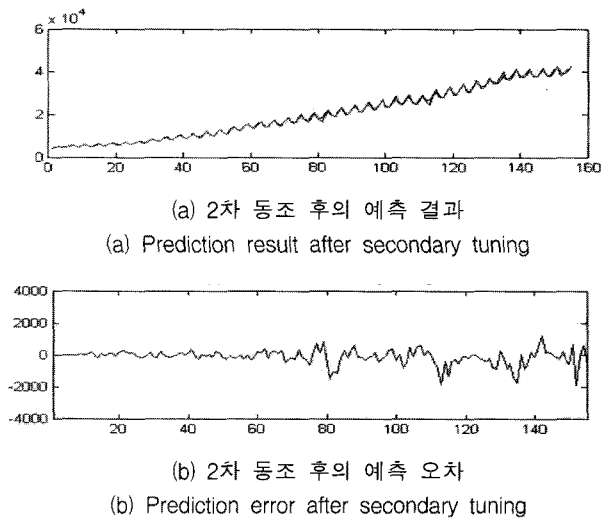


그림 8 2차 동조과정 후의 예측 결과  
Fig. 8 Prediction results after secondary tuning

아래의 표 3은 1,2차 동조 과정의 성능 비교표이며, 성능 평가 지수로는 다우지수와 같이 MRE를 사용하였다.

표 3 1,2 차 동조 후의 성능 비교  
Table 3 MRE comparison between primary and secondary tuning

MRE		최적 차분 간격 값
1차 동조 후	2차 동조 후	
1.7077	1.6336	4

첫 번째 시뮬레이션과 마찬가지로 2차 동조 후의 결과가 1차 동조 후의 결과보다 상당히 개선됨을 알 수 있다. 아래의 표 4는 제안된 퍼지시스템의 성능과 다른 기법들[2,5,6,8]을 비교한 것이다.

표 4 다른 방법들과의 성능 비교  
Table 4 Comparison of performances between the proposed method and other methods

모델 지수	Mamdani 퍼지모델	다중 퍼지 모델	Fuzzy AR	GS-RS 방식	proposed method
MRE	7.821	3.712	3.125	1.810	1.633

표 4에서 보이듯 제안된 시스템의 성능이 다른 시스템들의 성능에 비해 보다 우수함을 알 수 있다.

마지막 시뮬레이션 데이터로는 비선형 데이터로 많이 사용되는 Mackey-Glass 데이터로 식 (26)에 의해 생성할 수 있으며 생성된 데이터 중 1000개의 데이터를 추출하여 그 중 500개의 데이터를 시스템 모델링을 위해 사용하고 나머지 데이터를 시스템의 성능 검증을 위해 사용하였다.

$$\frac{dx(t)}{dt} = \frac{0.2x(t-17)}{1+x^{10}(t-17)} - 0.1x(t) \quad (26)$$

아래의 그림은 제안된 기법에 따라 설계된 퍼지시스템의 최종 출력을 보여 주며, 검은 색은 원형데이터이고 파란색이 시스템의 출력을 의미한다.

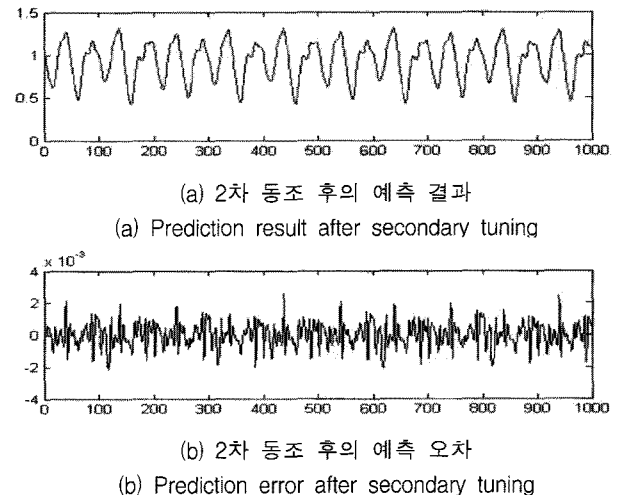


그림 9 2차 동조과정 후의 예측 결과  
Fig. 9 Prediction results after secondary tuning

그림 9를 살펴보면 실제 데이터의 값과 시스템의 출력 값이 거의 중복되어 있으며, 오차 결과 또한 매우 작은 값을 보이고 있다. 아래의 표 5는 제안된 시스템과 다른 시스템 [11-13]과의 성능을 비교한 것으로, 성능 평가 지수로는 다음과 같이 RMSE(root mean square error)를 사용하였다.

표 5 다른 방법들과의 성능 비교(단위: E-4=0.0001)

Table 5 Comparison of performances between the proposed method and other methods

모델 지수	WANG	MCM-1	MCM-2	ANFIS2	proposed method	
					1차동조	2차동조
RMSE	0.0108	0.0108	0.0105	0.0012	7.560E-4	7.246E-4

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2} \quad (27)$$

표 5를 살펴보면 제안된 시스템의 성능이 1차 동조과정에 서도 우수하였지만, 2차 동조과정을 거치면서 그 성능이 더욱 우수해 짐을 알 수 있다.

### 7. 결 론

본 논문에서는 비선형 데이터들의 특성을 충분히 반영하면서도 효율적으로 퍼지 시스템을 구현할 수 있는 방법과 구현된 시스템을 더욱 최적화 시킬 수 있는 방법을 제안하였다. 1차 동조 과정에서 데이터의 전처리 과정과 다중 퍼지시스템을 구현하여, 비선형 데이터의 유동적 특성에 시스템이 유연하게 대처할 수 있도록 하였다. 또한, 1차 동조 과정에 사용된 k-means 클러스터링 기법에 의해 탐색된 퍼지 집합들은 2차 동조과정의 실수 코딩 유전알고리즘의 기반이 됨으로써 최적해의 탐색을 위한 넓은 탐색공간의 부담을 경감할 수 있도록 하였으며, 다양한 개선된 유전연산자들은 지역 최적해로의 수렴이나 탐색 시간에 대한 부담 등을 충분히 극복하면서도 최적의 해를 탐색할 수 있도록 하였다. 시뮬레이션의 1차 동조 과정 후의 결과들은 본 논문에 제안된 기법들이 비선형 데이터에 대한 시스템의 구현을 위해 매우 효율적임을 보여 주었고, 2차 동조과정을 통해 그 성능이 더욱 강화될 수 있음을 보여 주었다. 따라서 제안된 기법들은 비선형 데이터를 다루는 다양한 분야에 적용될 수 있을 것으로 사료된다. 향후 연구 과제로는 다중 퍼지시스템의 각각의 성능들을 충분히 고려할 수 있는 적응 모델 선택 방법에 관한 연구가 필요할 것으로 생각된다.

### 감사의 글

본 과제(결과물)는 지식경제부의 지원으로 수행한 에너지 자원 인력 양성사업의 연구결과입니다.

### 참 고 문 헌

[1] K. J. Kim, C. H. Kim, "Using a Clustering Genetic Algorithm to Support Customer Segmentation for Personalized Recommender Systems", LNCS, vol. 3397, pp. 409-415, 2005  
 [2] S. S. Cheng, Y. H. Chao, H. M. Wang, H. C. Fu, "A Prototype-Embedded Genetic K-means Algorithm", ICPR. 2006. 18th International Conference on Pattern Recognition, vol. 2, pp.724-727, 2006  
 [3] K. Venkatalakshmi, P. A. Praisly, R. Maragathavalli, S. M. Shalinie, "Multispectral Image Clustering

Using Enhanced Genetic k-Means Algorithm", Information Technology Journal, vol.6, pp.557-560, 2007

[4] 진강규, 유전알고리즘과 그 응용, 교우사, 2004  
 [5] Y. K. Bang, C. H. Lee "Fuzzy Time Series prediction with Data Preprocessing and Error Compensation Based on Correlation Analysis", International Conference on Convergence and Hybrid Information Technology, vol.2, pp.714-721, 2008  
 [6] D. J. Kim, C. H. Kim, "Forecasting Time Series with Genetic Fuzzy Predictor Ensemble". IEEE Trans. on Fuzzy Systems, vol. 5, pp.523-535, 1997  
 [7] <http://www.personal.buseco.monash.edu.au/hyndman>  
 [8] K. Ozawa, T. Niimura, "Fuzzy Time-Series Model of Electric Power Consumption". IEEE Canadian Conference on Electrical and Computer Engineering, vol. 2, pp.1195-1198, 1999  
 [9] 주용석, 유전알고리즘과 러프집합을 이용한 퍼지 시스템 모델링, 강원대학교 석사학위논문, 2003  
 [10] I. T. Kim, S. R. Lee "A Fuzzy Time Series Prediction Method Based on Consecutive Values", IEEE International Conference on Fuzzy Systems, vol.2, pp. 703-707, 1999  
 [11] L. X. Wang, J. M. Mendel, "Generating fuzzy rules from numerical data, with applications", IEEE Trans. on Systems, Man, and Cybern, 22 No.6, pp1414-1427, 1992  
 [12] 김인택, 공창욱, "시계열 예측을 위한 퍼지 학습 알고리즘", 한국 퍼지 지능시스템 학회, vol.7, No.3, pp. 34-42, 1997  
 [13] M. El-Koujok, R. Gouriveau, N. Zerhouni, "Towards a Neuro-Fuzzy System for Time Series Forecasting in Maintenance Applications," 17th Triennial World Congress of the International Federation of Automatic Control, hal-00298361, version 1, 2008

## 저 자 소 개



### 방 영 근 (方 寧 根)

1974년 7월 4일생. 2000년 삼척대학교 졸업(학사). 2001~2003년 동대학교 전기공학과 졸업(석사). 2004년~현재 강원대학교(춘천) 전기전자공학과 박사과정  
 E-mail : b2y2c1@hanmail.net



### 이 철 희 (李 哲 熙)

1959년 7월 4일생. 1983년 서울대학교 전기공학과 졸업(학사). 1989년 동 대학원 전기공학과 졸업(공학박사). 1990년~현재 강원대학교 전기전자공학부 교수  
 관심분야 : Intelligent System & Soft computing, 제어 및 신호처리  
 E-mail : chlee@kangwon.ac.kr