

## RCGKA를 이용한 최적 퍼지 예측 시스템 설계

### Design of the Optimal Fuzzy Prediction Systems using RCGKA

방 영 근\*      심 재 선\*\*      이 철 희\*\*\*  
Bang, Young-Keun      Shim, Jae-Son      Lee, Chul-Heui

---

#### Abstract

In the case of traditional binary encoding technique, it takes long time to converge the optimal solutions and brings about complexity of the systems due to encoding and decoding procedures. However, the ROGAs (real-coded genetic algorithms) do not require these procedures, and the k-means clustering algorithm can avoid global searching space. Thus, this paper proposes a new approach by using their advantages.

The proposed method constructs the multiple predictors using the optimal differences that can reveal the patterns better and properties concealed in non-stationary time series where the k-means clustering algorithm is used for data classification to each predictor, then selects the best predictor. After selecting the best predictor, the cluster centers of the predictor are tuned finely via RCGKA in secondary tuning procedure. Therefore, performance of the predictor can be more enhanced.

Finally, we verifies the prediction performance of the proposed system via simulating typical time series examples.

키워드 : RCGA, RCGKA, 다중 예측기, 데이터 분류, 차분

Keywords : RCGA, RCGKA, multiple predictors, data classification, differences

---

#### 1. 서론

퍼지이론, 신경망, 유전 알고리즘과 같은 soft computing 기법들은 고전 선형 모델에서 나타나는 한계점들을 적절히 취급할 수 있어 많은 연구가 진행되고 있다. 퍼지 이론은 언어적 모호함을 퍼지 함수 관계에 따른 규칙으로 표현하고 모호함의 정도를 수치적으로 표현함으로써 이를 해결하려고 하였으며, 퍼지 이론에서 나타나는 다양한 문제점들은 다른 soft computing 기법들에 의해 보완되

고 개선되고 있는 중이다[1-2]. 이러한 기술들은 시스템 제어, 데이터 마이닝, 의료진단, 데이터 처리, 시계열 예측 등 많은 분야에서 그 성능이 입증되었을 뿐만 아니라 더욱 개선되고 있다. 또한 시계열 예측 분야는 시스템 제어, 경제 계획, 사업계획, 수요와 공급, 기상예측, 고장진단 등과 같이 광범위하게 적용될 수 있으며, 본 논문에서도 제안된 퍼지 시스템을 시계열 예측 분야에 적용 하였다.

시계열 예측 시스템을 모델링하기 위하여 본 논문은 원 시계열의 차분 데이터를 이용하였다. 이는 시계열의 통계적 특성이 원시계열 보단 차분 시계열에서 좀 더 안정되어 나타나며, 이전의 연구들에서 그들은 좋은 예측성과를 나타내었기 때문이다 [3-4]. 또한, 퍼지시스템과 유전알고리즘을 적용하여 시스템의 성능을 보다 향상시킨 연구들도 활발

---

\* 강원대학교 대학원 전기전자공학과 박사과정

\*\* 강원대학교 전기제어공학부 교수

\*\*\* 강원대학교 전기전자공학부 교수, 교신저자

히 진행 중이며, 본 논문에서도 유전자 알고리즘을 적용하여 예측 시스템의 최적화를 꾀하였다. 본 논문은 크게 2개의 동조 과정으로 이루어진다. 첫 번째 동조 과정은 비정상 시계열의 특성들을 보다 잘 드러내는 차분 시계열들을 생성하여 이로부터 퍼지 다중 예측 시스템을 구현하게 된다. 퍼지 예측 시스템의 설계를 위해 사용된 T-S 퍼지모델의 규칙기반을 위해 알고리즘의 구현이 쉬운 k-means 클러스터링 방법이 사용되며, 이렇게 구현된 다중 예측기들 중 성능 검증을 통해, 2차 동조 과정에 필요한 가장 우수한 예측기를 선택하게 된다. 2차 동조 과정은 1차 동조과정의 k-means 클러스터링 알고리즘에 의해 선택된 클러스터 중심값을 RCGA(real-coded genetic algorithm)를 이용하여 더욱 세밀히 동조하게 된다. 이러한 RCGKA(real-coded genetic k-means algorithm)기법은 k-means 클러스터링 알고리즘의 통계적 특성뿐만 아니라 유전 알고리즘의 진화적 특성까지 고려할 수 있으므로 클러스터링이 최적화 되는 것을 가능하게 하였으며, 이를 통해 더욱 우수한 예측성능을 이끌어 낼 수 있었다. 또한 2차 동조 과정에 사용된 유전 알고리즘에는 개선된 성능의 유전 연산자들을 적용함으로써 알고리즘의 성능을 더욱 향상시킬 수 있도록 하였으며, 각각의 과정들은 추후 상세히 설명될 것이다. 마지막으로 다양한 시계열 데이터의 시뮬레이션을 통해 제안된 퍼지 시스템의 성능을 검증하였다.

## 2. 1차 동조 과정

본 논문에 사용된 T-S 퍼지 모델은 전반부의 퍼지집합과 후반부의 선형 수식을 통해 데이터의 비선형 또는 불확실성을 적절히 취급함과 동시에 고전 선형 회귀 모델의 이점을 취할 수 있어 많이 선호되는 모델이며 다음의 수식과 같이 하나의 규칙을 표현할 수 있다.

$$R: \text{If } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ and } \dots \text{ and } x_n \text{ is } A_n \quad (1)$$

$$\text{Then } y = p_0 + p_1x_1 + p_2x_2 + \dots + p_nx_n$$

T-S 퍼지 모델은 수식(1)의 전반부와 같이 규칙을 생성하기 위한 입력공간의 퍼지 분할과 후반부와 같이 출력을 위한 파라미터 식별을 필요로 하게 된다. 1차 동조 과정의 입력공간 퍼지 분할을 위해 본 논문에서는 구조가 간단한 k-means 클러스터링 알고리즘을 사용하였으며, 파라미터 식별을 위해서는 LSM(least square method)를 사용하여 파라미터의 부정확성이 최소가 되도록 하였다. 또한 원시계열의 통계적 특성들을 잘 표현할 수 있는 차분 데이터를 이용하여 모델의 구현과 파라미터 추정의 효율성을 높일 수 있도록 하였다. 그림

1은 1차 동조과정의 흐름도 이다.

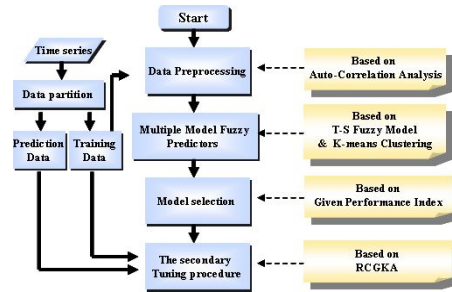


그림 1. 1차 동조 과정의 전체 순서도

데이터의 전처리 절차는 상관분석을 통해 최적의 차분 간격 후보군을 선별하게 되며, 그에 상응하는 차분 데이터를 생성하는 절차이다. 다중 모델 퍼지 예측기의 설계절차는 생성된 각각의 차분 데이터에 적합한 예측기들을 구현하는 절차이며, 모델 선택 절차에서는 2차 동조 과정을 위해 필요한 가장 우수한 성능의 예측기를 선택하는 절차이다. 이렇게 선택된 예측기의 k-means 클러스터링의 결과는 2차 동조 과정에서 유전 알고리즘과 결합하여 더욱 세밀히 동조될 것이다.

### 2.1 데이터의 전처리.

비정상 시계열의 경우 그들의 차분 데이터들은 보다 통계적 특성들이 안정되어 있으므로 시스템의 구현을 위해 더 효과적으로 쓰일 수 있다. 하지만 원 시계열에서 생성된 모든 차분 데이터들이 그 시계열의 패턴이나 규칙성을 잘 반영할 수 있는 것은 아니기 때문에 이에 대한 적절한 판별법이 요구된다. 따라서 본 논문에서는 상관 분석을 통해 원 시계열의 특성을 잘 반영하고 있는 일부를 추출하여 최적 차분 후보군으로 선정하였다. 수식 (2)은 자기 상관 함수로써 차분 간격 값에 따른 차분 시계열과 원시계열의 상관성을 분석하기 위해 사용되었다.

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

여기서,  $N$ 은 훈련데이터의 길이이고,  $j$ 는 차분 간격 값이다. 또한  $y(i)$ 는  $i$ 번째 훈련 데이터이며,  $\bar{y}$ 는 훈련데이터의 평균이다. 다음으로, 구하여진 상관계수들에 랭킹 순으로 나열한 뒤 그들의 차를 계산하고, 계산된 결과에서 가장 큰 폭의 변화를 나타내는 차 연산 결과에 해당하는 랭킹 이상

에 해당하는 차분 간격 값들을 최적 차분 후보군으로 선별하다. 이렇게 선택된 간격 값들을 이용하여, 퍼지 예측기의 구현을 위해 사용될 차분 데이터들은 수식 (4)에 의해 생성된다.

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

여기서  $m(i)$ 는 선택된 차분 간격 값이며, 생성된 각각의 차분 데이터들은 각각의 퍼지 예측기를 위한 규칙기반과 파라미터 식별을 위해 사용된다.

## 2.2 다중 퍼지 예측기의 설계.

다중 모델 퍼지 예측기의 규칙기반의 생성과 파라미터 식별을 위해 사용된 데이터로는 4개의 연속되는 차분 데이터들을 이용하였다. 따라서 각각의 차분 간격  $m(i)$ 에 의해 생성될 수 있는 총 입출력 데이터 쌍은 수식 (4)와 같이 생성된다.

$$sets = \{ \Delta_{m(i)}t_{k+1}, \Delta_{m(i)}t_{k+2}, \Delta_{m(i)}t_{k+3}, \nabla_k \} \quad (4)$$

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

여기서,  $N-m(i)-3$ 은 총 생성될 수 있는 입출력 데이터 쌍의 개수이다. 또한, 수식(4)의  $\nabla_k$ 는  $\Delta_{m(i)}t_k$ 를 의미하며 파라미터 추정을 위한 출력 값을 의미한다. 따라서, 3개의 차분데이터가 하나의 입력 쌍을 이루므로 차분 간격  $m(i)$ 에 대한 퍼지 예측기의  $j$ 번째 규칙  $R_j$ 는 수식 (5)과 같이 다시 정의 된다.

Rule ( $R_j$ )

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

Then  $\nabla_k = p_0^j + p_1^j \Delta_{m(i)}t_{k+1} + p_2^j \Delta_{m(i)}t_{k+2} + p_3^j \Delta_{m(i)}t_{k+3}$  (5)

## 2.3 퍼지 분할과 소속함수.

차분 간격  $m(i)$ 에 대하여 생성된 차분 데이터들 중 최소값과 최대값 사이를 퍼지 분할의 전체 영역으로 하고, 입력 데이터들에 대하여 k-means 클러스터링 알고리즘을 이용하여 구한 클러스터 중심값을 삼각형 소속 함수의 중심 값으로 채택하여 5개의 퍼지 집합으로 분할한다.

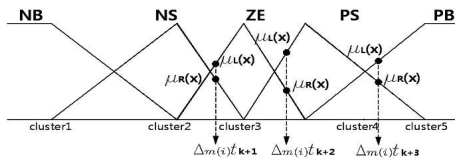


그림 2. 퍼지분할 및 소속함수

그림 2와 같이, 주어진 퍼지 집합에 대한 각각의 입력데이터들이 갖는 소속함수는 수식 (6)과 같이 구할 수 있다.

$$\begin{aligned} \text{If } x \in duster(1) \\ \mu_L(x) &= 1 \end{aligned} \quad (6)$$

$$\mu_R(x) = \frac{duster(z+1) - x}{duster(z+1) - duster(z)}$$

Elseif  $x \in duster(z)$

$$\mu_R(x) = \frac{duster(z+1) - x}{duster(z+1) - duster(z)}$$

$$\mu_L(x) = \frac{x - duster(z-1)}{duster(z) - duster(z-1)}$$

Elseif  $x \in duster(5)$

$$\mu_R(x) = 1$$

$$\mu_L(x) = \frac{x - duster(z-1)}{duster(z) - duster(z-1)}$$

여기서,  $duster(z)$ 는 입력 값  $x$ 가 만족하는 클러스터를 의미하며,  $\mu_L(x)$ 는 입력값  $x$ 가 만족하는 클러스터 중심값으로부터 왼쪽부분의 소속 값이며,  $\mu_R(x)$ 는 오른쪽의 소속함수 값을 의미한다. 이러한 퍼지분할과 소속함수의 정의는 차분 간격 각각에 대하여 수행되며, 구현되는 예측기의 수와 같을 것이다.

## 2.4 퍼지 규칙 생성.

TS 퍼지 규칙의 전반부에 사용되는 입력변수가 3개이고 5개의 퍼지 집합을 사용하였으므로 총 125개의 규칙 생성이 가능하지만, 생성된 규칙의 전반부를 만족 시키는 입력데이터 쌍이 없는 경우 그 규칙의 후반부 파라미터 식별은 불가능하다. 따라서 본 논문에서는 입력데이터 쌍들이 생성하는 규칙들만을 규칙 기반으로 사용하며, 또한 규칙 생성과정에서 중복되는 규칙은 제거하면서 규칙을 생성하였다. 그림 2와 같은 경우의 각각의 입력 데이터 쌍은 2개의 퍼지 집합을 만족하므로 수식 (7)과 같이 8개의 퍼지 규칙을 생성한다.

$$R_1 : \Delta_1 \text{ is NS and } \Delta_2 \text{ is ZE and } \Delta_3 \text{ is PS then } \dots \quad (7)$$

$$R_2 : \Delta_1 \text{ is NS and } \Delta_2 \text{ is ZE and } \Delta_3 \text{ is PB then } \dots$$

$$\vdots$$

$$R_7 : \Delta_1 \text{ is ZE and } \Delta_2 \text{ is PS and } \Delta_3 \text{ is PS then } \dots$$

$$R_8 : \Delta_1 \text{ is ZE and } \Delta_2 \text{ is PS and } \Delta_3 \text{ is PB then } \dots$$

여기서  $\Delta$ 는 수식 4의 각각의 차분간격에 따른 차분 데이터들을 의미하며, 이러한 규칙 생성 작업은 다중 퍼지 예측기들의 각각에 대하여 독립적으로 수행된다.

## 2.5 퍼지 규칙 파라미터 식별.

퍼지 규칙의 후반부 선형 수식의 파라미터 식

은 전반부의 규칙을 만족하는 모든 입력 데이터쌍들을 이용하여 LSM을 적용하여 추정된다. 따라서 차분 간격  $m(i)$ 에 대한 예측기의  $j$ 번째 규칙  $R_j$ 를 만족하는 입력쌍이  $n$ 개라면, 이들에 의한 후반부 선형식은 수식(8)과 같이  $n$ 개의 연립방정식이 될 것이다.

$$\begin{aligned} \nabla_{k1}^j &= p_0^j + p_1^j \Delta_{m(i)}^j t_{k+1}^1 + p_2^j \Delta_{m(i)}^j t_{k+2}^1 + p_3^j \Delta_{m(i)}^j t_{k+3}^1 \\ \nabla_{k2}^j &= p_0^j + p_1^j \Delta_{m(i)}^j t_{k+1}^2 + p_2^j \Delta_{m(i)}^j t_{k+2}^2 + p_3^j \Delta_{m(i)}^j t_{k+3}^2 \\ &\vdots \\ \nabla_{kn}^j &= p_0^j + p_1^j \Delta_{m(i)}^j t_{k+1}^n + p_2^j \Delta_{m(i)}^j t_{k+2}^n + p_3^j \Delta_{m(i)}^j t_{k+3}^n \end{aligned} \quad (8)$$

이를 다시 벡터-행렬식으로 아래의 수식들처럼 표현될 수 있다.

$$\begin{bmatrix} \nabla_{k1}^j \\ \nabla_{k2}^j \\ \vdots \\ \nabla_{kn}^j \end{bmatrix} = \begin{bmatrix} 1 & \Delta_{m(i)}^j t_{k+1}^1 & \Delta_{m(i)}^j t_{k+2}^1 & \Delta_{m(i)}^j t_{k+3}^1 \\ 1 & \Delta_{m(i)}^j t_{k+1}^2 & \Delta_{m(i)}^j t_{k+2}^2 & \Delta_{m(i)}^j t_{k+3}^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta_{m(i)}^j t_{k+1}^n & \Delta_{m(i)}^j t_{k+2}^n & \Delta_{m(i)}^j t_{k+3}^n \end{bmatrix} \begin{bmatrix} p_0^j \\ p_1^j \\ p_2^j \\ p_3^j \end{bmatrix} \quad (9.a)$$

$$Y_j = D_j P_j \quad (9.b)$$

여기서,  $Y_j$ 는 출력벡터,  $D_j$ 는 입력 벡터,  $P_j$ 는 파라미터 벡터이다. 식 (9.b)의 파라미터 벡터는 최소자승법을 이용하여 다음과 같이 구할 수 있다.

$$\hat{P}_j = (D_j^T D_j)^{-1} D_j^T Y_j \quad (10)$$

이렇게 구하여진 파라미터들은 새로운 입력데이터의 예측 출력을 위해 사용된다. 만약 입력 데이터 쌍이 총  $q$ 개의 퍼지 규칙을 만족한다면, 시스템의 출력은 규칙  $R_i$ 의 전반부에서 결정되는 적합도  $\mu_i$ 와 수식(5)에 의해 결정된  $\hat{\nabla}_k^i$ 로부터 다음과 같은 가중합으로 구할 수 있다.

$$\hat{\nabla}(t) = \frac{\sum_{i=1}^q \mu_i \hat{\nabla}_k^i}{\sum_{i=1}^q \mu_i} \quad (11)$$

퍼지 예측기의 출력  $\hat{\nabla}(t)$ 는 시간  $t$ 에 대하여 예측된 미래 값의 증가분을 의미하므로 원하는 시계열의 최종 예측값  $\hat{y}(t+p)$ 는 다음과 같이 구할 수 있다.

$$\hat{y}(t+p) = y(t) + \hat{\nabla}(t) \quad (12)$$

## 2.6 최적 예측기 및 퍼지 분할 결정.

다중 모델 퍼지 예측기들 중에서 가장 최적의

예측기를 찾는 방법은 성능평가 지수를 이용하여 해결 할 수 있다. 본 논문에서는 각각의 예측기들에 대한 규칙과 파라미터들을 이용하여 훈련데이터의 예측을 수행하고, 식 (13)과 같이 주어진 자승 오차 평균(MSE: squared error)를 최소화 하는 예측기를 선택하였다.

$$MSE = \frac{1}{N-m(i)-3} \sum_{n=m(i)+3}^N (y(n) - \hat{y}(n))^2 \quad (13)$$

여기서  $\hat{y}(n)$ 은 퍼지 예측기에서 출력한  $y(n)$ 의 예측 값이며, 차분 간격  $m(i)$ 인 퍼지 예측기 모델에서는 예측에 사용되는 입·출력 데이터쌍의 개수가  $N-m(i)-3$ 이므로 평균을 구할 수 있다.

## 3. 2차 동조 과정

2차 동조 과정은 1차 동조과정에서 선택된 최적의 퍼지 예측기를 더욱 세밀하게 동조하는 과정이다. 선택된 예측기의 1차 동조 과정에서 k-means 클러스터링의 통계적 특성에 기반하여 찾아진 클러스터 중심값들에 대하여 2차 동조 과정에서는 유전 알고리즘의 적자생존의 법칙에 따라 최적해를 탐색할 수 있는 능력과 탐색영역이나 정밀도를 고려한 실수코딩기법(real encoding)을 적용한 RCGKA(real-coded genetic k-means algorithm) 기법을 이용하여 더욱 세밀하게 클러스터들을 동조하게 된다. 또한 유전 알고리즘의 성능을 더욱 향상시키기 위하여, 본 논문에서는 재생산 연산자(regeneration operator)[Jin 1996], 엘리트 전략(elitist strategy) [De Jong 1975], 수정단순교배(modified simple crossover)[Jin 2000]와 동적 돌연변이(dynamic mutation)[Janikpw 1991, Micgakewicz 1996] 연산자가 적용 되었다.

### 3.1 초기집단의 생성과 실수 코딩

$g$ 세대( $g$ 번째 반복횟수)에서의 집단  $Pop(g)$ 는 수식(14)과 같이 염색체로 특정 지어지는  $L$ 개 개체들의 집합으로 정의 될 수 있다.

$$Pop(g) = \{s_1(g), s_2(g), \dots, s_L(g)\} \quad (14)$$

여기서,  $s_i(g)$ 는  $i$ 번째의 염색체로 탐색 공간상의 한점을 나타내며,  $L(>1)$ 은 집단의 크기를 의미한다. 초기 집단은 1차 동조 과정에서 k-means 클러스터링에 의해 구하여진 중심값들을 기반으로 하여 실수코딩 기법에 의해 임의적으로 집단의 크기만큼 생성된다. 수식(15)은  $i$ 번째 염색체에 대한 실수코딩의 한 예이며,  $z$ 는 클러스터의 개수를 의미한다.

$$s_i(g) = (v_{i1}(g) v_{i2}(g) \cdots v_{iz}(g)) \quad (15)$$

식(15)과 같은 실수 코딩은 염색체의 유전자와 중심값들이 정확히 일대 일의 실수 값으로 표현되므로 상당히 넓은 영역을 표현할 수 있고, 염색체의 길이를 증가할 필요가 없으므로 탐색을 위한 알고리즘의 능률을 더욱 개선할 수 있다.

### 3.2 적합도 함수와 재생산 연산자

클러스터 내에 분류된 데이터들과 클러스터 중심과의 분산이 작을수록 데이터와 클러스터의 적합성이 높은 것으로 볼 수 있다. 따라서 본 논문에서 사용된 목적함수(objective function)는 세대의 진화에 따라 수식(16)로 주어지는 전체 클러스터 내 분산(TWCV:total within-cluster validation)을 최소화 하는 방향으로 유전 알고리즘을 수행한다.

$$OF(s) = \sum_{c=1}^z \left( \frac{1}{h} \sum_{u=1}^h (x_u - v_c)^2 \right) \quad (16)$$

여기서,  $x_u$ 는  $v_c$ 를 중심값으로 하는 클러스터에 포함된 데이터이며,  $h$ 는 포함된 데이터들의 개수이다. 또한 자손을 생성하기 위해 필요한 적합도 함수는 수식(17)과 같은 선형 스케일링 기법을 적용한다.

$$Fit(s) = OF(s) + C_b \quad (17)$$

구하여진 적합도를 기반으로 재생산 연산자에서는 자손을 생성하게 된다. 자손을 생성하기 위한 탐색 점은 각 개체들의 적합도와 가장 우수한 개체의 적합도를 이용하여 다음과 같이 계산된다.

$$\bar{s}_{ic}(g+1) = s_{ic}(g) + \eta \frac{[Fit_{best}(g) - Fit_i(g)]}{Fit_{best}(g)} [s_{best(c)}(g) - s_{ic}(g)] \quad (18)$$

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

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.

### 3.3 수정단순교배, 동적돌연변이, 엘리트전략

수정단순교배(modified simple crossover)[진 2000]는 이진 코딩의 일점교배를 묘사한 연산자로 두 염색체에 대한 교배점이  $c \in [1, z-1]$ 인 구간에서 무작위로 선택되면 식(19)에 의하여 일차 결합되고, 그 후 교배점을 기준으로 서로 교환되어 자손을 생성하게 된다.

$$\begin{aligned} \tilde{v}_c^a &= \lambda \tilde{v}_c^b + (1-\lambda) \tilde{v}_c^a \\ \tilde{v}_c^b &= \lambda \tilde{v}_c^a + (1-\lambda) \tilde{v}_c^b \quad (1 \leq c \leq z) \end{aligned} \quad (19)$$

여기서,  $\lambda \in [0,1]$ 로 제한되는 값을 가진다. 이는 항상 적합한 형태의 자손을 생성하기 위한 것이다. 또한, 본 논문은 동적 돌연변이(dynamic mutation)와 엘리트 전략(elitist strategy)을 적용하였다. 동적 돌연변이는 진화의 과정에서 최적 개체 쪽으로의 수렴 속도의 향상을 위한 것이고, 엘리트 전략은 이전 세대의 최적 개체를 기억하였다가 현 세대에서 소멸된 것이 확인 될 경우 현세대의 가장 약한 적합도의 개체와 태체함으로써 유전알고리즘의 성능을 더욱 개선하기 위한 것이다.

## 4. 컴퓨터 시뮬레이션 및 검토

본 논문에서 사용된 시뮬레이션 예로는 비정상 및 혼돈 비선형 특성을 갖는 대표적인 시계열 데이터를 선택하였으며, 성능 평가지수에 의해 제안된 시스템의 성능을 분석하고, 기존의 연구 결과들과 비교하였다. 아래의 표 1은 제안된 시스템에 사용된 파라미터 값들이다.

표 1 제안된 시스템의 변수

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 ( $C_b$ )	$\min(OF)$
세대의 가중치 ( $\eta$ )	1.7
불균등 정도 ( $\alpha$ )	5

첫 번째 시뮬레이션 시계열로는 다우지수 데이터[10]로, 총 292개의 데이터 중 200개의 데이터를 훈련데이터로 사용하고 나머지 데이터를 예측 데이터로 사용하였다. 그림 3는 1차 동조 과정 후의 예측 결과를 보여준다.

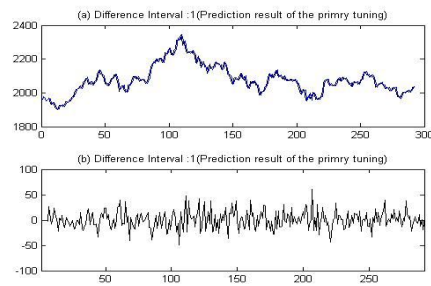


그림 3. 1차 동조과정 후의 예측 결과

그림 3에서, 검은 색은 원시계열 데이터 값이고 파란색은 예측된 결과를 나타낸다. 그림 4의 아래 오차 그림을 살펴보면, 제안된 시스템의 1차 동조 과정 만으로도 좋은 예측 결과를 보임을 알 수 있다. 아래의 그림 4는 2차 동조 과정을 통해 목적함수와 클러스터 중심이 최적화 되는 것을 보여 주며, 그림 5는 2차 동조 과정 후의 최종 예측 결과를 보여 준다.

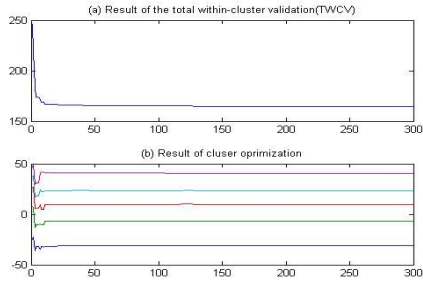


그림 4. 목적함수와 파라미터의 최적화 결과

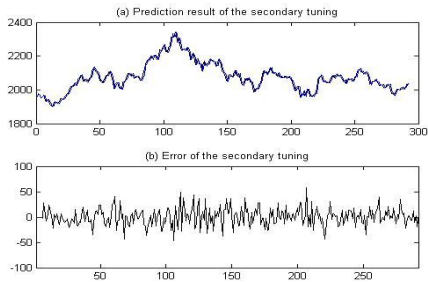


그림 5. 2차 동조과정 후의 예측 결과

1차 동조 후와 2차 동조 후의 성능비교를 위하여 다음과 같이 정의되는 MRE(mean relative error)를 이용하였다.

$$MRE = \frac{1}{N} \sum_{i=1}^N \left( \frac{|y(i) - \hat{y}(i)|}{|y(i)|} \times 100 \right) \% \quad (20)$$

아래의 표 2는 1,2차 동조 후의 성능을 비교한 것이다.

표 2 1,2차 동조 후의 성능 비교

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

표 2에서 보듯이 RCGKA기법을 이용한 2차 동조 후의 결과가 1차 동조 후의 결과 보다 더 우수함을 알 수 있다. 이것은 제안된 RCGKA이 시스

템 최적화 기법으로 유용함을 보여주는 것이다.

다음의 시뮬레이션 데이터로는 호주의 전력생산량 데이터[10]를 이용하였다. 호주의 전력생산량 데이터는 총 155개의 데이터로 구성되며, 다른 논문들과의 비교를 위해 70개의 데이터를 훈련데이터로, 나머지 데이터를 예측 데이터로 사용하였다. 아래의 그림 6은 최종 예측 결과를 보여준다.

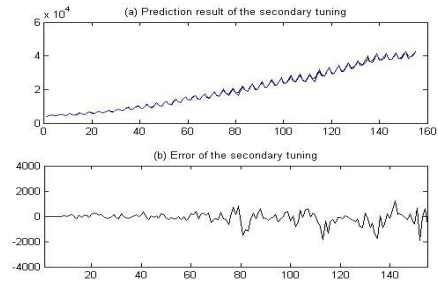


그림 6. 2차 동조과정 후의 예측 결과

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

표 3 1,2차 동조 후의 성능 비교

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

첫 번째 시뮬레이션과 마찬가지로 RCGKA를 이용한 2차 동조 후의 결과가 1차 동조 후의 결과 보다 상당히 개선됨을 알 수 있다. 아래의 표 4는 제안된 예측 시스템과 다른 기법들[2,5,6,8]을 비교한 것이다.

표 4 제안된 기법과 다른 기법들과의 성능 비교

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

표 4에서 알 수 있듯이, 제안된 퍼지 예측 시스템은 1차 동조 과정으로도 우수한 예측을 수행할 수 있으며, 2차 동조과정을 거치면서 그 성능이 더욱 개선됨을 알 수 있다.

## 5. 결론

본 논문에서는 비정상 시계열 데이터들의 특성을 충분히 반영하면서도 효율적으로 성능을 개선할 수 있는 퍼지 예측 시스템의 설계 방법을 제안하였다. 먼저, 비정상 시계열 속에 내재된 여러 패턴이나 규칙성들을 원시계열로부터 생성 가능한 차

분시계열을 이용하여 보다 효율적으로 분석하고, 이를 토대로 각각의 차분 시계열에 적합한 다중 모델 예측 시스템을 구현함으로써 효과적인 예측이 가능하도록 하였다. 또한, 예측 시스템의 구현에는 구조가 간단한 k-means 클러스터링 알고리즘을 1차 동조 과정에 적용하여, 2차 동조 과정의 유전 알고리즘 수행에 대한 넓은 탐색공간의 부담을 줄일 수 있도록 하였다. 제안된 논문의 RCGKA기법은 기존의 이진 코딩 기법에서 야기되는 염색체 길이에 대한 부담과 부호화와 복호화 같은 알고리즘 구현의 복잡성을 피하면서도 효율적인, 실수코딩 기법을 적용하였고, 성능이 우수한 다양한 유전 연산자들을 적용함으로써 비정상 시계열 데이터에 대한 보다 우수한 예측이 가능하도록 하였다. 시뮬레이션 결과들은 제안된 시스템의 성능이 매우 우수함을 보여준다. 따라서, 제안된 방법들은 시계열 예측뿐만 아니라 제어, 데이터 처리 등 비정상 데이터를 다루는 다양한 분야에서 유용하게 적용될 수 있을 것으로 사료된다.

- [7] Carl G. Looney, *Pattern Recognition using Neural Networks*, Oxford University Press, 1997.
- [8] 주용석, "유전알고리즘과 러프집합을 이용한 퍼지시스템 모델링", *강원대학교 석사학위논문*, 2003.
- [9] 진강규, *유전 알고리즘과 그 응용*, 교우사, 2004.
- [10] <http://www-personal.buseco.monash.edu.au/hyndman/TSDL/>

### 참 고 문 헌

- [1] Juhong Nie, "Nonlinear Time-Series Forecasting: A Fuzzy Neural Approach", *Neuro Computing*, Vol.16, pp.63-76, MacMaster University, 1997.
- [2] K.Ozawa, T.Nimura, "Fuzzy Time-Series Model of Electric Power Consumption", *IEEE Canadian Conference on Electrical and Computer Engineering*, pp.1195-1198, 1999.
- [3] Chul-Heui Lee, Sang-Hun Yoon, "Fuzzy Nonlinear Time Series Forecasting with Data Preprocessing and Model Selection", *Journal of Telecommunications and Information*, Vol.5, pp.232-238, 2001.
- [4] Young-Keun Bang, Chul-Heui, Lee, "Desigh of Multiple Model Fuzzy Predictors using Data Preprocessing and its Application", *Trans. Korean institute of electrical engineers*, 2009.
- [5] Dai-jin Kim, Chul-hyun Kim, "Forecasting Time Series with Genetic Fuzzy Predictor Ensemble". *IEEE Trans. on Fuzzy Systems*, Vol.5, pp.523-535, 1997.
- [6] Inteak Kim, Song-Rock Lee, "A Fuzzy Time Series Prediction Method Based on Consecutive Values", *IEEE International Conference on Fuzzy Systems*, Vol.2, pp.703 -707, 1999.