

러프 집합 기반 적응 모델 선택을 갖는 다중 모델 퍼지 예측 시스템 구현과 시계열 예측 응용

Multiple Model Fuzzy Prediction Systems with Adaptive Model Selection Based on Rough Sets and its Application to Time Series Forecasting

방영근* · 이철희**

Young-Keun Bang* and Chul-Heui Lee**

* 강원대학교 대학원 전기전자공학과

** 강원대학교 IT특성화학부대학 전기전자공학부

요 약

최근 시계열 예측에 결론부에 선형식을 갖는 TS 퍼지 모델이 많이 이용되고 있는데, 이의 예측 성능은 정상성과 같은 데이터의 특성과 밀접한 관련이 있다. 그러므로 본 논문에서는 특히 비정상 시계열 예측에 매우 효과적인 새로운 예측 기법을 제안하였다. 시계열의 패턴이나 규칙성을 잘 끌어내기 위한 데이터 전처리 과정을 도입하고 다중 모델 TS 퍼지 예측기를 구성한 뒤, 러프집합을 이용한 적응 모델 선택 기법에 의해 입력 데이터의 특성에 따라 가변적으로 적합한 예측 모델을 선택하여 시계열 예측이 수행되도록 하였다. 마지막으로 예측 오차를 감소시키기 위하여 오차 보정 메커니즘을 추가함으로써 예측 성능을 더욱 향상시켰다. 시뮬레이션을 통해 제안된 기법의 성능을 검증하였다. 제안된 기법은 예측 모델 구현과 예측 수행 과정에서 시계열 데이터의 특성들을 잘 반영할 수 있으므로 불확실성과 비정상성을 갖는 시계열의 예측에 매우 효과적으로 이용될 수 있을 것이다.

키워드 : 비정상 시계열, 데이터 전처리, TS퍼지 모델, 러프집합, 적응모델 선택, 오차 보정

Abstract

Recently, the TS fuzzy models that include the linear equations in the consequent part are widely used for time series forecasting, and the prediction performance of them is somewhat dependent on the characteristics of time series such as stationariness. Thus, a new prediction method is suggested in this paper which is especially effective to nonstationary time series prediction. First, data preprocessing is introduced to extract the patterns and regularities of time series well, and then multiple model TS fuzzy predictors are constructed. Next, an appropriate model is chosen for each input data by an adaptive model selection mechanism based on rough sets, and the prediction is going. Finally, the error compensation procedure is added to improve the performance by decreasing the prediction error. Computer simulations are performed on typical cases to verify the effectiveness of the proposed method. It may be very useful for the prediction of time series with uncertainty and/or nonstationariness because it handles and reflects better the characteristics of data.

Key Words : nonstationary time series, data preprocessing, TS fuzzy model, rough sets, adaptive model selection, error compensation

1. 서 론

시계열 데이터는 시간의 흐름에 따라 일정한 간격으로 관측하여 기록한 자료로서, 이들 데이터로부터 특징적인 패턴을 규명하거나 일련의 규칙성을 발견하여 이를 모델링한다면 미래에 대한 예측이 가능할 것이다. 따라서 관측된 데

이터로부터 데이터의 이면에 내재해 있는 특성과 규칙성 등을 분석하여 모델을 개발하는 것이 필요하다.

시계열의 모델로는 AR(Auto-Regressive), MA(Moving Average), ARMA(Auto-Regressive Moving Average)와 같은 선형 회귀 모델이나 Box-Jenkins 방법(Auto-Regressive Integrated Moving Average : ARIMA) 등이 널리 사용되고 있으며[1-2], 최근에는 선형적 접근 방법들의 문제점들을 극복하고자 퍼지이론, 신경망, 유전 알고리즘 등과 같은 soft computing 기법을 이용하여 시계열 데이터의 불확실성과 비선형 특성을 적절하게 취급할 수 있는 방법들에 대한 연구가 많이 이루어지고 있다 [3-5].

또한 비정상(nonstationary) 시계열 예측의 경우 통계량

접수일자 : 2008년 7월 31일

완료일자 : 2008년 10월 15일

** 교신저자

본 논문은 강원대학교 연구년 교수 지원과 지식경제부가 지원한 에너지자원 인력양성사업의 연구결과입니다

이 시간에 따라 변화하는 특성을 반영할 수 있도록 원 시계열 데이터의 1차 차분 데이터를 사용하는 방법이 연구되었으나[6], 비정상 시계열 데이터의 유동적 특성을 잘 표현하는 문제에 있어 그 효과와 활용이 제한적이므로 우리는 데이터의 전처리와 모델 선택을 통해 이를 극복하고자 한 바 있다[7]. 그러나 이 방법은 다중 모델을 구성한 의도를 100% 반영한 효율적인 방법이라고는 할 수 없다.

따라서 본 논문에서는 러프 집합을 이용하여 훈련 모드에서의 결과들을 기반으로 모델 선택에 필요한 지식을 발견하고 적당한 규칙을 생성함으로써 데이터 특성에 따라 모델 선택이 가변적인 적응 메커니즘을 갖는 모델 선택 기법을 제안하였다. 러프 집합[8-9]은 모든 데이터들로부터 동질성을 갖는 데이터들을 분류하고 이들의 특징을 추출하여 연관성에 따른 정형화된 지식으로 표현할 수 있으므로 매우 유용한 기법이다.

제안된 방법은 데이터 전처리 과정에서 상관해석에 기반하여 최적 차분 간격 후보군을 적절히 선택하여 상응하는 차분 데이터를 입력으로 하는 퍼지 예측기를 다중 모델로 구성한다. 그리고 성능 지표에 의해 가장 최적인 차분 간격을 선정하여 이에 상응하는 차분 데이터를 러프 집합에 의한 모델 선택 규칙의 조건부 속성으로 사용함으로써 현재 순간의 데이터에 대한 최적 모델이 선택되어 예측이 수행되도록 하였다. 또한, 예측기 입력과 규칙 생성에 기여한 훈련 데이터들과의 상관 해석에 기반한 오차 보정 기법을 추가함으로써 예측 성능을 더욱 향상시킬 수 있도록 하였다.

시계열 예측의 전형적인 사례로 꼽히는 3개의 시계열 데이터를 이용하여 제안된 방법의 효용성을 검증하였다.

2. 다중 모델 퍼지 예측 시스템의 구성

서론에서 언급한 바와 같이, 차분 데이터를 이용하면 비정상 또는 혼돈 시계열과 같이 유동적 비선형성이 존재하는 시계열들의 통계량을 안정화시키고 내재된 패턴이나 규칙성을 보다 효과적으로 찾아낼 수 있다. 그런데 시계열의 패턴이나 규칙성을 가장 잘 뽑아낼 수 있는 차분 간격이 얼마인지를 알 수 없으므로, 본 논문에서는 훈련 데이터를 이용하여 높은 가능성을 지니는 다수의 차분 간격 후보군을 선정하고(데이터의 전처리), 선정된 차분 간격들을 갖는 차분 데이터들을 입력으로 하는 퍼지 예측 모델을 각각 독립적으로 구성하게 된다(다중 모델). 이렇게 구성된 다중 모델들은 입력 데이터의 특성에 따라 최상의 모델이 선택되어(모델선택) 실제적인 시계열 예측을 수행하게 된다. 제안된 퍼지 시계열 예측 시스템의 전체적인 구성은 그림 1과 같다.

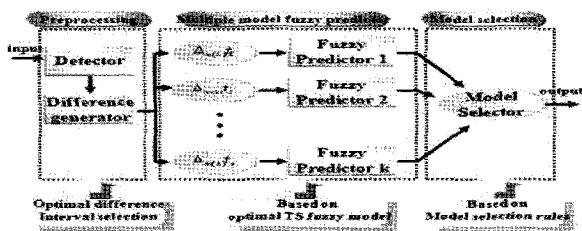


그림 1. 제안된 퍼지 시계열 예측 시스템의 구조
Figure 1. Structure of proposed fuzzy predictor

그림 1과 같은 구조의 제안된 퍼지 시계열 예측 기법은

크게 최적의 예측기를 설계하기 위한 훈련 모드와 설계된 예측기를 이용하여 예측을 수행하는 동작 모드로 나눌 수 있다. 이때 훈련 데이터는 초기 시계열 데이터 중에서 모든 TS 퍼지 모델의 파라미터 식별이 제대로 이루어질 수 있을 정도의 양을 선택한다. 아래의 그림 2는 제안된 기법의 전체적인 진행 순서도이며 방법은 다음과 같이 진행된다.

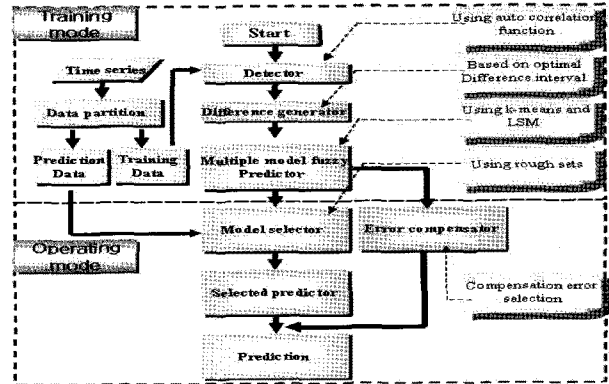


그림 2. 제안된 알고리즘의 순서도
Figure 2. Flow chart of proposed fuzzy predictor

훈련 모드

- 1단계) 자기 상관 계수를 이용한 차분 간격 후보군 선정 및 차분 데이터 생성(데이터 전처리)
- 2단계) 각 차분 간격에 대해 TS 퍼지 규칙 도출(다중 모델 퍼지 예측기 구성)
- 3단계) 모델 선택을 위한 의사결정표의 조건부 속성 생성
- 4단계) 구성된 다중 모델 퍼지 예측기에 훈련 데이터를 예측 입력으로 사용하여 시계열 예측 수행
- 5단계) MSE를 이용한 최적 차분 간격 선정
- 6단계) 모델 선택을 위한 의사결정표의 의사결정부 속성 생성
- 7단계) 러프집합을 이용한 모델 선택 규칙 도출

동작모드

- 8단계) 현재 순간 데이터에 대해 최적 차분 간격에 상응하는 차분 데이터 생성
- 9단계) 모델 선택 규칙에 의한 최적 퍼지 예측기 선택
- 10단계) 선택된 퍼지 예측기에 의한 시계열 예측 수행
- 11단계) 교차 상관 계수를 이용한 오차 보정값 결정 및 오차 보정

그림 2에서 검출기(detector)는 데이터에 대한 전처리가 이루어지는 부분으로서, 상관 해석에 의거하여 최적의 차분 간격에 대한 적절한 후보군을 선택하게 되며 자세한 설명은 3장에서 다룬다. 검출기를 통해 생성된 차분데이터들은 그림 1에 나타난 것처럼 퍼지 예측기 다중 모델을 형성하게 되며, 각각의 모델로는 규칙의 조건부에는 퍼지 집합을 이용함으로써 데이터의 불확실성이나 비선형성을 능동적으로 수용하면서도 결론부의 선형 수식을 통해 고전적인 선형 회귀 모델의 이점도 동시에 취할 수 있으며, 데이터 표현에 대한 적합성과 예측기 구현의 용이성도 꾀할 수 있는 TS 퍼지모형을 사용하였다. 아래의 수식은 TS 퍼지 모델의 규칙을 표현한 것이다.

$$R: \text{If } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ and } \dots \text{ and } x_n \text{ is } A_n \\ \text{Then } y = a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (1)$$

TS 퍼지 모델 규칙의 입력 공간 퍼지 분할에는 k-means 알고리즘[10]을 이용한 clustering 기법을 적용하여 데이터 표현에 대한 적합도를 향상시켰으며, 결론부 파라미터 식별은 최소자승법(Least Squares Method)을 이용하여 파라미터 식별의 부정확성을 최소화하도록 하였다. 그림 2의 모델선택기(model selector)에서는 새로운 시계열 데이터가 들어올 때마다 다중 모델 중에서 그 데이터에 최적의 퍼지 예측기를 선택하는 적응 메커니즘을 채택하여, 매 시간 순간마다 최적의 퍼지 예측기를 통해 예측이 수행되도록 하였으며 자세한 설명은 5장에서 다루게 된다.

마지막으로, 그림 2의 오차보정기(error compensator)는 예측 정확도를 향상시킬 목적으로 차분 입력 쌍과 각 퍼지 규칙 R_j 의 생성에 기여한 훈련 데이터들과의 교차 상관 계수를 이용하여 유사성을 판별하고, 유사성이 가장 큰 훈련 데이터의 예측 오차를 이용하여 입력의 특성이 반영된 오차보정이 이루어지도록 한 것으로 자세한 설명은 6장에서 다루게 될 것이다.

3. 데이터의 전처리

데이터의 전처리는 데이터의 유동적 비선형성 뒤에 내재된 패턴이나 규칙성이 예측 시스템에 잘 표착될 수 있는 최적의 차분 간격 후보군을 선별하고 이에 상응하는 차분 데이터들을 생성하는 과정이다.

먼저 주어진 시계열 데이터의 훈련 데이터로부터 다음과 같이 정의되는 자기 상관 계수를 구한다.

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

여기서, $y(i)$ 는 i 번째 훈련 데이터이며, \bar{y} 는 훈련 데이터의 평균이다. 또한, N 은 훈련 데이터의 길이이고, j 는 차분 간격 값이다. 자기 상관 계수 값이 큰 것부터 작은 순서대로 나열한 뒤 인접한 두 상관 계수의 값의 차가 가장 크게 나는 것에서 잘라 차분 간격의 후보군으로 선정한다. 예를 들어, 다음과 같이 자기 상관 계수가 크기순으로 나열된다면

$$r_{m(1)} = 0.96, r_{m(2)} = 0.94, r_{m(3)} = 0.90, r_{m(4)} = 0.83, \\ r_{m(5)} = 0.62, r_{m(6)} = 0.55, r_{m(7)} = 0.46, r_{m(8)} = 0.38$$

$r_{m(4)} - r_{m(5)} = 0.21$ 로 가장 차이가 크므로 $r_{m(1)}, r_{m(2)}, r_{m(3)}, r_{m(4)}$ 의 4개의 상관 계수가 선택되고 이에 대응되는 차분 간격 $m(1), m(2), m(3), m(4)$ 가 최적 차분 간격의 후보군으로 선정된다. 최적 차분 간격의 후보군이 선정되면, 그림 1의 차분데이터발생기(Difference Generator)에서 이들 차분 간격 $\{m(i)\}$ 을 갖는 차분 데이터들을 원 시계열의 훈련 데이터로부터 생성하여 퍼지 예측 모델에 입력으로 사용하게 된다. 차분 간격 $m(i)$ 에 대해 시간 t 순간에 생성되는 차분 데이터는 다음과 같다.

$$\Delta_{m(i)}t_0 = y(t) - y(t - m(i)) \\ \Delta_{m(i)}t_1 = y(t-1) - y(t-1 - m(i)) \\ \vdots \\ \Delta_{m(i)}t_n = y(t-n) - y(t-n - m(i)) \\ \vdots \\ \Delta_{m(i)}t_{t-m(i)-1} = y(m(i)+1) - y(1) \quad (3)$$

(3)에서 보면, N 개의 훈련 데이터로부터 생성되는 차분 데이터의 수가 차분 간격 값 $m(i)$ 에 따라 달라지는데, 이는 차분 데이터 생성에 유효 데이터만을 이용하기 때문이다. 그러므로 $y(n) = 0, n \leq 0$ 라는 가정을 쓰지 않아 예측 과정에 편향(bias)이 생기는 것을 방지하는 효과를 가져다 준다. 이렇게 생성된 각 차분 간격 값에 대한 차분 데이터들은 퍼지 예측기의 입력으로 사용될 뿐만 아니라 TS 퍼지 규칙 기반의 결론부 선행 수식 파라미터의 식별에도 이용된다. 또한 러프 집합을 이용하여 모델 선택 규칙을 도출하기 위한 의사결정표의 조건부 속성을 생성하는 데에도 사용된다.

4. 다중 모델 퍼지 예측기의 구현

제안된 방법의 예측기 모델에 사용된 TS 퍼지 모델 입력 변수와 퍼지 집합의 개수는 여러 시뮬레이션 분석 결과를 토대로 요구되는 예측 성능을 충족시키면서 예측기 구현이 용이한 최저 수준으로 선정하였다. 선정된 k 개의 차분 간격 $\{m(i)\}$ 에 대해 각각 퍼지 예측 모델을 구성하며, 각 퍼지 예측 모델마다 예측이 수행되는 시간 t 에서 가장 최근의 차분 데이터 $\Delta_{m(i)}t_0, \Delta_{m(i)}t_1, \Delta_{m(i)}t_2$ 의 3개를 입력 변수로 하고 이들에 대해 각각 NB, NS, ZE, PS, PB의 5개 퍼지 집합으로 입력 공간을 퍼지 분할하였다.

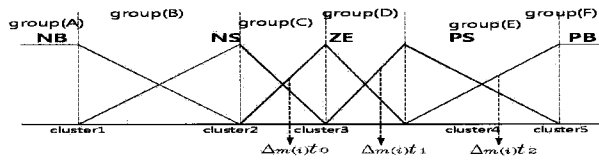
따라서 차분 간격 $m(i)$ 에 해당하는 j 번째 TS 퍼지 예측 모델의 언어적 규칙은 다음과 같은 형태가 된다.

$$R_j: \text{if } \Delta_{m(i)}t_0 \text{ is } A_j \text{ and } \Delta_{m(i)}t_1 \text{ is } B_j \text{ and } \Delta_{m(i)}t_2 \text{ is } C_j \\ \text{then } \hat{y}_t^j = a_0^j \Delta_{m(i)}t_0 + a_1^j \Delta_{m(i)}t_1 + a_2^j \Delta_{m(i)}t_2 \quad (4)$$

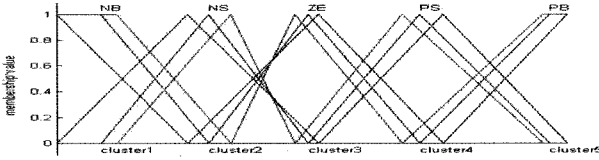
여기서 결론부의 출력은 $\hat{y}_t^j = \hat{y}^j(t+p) - y(t)$ 로서 현재 순간의 데이터 $y(t)$ 와 j 번째 규칙에 의해 결정되는 예측하고자 하는 p 스텝 앞의 예측 값 $\hat{y}^j(t+p)$ 와의 차분 값으로 현재의 데이터로부터 예측되는 미래 값의 증가분의 예측 값이 된다.

4.1 입력 공간의 퍼지 분할

퍼지 예측기의 입력변수 $\Delta_{m(i)}t_0, \Delta_{m(i)}t_1, \Delta_{m(i)}t_2$ 에 대해 다음과 같이 퍼지 영역으로 분할한다. 먼저 N 개의 훈련 데이터로부터 차분 간격 $m(i)$ 에 대해 생성된 차분 값 $\Delta_{m(i)}t_0, \dots, \Delta_{m(i)}t_{N-m(i)-1}$ 의 최소값과 최대값 사이를 퍼지 분할의 전체 영역(universe of discourse)으로 하고, 입력 데이터에 대해 k-means 클러스터링 알고리즘[10]을 이용하여 구한 클러스터 중심값을 삼각형 소속 함수의 중심값으로 채택하여 그림 3(a)와 같이 NB, NS, ZE, PS, PB의 5개 퍼지 집합으로 분할하였다.



(a) 입력 변수에 대한 퍼지 집합과 소속함수
(a) fuzzy sets and membership function for input variables



(b) k-means 알고리즘에 의한 소속 함수의 형태
(b) execution of k-means clustering algorithm

그림 3. 입력 공간의 퍼지 분할과 소속 함수 결정
Figure 3. Fuzzy partition and membership function

그림 3(b)는 k-means 알고리즘을 이용해 차분 데이터에 적합화시킨 소속 함수의 한 예로 클러스터링 이전에 따른 소속 함수의 변화를 살펴볼 수 있다. 각각의 퍼지 예측기마다 사용되는 차분 데이터가 다르므로, 이러한 과정은 k 개의 퍼지 예측 모델 각각에 대해 독립적으로 수행된다.

그림 3과 같은 소속 함수에 대해서 퍼지 예측기의 입력 값 x 에 대한 소속 함수 값은 다음과 같이 구할 수 있다.

$$\begin{aligned}
 &A, F \text{ 구간} : \mu_{NB}(x) \text{ or } \mu_{PB}(x) = 1 \\
 &\mu_L(x) = \frac{C_R - x}{C_R - C_L} \\
 &B, C, D, E \text{ 구간} : \\
 &\mu_R(x) = \frac{x - C_L}{C_R - C_L}
 \end{aligned} \tag{5}$$

여기서 $\mu_L(x)$ 는 입력 값 x 가 속한 구간의 왼쪽에 위치한 클러스터 중심값 C_L 을 중심값으로 하는 퍼지 집합에 대한 소속 함수 값을 나타내며, $\mu_R(x)$ 는 오른쪽에 위치한 클러스터 중심값 C_R 을 중심값으로 하는 퍼지 집합에 대한 소속 함수 값을 나타낸다. 예를 들어, 그림 3(a)에서 입력 값이 $x = \Delta_{m(i)}t_1$ 인 경우 $C_L = \text{cluster3}$, $C_R = \text{cluster4}$ 가 되고 $\mu_L(x) = \mu_{ZE}(x)$, $\mu_R(x) = \mu_{PS}(x)$ 가 된다.

4.2 퍼지 규칙 생성

TS 퍼지 규칙은 실제 데이터로부터 도출되므로 생성된 차분 $\Delta_{m(i)}N_0, \dots, \Delta_{m(i)}N_{N-m(i)-1}$ 이 만족시키는 경우에 한정하여 규칙이 생성된다. 따라서 얻어진 규칙의 수는 이론적인 상한보다 적을 수 있다.

우선 서로 인접한 차분 값 3개가 한 쌍을 이룬 각 입력 쌍 $\{\Delta_{m(i)}N_{N-m(i)+1}, \Delta_{m(i)}N_{N-m(i)}, \Delta_{m(i)}N_{N-m(i)-1}\}, \dots, \{\Delta_{m(i)}N_1, \Delta_{m(i)}N_2, \Delta_{m(i)}N_3\}, \{\Delta_{m(i)}N_0, \Delta_{m(i)}N_1, \Delta_{m(i)}N_2\}$ 에 대해 대응되는 규칙의 조건부를 생성하여 중복되는 것들을 제거하면 퍼지 예측 모델의 규칙기반을 구성할 모든 퍼지 규칙들이 생성된다. 예를 들어, 그림 3(a)에 표시한 것과 같은 입력 쌍 $\{\Delta_{m(i)}t_0, \Delta_{m(i)}t_1, \Delta_{m(i)}t_2\}$ 의 경우, $\Delta_{m(i)}t_0$ 는 퍼지 집합 NS 또는 ZE에 속하고, $\Delta_{m(i)}t_1$ 는 ZE 또는 PS에, 그리고

$\Delta_{m(i)}t_2$ 는 PS 또는 PB에 속하므로 이 입력 쌍에 대해서는 다음과 같이 8개의 규칙이 생성된다.

- R_1 : if $\Delta_{m(i)}t_0$ is NS and $\Delta_{m(i)}t_1$ is ZE and $\Delta_{m(i)}t_2$ is PS then ~
- R_2 : if $\Delta_{m(i)}t_0$ is NS and $\Delta_{m(i)}t_1$ is ZE and $\Delta_{m(i)}t_2$ is PB then ~
- R_3 : if $\Delta_{m(i)}t_0$ is NS and $\Delta_{m(i)}t_1$ is PS and $\Delta_{m(i)}t_2$ is PS then ~
- R_4 : if $\Delta_{m(i)}t_0$ is NS and $\Delta_{m(i)}t_1$ is PS and $\Delta_{m(i)}t_2$ is PB then ~
- R_5 : if $\Delta_{m(i)}t_0$ is ZE and $\Delta_{m(i)}t_1$ is ZE and $\Delta_{m(i)}t_2$ is PS then ~
- R_6 : if $\Delta_{m(i)}t_0$ is ZE and $\Delta_{m(i)}t_1$ is ZE and $\Delta_{m(i)}t_2$ is PB then ~
- R_7 : if $\Delta_{m(i)}t_0$ is ZE and $\Delta_{m(i)}t_1$ is PS and $\Delta_{m(i)}t_2$ is PS then ~
- R_8 : if $\Delta_{m(i)}t_0$ is ZE and $\Delta_{m(i)}t_1$ is PS and $\Delta_{m(i)}t_2$ is PB then ~

(6)

이러한 규칙 생성 또한 다중 모델을 구성하는 k 개의 퍼지 예측 모델 각각에 대해 독립적으로 수행된다.

4.3 퍼지 규칙 파라미터 식별

퍼지 규칙의 결론부의 선형 수식의 파라미터 식별에는 그 규칙의 조건부를 생성하는 모든 입력 쌍들을 데이터로 사용하게 되며, 본 논문에서는 파라미터 식별 기법으로 최적해를 제공하는 최소 자승법을 적용한다. 차분 간격 $m(i)$ 에 대한 i 번째 TS 퍼지 예측기의 j 번째 퍼지 규칙 R_j 의 생성에 기여한 입력 쌍 데이터들이 n 쌍이라면, 이들에 의한 결론부 선형식은 다음과 같이 n 개의 연립 방정식이 된다.

$$\begin{aligned}
 \widehat{\nabla}_{i1}^j &= a_0^j \Delta_{m(i)}^j t_{10} + a_1^j \Delta_{m(i)}^j t_{11} + a_2^j \Delta_{m(i)}^j t_{12} \\
 \widehat{\nabla}_{i2}^j &= a_0^j \Delta_{m(i)}^j t_{20} + a_1^j \Delta_{m(i)}^j t_{21} + a_2^j \Delta_{m(i)}^j t_{22} \\
 &\vdots \\
 \widehat{\nabla}_{in}^j &= a_0^j \Delta_{m(i)}^j t_{n0} + a_1^j \Delta_{m(i)}^j t_{n1} + a_2^j \Delta_{m(i)}^j t_{n2}
 \end{aligned} \tag{7}$$

이를 벡터-행렬식으로 표현하면

$$\begin{bmatrix} \widehat{\nabla}_{i1}^j \\ \widehat{\nabla}_{i2}^j \\ \vdots \\ \widehat{\nabla}_{in}^j \end{bmatrix} = \begin{bmatrix} \Delta_{m(i)}^j t_{10} & \Delta_{m(i)}^j t_{11} & \Delta_{m(i)}^j t_{12} \\ \Delta_{m(i)}^j t_{20} & \Delta_{m(i)}^j t_{21} & \Delta_{m(i)}^j t_{22} \\ \vdots & \vdots & \vdots \\ \Delta_{m(i)}^j t_{n0} & \Delta_{m(i)}^j t_{n1} & \Delta_{m(i)}^j t_{n2} \end{bmatrix} \begin{bmatrix} a_0^j \\ a_1^j \\ a_2^j \end{bmatrix} \tag{8.a}$$

$$Y_j = X_j \Theta_j \tag{8.b}$$

여기서 Y_j 는 출력 벡터, X_j 는 입력 데이터 행렬, Θ_j 는 계수 벡터이다. 식(8.b)의 계수 벡터 Θ_j 는 최소 자승법을 이용하여 다음과 같이 구할 수 있다.

$$\widehat{\Theta}_j = (X_j^T X_j)^{-1} X_j^T Y_j \tag{9}$$

식(9)와 같이 데이터로부터 구한 추정 계수는 다음과 같이 정의되는 오차 파워의 합을 최소화하는 최적해이다.

$$E_j = (Y_j - X_j \widehat{\Theta}_j)^T (Y_j - X_j \widehat{\Theta}_j) \tag{10}$$

5. 적응 모델 선택 메커니즘

앞서 구현된 다중 모델 예측기들은 입력 데이터의 특성이 시간에 따라 달라질 수 있기 때문에 각각의 순간에 대하여 예측 성능이 변화할 것이다. 따라서 본 논문에서는 데이터 특성에 따라 모델 선택이 가변적인 적응 메커니즘을 갖는 모델 선택 기법을 제안한다. 적응 모델 선택 기법의 어려움은 모델 선택을 위한 효율적이면서도 타당한 판별 기준을 어떻게 구현할 것이냐에 있는데, 서론에서 언급하였듯이

러프 집합을 이용하여 훈련 모드에서의 결과들을 기반으로 모델 선택에 필요한 지식을 발견하고 적당한 규칙을 생성함으로써 이를 해결하였다.

5.1 모델 선택을 위한 의사결정표 생성

러프 집합 이론은 전체 집합에 속한 모든 객체들이 어떤 정보(지식)와 연관되어 있다는 가정에서부터 출발하며, 이러한 지식을 간결하게 표현한 것이 의사결정표로서 각 객체에 대한 조건부 속성 값과 의사결정부 속성 값으로 표현된다.

의사 결정표의 조건부 속성으로는 입력 차분 데이터들이 퍼지 분할되는 입력 공간(전체 집합)에서의 위치를 사용하며, 그 속성 값은 그림 3(a)에 나타난 것처럼 A, B, C, D, E, F의 6가지 중 하나로 group(A), group(B), ..., group(F)로 주어진다. 의사결정부 속성은 조건부 속성으로 사용되는 입력 차분 데이터에 대해 예측 오차가 최소인 예측 모델로서, 그 속성 값은 model(1), model(2), ..., model(k)의 k개 중 하나가 된다. 예를 들어, 원 시계열 데이터 $y(t)$ 에 대한 퍼지 예측기의 입력 차분 데이터 쌍이 그림 2(a)에 나타난 것과 같이 $\Delta_{m(i)}t_0, \Delta_{m(i)}t_1, \Delta_{m(i)}t_2$ 라면, 그에 대한 속성 값은 각각 group(D), group(E), group(F)가 되고, 이 입력쌍에 대해 3번째 퍼지 예측기의 예측 오차가 가장 작다면 의사 결정부 속성 값은 model(3)으로 주어지게 된다. 이러한 방식으로 시계열 데이터 $\{y(t)\}$ 에 대해 차례로 속성 값을 결정하면 의사결정표가 완성된다. 표1은 모델 선택을 위한 의사결정표의 한 예이다.

표 1. 모델 선택을 위한 의사결정표 예
Table 1. Example of decision table for model selection

속성 객체	$\Delta_{m(i)}t_0$	$\Delta_{m(i)}t_1$	$\Delta_{m(i)}t_2$	optimal model
1	group(A)	group(E)	group(A)	model (1)
2	group(A)	group(E)	group(E)	model (1)
3	group(E)	group(E)	group(E)	model (2)
4	group(A)	group(A)	group(A)	model (2)
5	group(A)	group(A)	group(D)	model (3)
6	group(D)	group(D)	group(D)	model (4)
7	group(D)	group(D)	group(D)	model (4)
8	group(D)	group(D)	group(D)	model (5)
...

그런데, 의사결정표를 만들 때 문제는 조건부 속성으로 사용되는 차분 데이터의 선정이다. 하나의 원 시계열 데이터 $y(t)$ 에 대해 차분 간격은 $m(1), m(2), \dots, m(k)$ 의 k개가 있으며 그에 따라 차분 데이터 입력 쌍도 k개가 가능하므로, 조건부 속성 값을 결정하기 전에 먼저 입력 쌍을 선정해야만 한다. 제안된 논문에서는 훈련 데이터에 대해 다음과 같이 정의되는 자승 오차 평균(Mean Squared Error : MSE)을 최소화하는 차분 간격을 조건부 속성 값으로 사용하였다.

$$MSE = \frac{1}{N - m(i)} \sum_{n=m(i)+1}^N (y(n) - \hat{y}(n))^2 \quad (11)$$

여기서, $\hat{y}(n)$ 은 퍼지 예측기에서 출력한 $y(n)$ 의 예측 값이며, 차분 간격 $m(i)$ 인 퍼지 예측기 모델에서는 예측에 사용되는 차분 데이터가 $N - m(i)$ 개이므로 위와 같이 평균이 구해진다.

5.2 비일관적인 규칙의 확률적 처리

표 1에서 6, 7번째의 규칙과 8번째의 규칙과 같은 비일관적인 규칙들은 규칙의 감축 과정에서 삭제되어야 하지만, 비일관적인 경우를 모두 배제하게 되면 유용한 정보의 손실을 초래할 수 있다. 따라서 비일관적인 규칙 중에 일정 빈도 이상으로 같은 규칙을 생성하게 되면 유용한 데이터로 간주하여 규칙 생성에 활용함으로써 데이터의 비일관성을 최대한 줄이면서도 유용한 데이터가 손실되는 것을 방지하는 방법이 제안된 바 있다[11-12]. 본 논문에서도 이를 채택하여 보다 효율적인 모델 선택 규칙을 생성하고자 하였다.

비일관적인 데이터의 유용성을 확률적으로 고려하기 위해 도입되는 비일관적 규칙의 빈도에 따른 확률은 다음과 같이 정의된다.

$$pr(r_i) = \frac{card(r_i)}{card(r_j)} \quad (12)$$

여기서, $card(r_j)$ 는 r_i 와 조건부 속성이 같은 규칙들의 개수이고, $card(r_i)$ 는 조건부 속성과 의사결정부 속성이 모두 같은 규칙들의 개수이다. $pr(r_i)$ 가 임계치 $\beta(0.5 < \beta \leq 1)$ 보다 크면 r_i 그룹에 속한 비일관적인 규칙들을 일관적인 규칙으로 간주하여 모델 선택 규칙 생성을 위한 의사결정표에 잔류시킨다. 예를 들어, 표 1에서 조건부 속성 값이 같은 비일관적 규칙 6, 7, 8 중 의사결정부 속성 값이 model(4)인 규칙은 6과 7의 2개이며, model(5)인 규칙은 8 하나뿐이다. 따라서 의사결정부 속성 값이 model(4)가 될 확률 빈도는 2/3이 되며, model(5)가 될 확률 빈도는 1/3이 되므로, $\beta = 0.6$ 으로 설정한다면 규칙 8은 불필요한 규칙으로 간주되어 제거되고 규칙 6과 7은 유용한 규칙으로 의사결정표에 잔류하여 모델 선택 규칙 생성에 이용된다.

5.3 코어(core)에 의한 규칙 감축

확률처리를 통해 감축된 규칙을 최소의 모델 선택 규칙 기반으로 만들기 위해선 중복되거나 불필요한 지식의 감축이 필요하며, 이에 대하여 기본적인 역할을 하는 것이 코어와 리덕트로서, 리덕트는 현재의 지식 내에서 나타나는 모든 기본 범주들을 정의하기에 충분한 지식의 필수적인 부분이고, 코어는 지식의 가장 특징적인 부분의 집합으로서 지식을 감축할 때 꼭 필요한 요소가 된다.

코어를 이용하여 규칙을 감축하는 과정을 표 1을 이용하여 살펴보기로 하자. 표 1에서 각각의 속성을 하나씩 배제하면서 규칙들을 비교하여 비일관적인 규칙이 발생하면 배제된 속성의 속성 값은 그 규칙의 코어가 된다. 예를 들어, 속성 $\Delta_{m(i)}t_2$ 를 배제했을 때, 속성 $\Delta_{m(i)}t_0$ 와 $\Delta_{m(i)}t_1$ 에 대해 각각의 규칙을 비교하여 보면 규칙 4와 5에서 비일관적인 규칙이 형성된다. 이는 속성 $\Delta_{m(i)}t_2$ 의 값을 배제하고는 규칙을 식별할 수 없다는 것을 의미하므로 규칙 4와 5의 속성 $\Delta_{m(i)}t_2$ 의 속성 값은 그 규칙을 표현하는 데 있어서 없어서는 안 될 중요한 속성, 즉 코어가 되며, 이는 규칙 표현의 최소 단위라고 할 수 있다. 같은 방법으로 코어를 구하게 되면 표 1의 굵게 표시된 속성 값들인 코어를 구할 수 있으며 이 코어를 이용하여 규칙을 감축한 결과는 표 2와 같다.

표 2. 코어에 의한 규칙의 감축

Table 2. Rule reduction by using CORE attribution

규칙	$\Delta_{m(i)t_0}$	$\Delta_{m(i)t_1}$	$\Delta_{m(i)t_2}$	optimal model
1		group(E)		model (1)
2	group(A)			model (1)
3	group(E)			model (2)
4		group(A)	group(A)	model (2)
5			group(D)	model (3)

표 2의 의사결정표가 최종적인 모델 선택 규칙기반으로서, 차분 데이터의 속성 값에 따라 예측을 수행할 최적 모델을 선택하게 해준다. 만약 조건부 속성을 결정하는 차분 데이터 쌍이 표 2의 모델 선택 규칙기반을 만족하지 못할 경우에는 MSE를 최소화하는 퍼지 예측기를 그대로 사용하여 예측이 수행되도록 한다.

6. 오차 보정 메커니즘

예측 오차는 일반적으로 정규 분포 백색 잡음의 특성을 갖는 확률 프로세스가 되므로 그 보정이 쉽지 않다. 본 논문에서는 예측을 위해 인가되는 입력 쌍이 만족시키는 퍼지 규칙 R_j 의 생성에 기여한 훈련 데이터들 중에서 입력 쌍과 가장 유사한 특성의 훈련 데이터로부터 발생된 예측 오차만큼을 보정 오차로 사용하게 된다. 이러한 방법은 완벽하진 않지만 아주 상당한 정도로 예측 오차를 줄일 수 있을 것이다. 이러한 유사성을 판별하는 방법으로는 교차 상관 계수 [13]를 이용하였다. 입력 쌍 $X_t = [\Delta_{m(i)t_0}, \Delta_{m(i)t_1}, \Delta_{m(i)t_2}]$ 와 훈련 데이터 $T_n = [\Delta_{m(i)n_0}, \Delta_{m(i)n_1}, \Delta_{m(i)n_2}]$ 의 교차 상관 계수 ρ_{XT} 는 다음과 같이 정의된다.

$$\rho_{XT} = \frac{C_{XT}}{\sqrt{C_{XX}} \sqrt{C_{TT}}} \quad (13)$$

여기서 C_{XX} 는 X_t 의 공분산, C_{TT} 는 T_n 의 공분산, 그리고 C_{XT} 는 X_t 와 T_n 의 교차 공분산으로서, X_t 의 평균을 \bar{X}_t , T_n 의 평균을 \bar{T}_n 이라고 하면 다음과 같이 계산된다.

$$C_{XX} = \sum_{l=0}^2 (\Delta_{m(i)t_l} - \bar{X}_t)^2 \quad (14)$$

$$C_{TT} = \sum_{l=0}^2 (\Delta_{m(i)n_l} - \bar{T}_n)^2 \quad (15)$$

$$C_{XT} = \sum_{l=0}^2 (\Delta_{m(i)t_l} - \bar{X}_t)(\Delta_{m(i)n_l} - \bar{T}_n) \quad (16)$$

식(13)의 교차 상관 계수는 $-1 \leq \rho_{XT} \leq 1$ 이고, $|\rho_{XT}|=0$ 이면 무상관을 의미하며, $|\rho_{XT}|=1$ 이면 완전한 선형 관계를 의미하게 된다. 그러므로 교차 상관 계수의 크기가 클수록 입력 쌍과 훈련 데이터 간의 유사성이 크다고 볼 수 있으므로 퍼지 규칙 R_j 의 생성에 기여한 훈련 데이터들 중 입력 쌍과의 교차 상관 계수가 가장 큰 훈련 데이터에 의해 발생하는 예측 오차를 선택하여 오차 보정 값 \hat{e}_t^j 로 사용한다. 그러면, 입력 쌍 X_t 에 대해 퍼지 규칙 R_j 의 결론부에 의한 출력은 다음과 같이 계산된다.

$$\hat{\nabla}_t^j = \hat{a}_0^j \Delta_{m(i)t_0} + \hat{a}_1^j \Delta_{m(i)t_1} + \hat{a}_2^j \Delta_{m(i)t_2} + \hat{e}_t^j \quad (17)$$

이때 최대 교차 상관 계수가 임계치(0.5) 이하일 경우에는 두 데이터 간의 유사성이 존재하지 않는 것으로 간주하여 오차 보정을 하지 않는다. 총 q 개의 퍼지 규칙을 갖는 퍼지 예측기의 입력 쌍 X_t 에 대한 출력 $\hat{\nabla}_t$ 은 각각의 규칙 R_j 의 전건부에서 결정되는 적합도 μ_j 와 후건부에서 (17)을 이용하여 결정된 $\hat{\nabla}_t^j$ 로부터 다음과 같이 가중 합으로 구해지며 원하는 시계열의 예측값은 식(19)와 같이 구할 수 있다.

$$\hat{\nabla}(t) = \frac{\sum_{i=1}^q \mu_i \hat{\nabla}_t^i}{\sum_{i=1}^q \mu_i} \quad (18)$$

$$\hat{y}(t+p) = y(t) + \hat{\nabla}(t) \quad (19)$$

7. 컴퓨터 시뮬레이션 및 검토

3개의 주요한 시계열을 이용하여 기존의 연구 결과들과 같은 크기의 훈련데이터, 예측데이터 및 같은 예측구간(one step ahead)으로 예측을 수행하여 성능을 비교하였다.

사례 1 : 분기별 호주의 전력 생산량 예측[14]

호주의 전력생산량 데이터의 총 데이터 155개 중 70개를 훈련 데이터로 사용하고 나머지를 예측데이터로 사용하였다. 훈련 모드에서 만들어진 의사결정표는 아래 표 3과 같고, 표 4는 확률처리에 의한 규칙 감축이며, 표 5는 코어를 이용한 최종 감축 규칙을 나타낸다.

표 3. 모델 선택을 위한 의사결정표

Table 3. Decision table for model selection

규칙	$\Delta_{m(i)t_0}$	$\Delta_{m(i)t_1}$	$\Delta_{m(i)t_2}$	model	규칙	$\Delta_{m(i)t_0}$	$\Delta_{m(i)t_1}$	$\Delta_{m(i)t_2}$	model
R1	group(E)	group(E)	group(D)	model(1)	R26	group(E)	group(E)	group(E)	model(4)
R2	group(D)	group(P)	group(P)	model(1)	R27	group(P)	group(P)	group(P)	model(4)
R3	group(E)	group(P)	group(E)	model(1)	R28	group(E)	group(E)	group(P)	model(4)
R4	group(E)	group(C)	group(C)	model(1)	R29	group(E)	group(E)	group(E)	model(4)
R5	group(P)	group(P)	group(D)	model(1)	R30	group(P)	group(P)	group(E)	model(4)
R6	group(E)	group(D)	group(D)	model(2)	R31	group(C)	group(C)	group(D)	model(4)
R7	group(P)	group(P)	group(E)	model(2)	R32	group(D)	group(D)	group(D)	model(4)
R8	group(D)	group(E)	group(P)	model(2)	R33	group(D)	group(D)	group(D)	model(5)
R9	group(D)	group(D)	group(E)	model(2)	R34	group(E)	group(E)	group(D)	model(5)
R10	group(D)	group(D)	group(D)	model(2)	R35	group(D)	group(D)	group(E)	model(5)
R11	group(E)	group(D)	group(P)	model(2)	R36	group(D)	group(D)	group(D)	model(6)
R12	group(E)	group(D)	group(E)	model(2)	R37	group(E)	group(E)	group(D)	model(6)
R13	group(E)	group(D)	group(P)	model(2)	R38	group(E)	group(P)	group(E)	model(6)
R14	group(D)	group(D)	group(D)	model(2)	R39	group(D)	group(D)	group(C)	model(6)
R15	group(P)	group(P)	group(P)	model(2)	R40	group(E)	group(E)	group(D)	model(7)
R16	group(D)	group(D)	group(D)	model(3)	R41	group(D)	group(E)	group(E)	model(7)
R17	group(P)	group(P)	group(P)	model(3)	R42	group(D)	group(D)	group(P)	model(7)
R18	group(E)	group(D)	group(E)	model(3)	R43	group(E)	group(P)	group(P)	model(8)
R19	group(E)	group(E)	group(D)	model(3)	R44	group(D)	group(E)	group(D)	model(8)
R20	group(P)	group(E)	group(E)	model(3)	R45	group(P)	group(P)	group(E)	model(8)
R21	group(P)	group(E)	group(E)	model(3)	R46	group(P)	group(E)	group(D)	model(8)
R22	group(P)	group(E)	group(E)	model(3)	R47	group(P)	group(P)	group(E)	model(8)
R23	group(C)	group(D)	group(P)	model(3)	R48	group(P)	group(P)	group(E)	model(8)
R24	group(C)	group(D)	group(D)	model(4)	R49	group(D)	group(E)	group(E)	model(8)
R25	group(P)	group(E)	group(D)	model(4)	R50	group(P)	group(P)	group(P)	model(8)

표 4. 확률적 처리에 의한 비일관적 규칙의 감축

Table 4. Inconsistent rule reduction by using probabilistic process

규칙	$\Delta_{m(i)t_0}$	$\Delta_{m(i)t_1}$	$\Delta_{m(i)t_2}$	model	규칙	$\Delta_{m(i)t_0}$	$\Delta_{m(i)t_1}$	$\Delta_{m(i)t_2}$	model
R1	group(D)	group(P)	group(P)	model(1)	R11	group(E)	group(E)	group(P)	model(4)
R2	group(B)	group(C)	group(C)	model(1)	R12	group(C)	group(C)	group(D)	model(4)
R3	group(P)	group(P)	group(D)	model(1)	R13	group(D)	group(D)	group(D)	model(4)
R4	group(E)	group(E)	group(E)	model(2)	R14	group(D)	group(D)	group(B)	model(5)
R5	group(D)	group(D)	group(E)	model(2)	R15	group(B)	group(B)	group(C)	model(6)
R6	group(E)	group(D)	group(P)	model(2)	R16	group(E)	group(E)	group(D)	model(7)
R7	group(P)	group(E)	group(E)	model(3)	R17	group(P)	group(P)	group(D)	model(7)
R8	group(E)	group(P)	group(P)	model(3)	R18	group(D)	group(D)	group(D)	model(8)
R9	group(C)	group(D)	group(P)	model(3)	R19	group(D)	group(P)	group(E)	model(8)
R10	group(C)	group(D)	group(P)	model(3)	R20	group(D)	group(E)	group(D)	model(8)
R21	group(E)	group(E)	group(E)	model(4)					

표 5. 코어를 이용한 규칙기반의 감축

Table 5. rule base reduction by using CORE attribution

속성 규칙	Δ_{m1}	Δ_{m2}	Δ_{m3}	model	속성 규칙	Δ_{m1}	Δ_{m2}	Δ_{m3}	model
R1	group(D)	group(F)		model(1)	R9	group(E)			model(4)
R2		group(C)		model(1)	R10	group(E)	group(B)		model(4)
R3		group(F)		model(1)	R11		group(D)	group(B)	model(4)
R4	group(D)	group(E)	group(B)	model(2)	R12	group(D)	group(B)	group(E)	model(2)
R5		group(D)	group(E)	model(2)	R13		group(D)		model(2)
R6	group(E)	group(D)		model(2)	R14	group(E)	group(P)		model(6)
R7	group(F)			model(3)	R15		group(D)		model(6)
R8	group(C)			model(3)					

이상의 결과에서 볼 수 있듯이, 러프 집합을 이용한 규칙 감축이 매우 효율적임을 알 수 있다. 그림 4는 제안된 예측 기법을 이용한 예측 결과로서, 검은색은 원 시계열의 값이며 파란색으로 표시된 값이 제안된 기법에 의한 예측 값으로 매우 양호한 예측 성능을 보이고 있음을 알 수 있다.

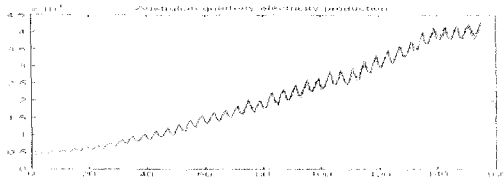


그림 4. 분기별 호주의 전력 생산량 예측
Figure 4. Australian quarterly electricity Production prediction

데이터 전처리 과정을 통해 선정된 최적 차분 간격 후보군은 표 6에 나타낸 9개로서, 이에 대해 구성된 다중 모델 퍼지 예측기 각각의 예측 성능과 제안된 적응 메커니즘을 이용한 예측 성능의 비교 결과를 표 6에 제시하였다. 성능을 비교 분석하기 위한 지표로는 다음과 같이 정의되는 상대 오차 평균(Mean Relative Error : MRE)을 사용하였다.

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

표 6. 모델 선택전과 모델선택 후의 예측 성능 비교
Table 6. Comparison of results before and after model selection

성능 지표	상관계수	사 모델에 의한 MRE	적용 모델 선택에 의한 MRE
차분 값 8	0.6729	1.6892	1.4930
차분 값 2	0.8514	2.0466	
차분 값 1	0.9188	2.4907	
차분 값 4	0.8370	2.9462	
차분 값 6	0.6836	3.1188	
차분 값 3	0.8826	3.5730	
차분 값 5	0.7576	3.3911	
차분 값 7	0.6853	4.4249	
차분 값 9	0.6946	5.6841	

표 6에서 보면, 적용 모델 선택을 한 경우가 각각의 예측 기들의 성능을 잘 반영함으로써 성능이 향상됨을 알 수 있다. 표 7은 기존의 예측 기법들[3, 5, 6, 15]과 제안된 기법의 예측 성능을 비교한 것으로, 이 경우에도 역시 제안된 기법이 예측 성능이 가장 양호함을 알 수 있다.

표7. 기존 기법들과의 예측 성능 비교
Table 7. Comparison between other methods and ours

	Mamdani 퍼지 모델	다중 퍼지 모델	Fuzzy AR	GA-RS 방법	제안된 기법
MRE	7.8123	2.7125	3.1254	1.8100	1.4930

사례 2 : 월별 Erie 호수의 수위[14]

월별 Erie 호수의 수위 데이터의 총 600개의 데이터 중 400개를 훈련 데이터로 사용하였다. 표 8은 제안된 기법에 의해 생성된 모델 선택 규칙이며, 그림 5는 제안된 기법에 의한 예측 결과로서 검은색은 원 시계열, 파란색은 예측 값이다.

표 8. 모델 선택 규칙

Table 8. Model selection rules

속성 규칙	Δ_{m1}	Δ_{m2}	Δ_{m3}	model
R1	group(D)	group(F)		model(1)
R2		group(C)		model(1)
R3		group(F)		model(1)
R4	group(D)	group(E)	group(E)	model(2)
R5		group(D)	group(E)	model(2)
R6	group(E)	group(D)		model(2)
R7	group(F)			model(3)
R8	group(C)			model(3)

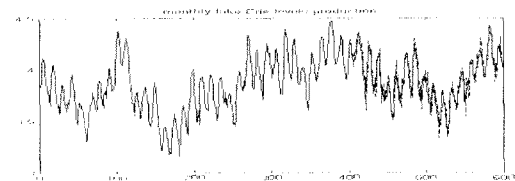


그림 5. 월별 Erie 호수의 수위 예측
Figure 5. monthly lake Erie levels prediction

표 9는 제안된 기법과 기존의 기법[16]의 예측 성능을 비교한 것으로, 성능 지표로는 다음과 같은 RMSE(Root Mean Square Error)를 사용하였다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (21)$$

표 9. 기존 기법들과의 예측 성능 비교
Table 9. Comparison between other methods and ours

	ANN	hybrid ARIMA-ANN	ours
RMSE	0.1045	0.0820	0.0806

표 9는 퍼지 전문가 시스템에 의한 38개의 규칙기반을 이용한 hybrid ARIMA-ANN방식 보다 제안된 기법이 훨씬 적은 수(8개)의 모델 선택 규칙에 의해서도 더 양호한 예측 성능을 보임을 알 수 있다.

사례 3 : Mackey-Glass 시계열 데이터

Mackey-Glass 시계열은 비선형 시계열 예측에 빈번히 쓰이는 데이터로 다음과 같이 정의되는 수식으로부터 발생된다.

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x^{10}(t-\tau)} - bx(t) \quad (22)$$

위 식의 파라미터 값으로 $a=0.2$, $b=0.1$, $\tau=17$ 을 사용하였다. 데이터는 $x(124)$ 부터 $x(1123)$ 까지 1000개의 데이터를 추출하여 500개의 데이터를 훈련 데이터로 사용하였다. 표 10은 최종적으로 생성된 모델 선택 규칙이며, 그림 6는 예측 결과이다.

표 10. 모델 선택 규칙

Table 10. Model selection rules

규칙 \ 측정	Δ_{m1}	Δ_{m2}	Δ_{m3}	model
R1		group(E)		model(1)
R2		group(D)		model(2)

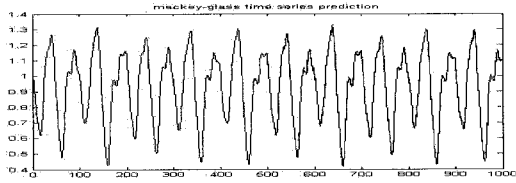


그림 6. Mackey-Glass 시계열 예측

Figure 6. Mackey-Glass Time Series prediction

표 11은 이 시계열 예측에 대해 Wang과 Mendel의 방법 [17], MCM-1과 MCM-2[18]와 제안된 방법의 성능을 비교한 것으로서 제안된 방법이 훨씬 적은 수의 입력 공간 퍼지 분할 영역(퍼지 집합)을 가지고서도 더 나은 예측 성능을 보임을 알 수 있다.

표 11. 기존 기법들과의 예측 성능 비교

Table 11. Comparison between other methods and ours

입력 퍼지 집합 수	RMSE			
	WANG	MCM-1	MCM-2	제안된 방법
5				0.0083
7	0.0372	0.0374	0.0340	-
11	0.0253	0.0265	0.0235	-
15	0.0191	0.0197	0.0187	-
19	0.0161	0.0162	0.0159	-
23	0.0135	0.0142	0.0131	-
27	0.0115	0.0115	0.0113	-
31	0.0108	0.0108	0.0105	-

8. 결 론

본 논문에서 제안한 데이터의 전처리 과정은 적절한 최적 차분 간격 후보군을 선정하여 이를 이용해 다중 모델 예측기를 모델링함으로써 비정상 또는 혼돈 시계열의 특성들을 잘 반영할 수 있도록 하였으며, 또한 이러한 다중모델들의 각각의 예측 성능들을 러프 집합을 이용해 도출된 모델 선택 규칙을 통해 매 데이터 순간마다 최적 모델이 선택되어 예측이 이루어지도록 한 적응 모델선택 방식을 통해 각 예측기의 특성을 최대한 고려될 수 있도록 하였다. 마지막으로 예측시 발생하는 오차에 대하여 상관해석을 기반으로 적절하게 오차를 보정함으로써 예측기의 성능을 높이고자 하였다. 시뮬레이션 결과들은 제안된 방법의 유효성을 보여주는 것으로 예상한대로 까다로운 시계열 예측 문제에 대해서도 좋은 예측 성능을 보이며 기존의 연구 결과들에 비해 개선된 예측 성능을 가짐을 나타내고 있다. 그러므로 제안된 방법은 비정상 또는 혼돈 시계열과 같이 유동적 비선형성을 나타내는 시계열 데이터 예측에 매우 유용하게 사용될 수 있을 것으로 기대된다.

참 고 문 헌

[1] George J. Klir and Bo Yuan, *Fuzzy Sets and*

Fuzzy Logic Theory and Applications, Prentice-Hall, 1995.

[2] G.Janazcek, L.Swift, *Time Series Forecasting, Simulation, Applications*, Ellis Horwood, 1993.

[3] K.Ozawa, T.Niimura, "Fuzzy Time-Series Model of Electric Power Consumption", *IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 1195-1198, 1999.

[4] Juhong Nie, "Nonlinear Time-Series Forecasting A Fuzzy Neural Approach", *Neuro computing*, vol.16, pp.66-76, MacMaster University, 1997.

[5] Daijin Kim, Chulhyun 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 Fuzzy Systems*, vol. 2, pp. 703-707, 1999.

[7] C. H. Lee, S. H. Yoon, "Fuzzy Nonlinear Time-Series Forecasting with Data Preprocessing and Model Selection", *Journal of Telecommunications and Information*, vol. 5, pp. 232-238, 2001.

[8] Z. Pawlak, "Rough Sets", *Int. J. of Computer and Information Sciences*, vol. 11, pp.341-356, 1982.

[9] 변증남, 방원철, *러프집합의 이론과 응용*, 청문각, 1999.

[10] Carl G. Looney, "Pattern Recognition Using Neural Networks", *Oxford University Press*, 1997.

[11] 권은아, 김흥기, "학률적 러프집합에 기반한 근사 규칙의 간결화", *정보처리학회 논문지(D) 제 8-D권 제 3호* pp. 203-210, 2001.

[12] C. H. Lee, S. H. Seo, "Efficient Extraction of Hierarchically Structured Rules Using Rough Sets", *International Journal of fuzzy Logic and Intelligent System*, Vol. 4, no. 2, pp. 205-210, september, 2004.

[13] 김해경, 김태수, *시계열 분석과 예측 이론*, 경문당, 2003.

[14] <http://www-personal.buseco.monash.edu.au>

[15] 주용석, *유전알고리즘과 러프집합을 이용한 퍼지 시스템 모델링*, 강원대학교 석사학위논문, 2003.

[16] O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L. J. Herrera, A. Guillen, L. Marquez, M. Pasadas, "Hybridization of intelligents and ARIMA models for time series prediction", *Fuzzy Sets and Systems*, vol. 159, pp. 821-845, April 2008.

[17] L. X. Wang, J. M. Mendel, "Generating fuzzy rules from numerical data, with applications", *IEEE Trans. on Systems, Man, and Cybern*, 22, No. 6, pp. 1414-1427, 1992.

[18] 김인택, 공창욱, "시계열 예측을 위한 퍼지 학습알고리즘", *한국 퍼지 지능시스템 학회*, vol.7, No.3, pp34-42, 1999.

저 자 소 개



방영근(Young-Keun Bang)
2000년: 강원대학교(삼척) 공학사.
2003년: 강원대학교(삼척) 공학석사
2004년~현재: 강원대학교(춘천) 전기전
자공학과 박사과정

관심분야 : 지능제어, 최적화, 데이터마이닝
E-mail : b2y2c1@hanmail.net



이철희(Chul-Heui Lee)
1983년: 서울대학교 공학사.
1985년: 서울대학교 공학석사
1989년: 서울대학교 전기공학과(공학박사)
1994년~1995년: 미국 IONA대 방문교수
1990년~현재: 강원대학교 전기전자공학부
교수

관심분야 : Intelligent System & Soft computing, 제어
및 신호처리
E-mail : chlee@kangwon.ac.kr