

러프집합을 이용한 모델선택을 갖는 퍼지 시계열 예측

방영근, 이철희
강원대학교

Fuzzy Time Series Forecasting with Model Selection by using Rough Set

Young-Keun Bang, Chul-Heui Lee
Kangwon National University

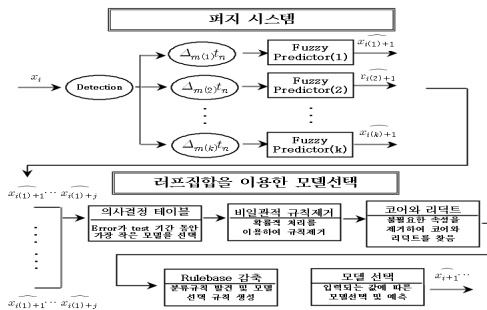
Abstract -본 논문에서는 유동적 비정상 시계열의 패턴과 규칙성을 잘 반영할 수 있는 최적의 차분 간격 후보군을 이용한 TS 퍼지 모델로 다중 퍼지 모델을 구현하였고, 각각의 모델들의 예측 특성을 반영하기 위하여 러프집합을 이용한 모델선택법을 제안하였다. 또한 TS퍼지 모델의 파라미터 식별에는 적절한 오차보정 메커니즘을 추가하여 더욱 예측 성능을 향상 시켰다.

1. 서 론

예측 시스템의 구현을 위해 퍼지 이론은 상당히 효과적이며, 비정상 시계열의 차분 데이터를 이용하면 더욱 예측성능을 향상시킬 수 있다.[1] 따라서 본 논문에서는 비정상 시계열의 전처리 과정을 통해 평균이나 분산의 변화가 안정화된 최적의 차분 간격 후보군을 선정하여 각각의 차분간격 후보군에 대한 적절한 오차보정 메커니즘을 갖는 TS 퍼지 모델들로 구성된 다중 퍼지 모델 예측 시스템을 구현하였다. 또한 러프집합[2]을 이용해 생성된 모델 선택 규칙에 따라 예측을 수행하는 퍼지 모델이 선택되게 함으로써 각각의 퍼지 모델의 예측 특성을 최대한 반영할 수 있도록 하여 성능을 향상 시켰다.

2. 제안된 퍼지 시계열 예측 시스템의 구조

그림 1은 제안된 알고리즘의 구조이며, 다음과 같이 수행된다.



<그림 1> 제안된 알고리즘의 구조

- 1단계 : 비정상 시계열의 전처리 및 최적 차분 간격 후보군 선정
- 2단계 : 오차보정 메커니즘 TS 모델에 의한 다중 퍼지 모델 구현
- 3단계 : 러프집합을 이용한 퍼지모델 선택 규칙 생성
- 4단계 : 모델선택 규칙에 따라 선택된 모델의 예측 수행

3. 데이터의 전처리

1단계에서는 자기 상관 계수를 구하여 두 상관 계수의 값의 차가 가장 큰 상관 값 이상의 차분간격을 최적 차분 간격 후보군으로 결정한다.

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (1)$$

여기서, N 은 훈련 데이터의 길이, k 는 차분 간격, x_i 는 i 번째 훈련 데이터, \bar{x} 는 훈련 데이터의 평균이다.

$$\begin{aligned} \Delta_{m(k)}t_1 &= x_t - x_{t-m(k)} \\ \Delta_{m(k)}t_2 &= x_{t-1} - x_{t-1-m(k)} \\ &\vdots \\ \Delta_{m(k)}t_n &= x_{t-n+1} - x_{t-n+1-m(k)} \end{aligned} \quad (2)$$

위의 식 (2)는 선정된 차분 간격값 $\{m(k)\}$ 에 대한 차분데이터들은 생성하는 방법이다. 이렇게 생성된 각각의 차분 간격 값을 갖는 데이터들은 다중 퍼지 모델의 각각의 입력으로 사용되며, 또한 TS모델 후건부의 파라미터 식별에 이용된다.

4. TS 퍼지 예측기 설계

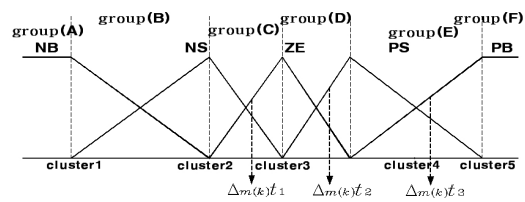
차분 간격 $m(k)$ 에 대해 예측이 수행되는 시간 t 에서 가장 최근의 차분 데이터 $\Delta_{m(k)}t_1, \Delta_{m(k)}t_2, \Delta_{m(k)}t_3$ 의 3개를 입력 변수로 하는 i 번째 TS 퍼지 예측 모델의 언어적 규칙은 다음과 같은 형태가 된다.

$$R_i : \text{if } \Delta_{m(k)}t_1 \text{ is } A_i \text{ and } \Delta_{m(k)}t_2 \text{ is } B_i \text{ and } \Delta_{m(k)}t_3 \text{ is } C_i \quad (3) \\ \text{then } \hat{\nabla}_t^i = a_0 + a_1 \Delta_{m(k)}t_1 + a_2 \Delta_{m(k)}t_2 + a_3 \Delta_{m(k)}t_3$$

여기서 후건부의 출력은 $\hat{\nabla}_t^i = \hat{y}_{(t+p)}^i - y_t$ 로서 현재 순간의 데이터 y_t 와 i 번째 규칙에 의해 결정되는 예측하고자 하는 p 스텝 앞의 예측 값 $\hat{y}_{(t+p)}^i$ 와의 차분 값이다.

4.1 퍼지 규칙 및 모델 선택 규칙 생성

데이터의 전처리과정에서 선정된 각각의 차분 데이터들에 대하여 최소값과 최대값 사이를 퍼지 분할의 전체 영역으로 하고, 입력 변수 데이터에 대해 k-means 클러스터링 알고리즘을 이용하여 그림 3과 같이 NB, NS, ZE, PS, PB의 5개 퍼지 집합으로 분할하였으며, 입력 값 x 에 대한 소속함수 값은 다음과 같이 구한다.



<그림 2> 입력 공간의 퍼지 분할과 소속함수 설정

$$\begin{aligned} \text{group(A, F)} &: \mu_{NB}(x) \text{ or } \mu_{PB}(x) = 1 \\ \text{group(B-E)} &: \mu_L(x) = \frac{C_R - x}{C_R - C_L}, \mu_R(x) = \frac{x - C_L}{C_R - C_L} \end{aligned} \quad (4)$$

여기서 $\mu_L(x)$ 는 x 가 속한 구간의 좌측 클러스터 중심값 C_L 을 중심값으로 하는 퍼지 집합에 대한 소속 함수 값을 나타내며, $\mu_R(x)$ 는 우측 클러스터 중심값 C_R 을 중심값으로 하는 퍼지 집합에 대한 소속 함수 값을 나타낸다. 또한, 그림 2의 입력 쌍은 8개의 퍼지 규칙을 생성하며, 다음과 같은 형태의 하나의 모델선택 규칙의 조건부를 생성한다.

$$\text{규칙1} : \{\Delta_{m(k)}t_1, \Delta_{m(k)}t_2, \Delta_{m(k)}t_3\} - \{\text{group(C), group(D), group(E)}\}$$

4.2 퍼지규칙 파라미터 식별

TS 퍼지모델 후건부의 선형 수식의 파라미터 식별에는 그 규칙에 포함되는 모든 입력 쌍들에 대하여 최소자승법을 적용한다. 차분 간격 $m(k)$ 에 대한 k 번째 TS 퍼지 예측기의 j 번째 퍼지 규칙 R_j 의 생성에 기여한 l 개의 입력 쌍 데이터들의 후건부 선형식은 다음과 같다.

$$\widehat{\nabla}_{t_n}^j = a_0^j \Delta_{m(k)}t_1^l + a_1^j \Delta_{m(k)}t_2^l + a_2^j \Delta_{m(k)}t_3^l \quad (5)$$

여기서 l 은 $\{1, 2, \dots, n\}$ 이다.
이를 벡터-행렬식으로 표현하면

$$\begin{bmatrix} \widehat{\nabla}_p^j(1) \\ \widehat{\nabla}_p^j(2) \\ \vdots \\ \widehat{\nabla}_p^j(n) \end{bmatrix} = \begin{bmatrix} \Delta_{m(k)} t_1^j(1) & \Delta_{m(k)} t_2^j(1) & \Delta_{m(k)} t_3^j(1) \\ \Delta_{m(k)} t_1^j(2) & \Delta_{m(k)} t_2^j(2) & \Delta_{m(k)} t_3^j(2) \\ \vdots & \vdots & \vdots \\ \Delta_{m(k)} t_1^j(n) & \Delta_{m(k)} t_2^j(n) & \Delta_{m(k)} t_3^j(n) \end{bmatrix} \begin{bmatrix} a_0^j \\ a_1^j \\ a_2^j \end{bmatrix} \quad (6)$$

$$Y_j = X_j \Theta_j \quad (7)$$

여기서 Y_j 는 출력 벡터, X_j 는 입력 데이터 행렬, Θ_j 는 계수 벡터이며 Θ_j 는 최소 자승법을 이용하여 다음과 같이 구할 수 있다.

$$\widehat{\Theta}_j = (X_j^T X_j)^{-1} X_j^T Y_j \quad (8)$$

4.3 오차 보정

오차의 보정은 입력 쌍과 입력쌍이 만족하는 규칙에 포함된 훈련 데이터들의 유사성을 교차 상관 계수를 이용하여 분석하여 상관성이 높은 오차를 적용하였다. 예측을 위한 입력 쌍 X_i 와 데이터쌍 T_n 의 교차 상관 계수 ρ_{XT} 는 다음과 같이 정의된다.

$$\rho_{XT}^i = \frac{C_{XT}^i}{\sqrt{C_{XX}^i} \sqrt{C_{TT}^i}} \quad (9)$$

여기서 C_{XT} 는 X_i 와 T_n 의 교차 공분산, C_{XX} , C_{TT} 는 각각의 공분산이며, 이렇게 구하여진 상관 계수 중 가장 큰 값의 오차를 오차 보정 값 \widehat{e}_i^j 로 사용하여 후건부 출력을 다음과 같이 계산한다.

$$\widehat{\nabla}_i^j = \widehat{a}_0^j \Delta_{m(k)} t_1 + \widehat{a}_1^j \Delta_{m(k)} t_2 + \widehat{a}_2^j \Delta_{m(k)} t_3 + \widehat{e}_i^j \quad (10)$$

총 q 개의 퍼지 규칙을 갖는 입력 쌍 X_i 에 대한 출력 $\widehat{\nabla}_i$ 은 소속함수 값과 위 (10)식 이용하여 구할 수 있다.

$$\widehat{\nabla}_i = \frac{\sum_{i=1}^q \mu_i \widehat{\nabla}_i^i}{\sum_{i=1}^q \mu_i}, \quad \widehat{y}_{(t+p)} = y_t + \widehat{\nabla}_i \quad (11)$$

$\widehat{\nabla}_i$ 은 현재와 미래 값의 증가분이므로 최종 예측 값 $\widehat{y}_{(t+p)}$ 는 식(11)과 같이 구할 수 있다.

5. 모델 선택

4.1절에 의해 생성된 모델 선택 규칙들은 다음과 같은 방법으로 감축되어 최소의 규칙으로 최적성능의 모델을 선택할 수 있게 된다.

5.1 의사결정표

먼저 4.1절의 각각의 다중 모델에서 생성된 모델 선택 규칙 중 비교구간에서 MSE(mean square error)를 최소로 하는 모델에서 생성된 규칙을 조건부로 이용하며, 그때의 예측값들 중 가장 좋은 예측성능을 보이는 모델을 이용하여 의사결정부를 만들면 다음의 표와 같이 생성된다.

〈표 1〉 모델선택을 위한 의사결정표

속성 규칙	$\Delta_{m(k)}(1)$	$\Delta_{m(k)}(2)$	$\Delta_{m(k)}(3)$	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 (3)
7	group(D)	group(D)	group(D)	model (3)
8	group(D)	group(D)	group(D)	model (4)

5.2 비일관적 규칙의 확률적 처리 및 코어에 의한 규칙 감축

위의 표 1에서의 6,7번째의 규칙과 8번째의 규칙은 아래의 확률빈도를 이용하여 유용한 규칙과 불필요한 규칙으로 구분되며, 불필요한 규칙은 제거하고 유용한 규칙이 손실되는 것을 방지할 수 있다.

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

여기서, $card(r_i)$ 는 r_i 와 조건부 속성이 같은 규칙들 개수이고 $card(r_j)$ 는 조건부와 의사결정부가 모두 같은 규칙들 개수이며, 임계치 β ($0.5 < \beta \leq 1$)보다 작으면 제거하여 감축하게 되고, 감축된 규칙을 이용하여 코어를 탐색하며 방법은 다음과 같다. 먼저 규칙 1에 대하여 각각의 속성에 대한 식별 불가능한 규칙들을 찾으면 다음과 같다.

속성 집합 :

$$\{\Delta_{m(k)}(1), \Delta_{m(k)}(2), \Delta_{m(k)}(3)\} = \{\{1, 2, 4, 5\}, \{1, 2, 3\}, \{1, 4\}\}$$

의사결정 집합 : $Model = \{1, 2\}$

이렇게 찾아진 속성 집합과 의사결정 집합을 이용하여 코어를 계산하는 과정은 다음과 같다.

$$1) \Delta_{m(k)}(1) \cap \Delta_{m(k)}(2) = \{1, 2\}$$

$$2) \Delta_{m(k)}(2) \cap \Delta_{m(k)}(3) = \{1\}$$

$$3) \Delta_{m(k)}(1) \cap \Delta_{m(k)}(3) = \{1, 4\}$$

$$\text{코어} : \{\Delta_{m(k)}(1), \Delta_{m(k)}(2)\} \cap \{\Delta_{m(k)}(2), \Delta_{m(k)}(3)\}$$

이러한 방법으로 코어를 이용하여 최종 감축된 규칙은 아래의 표와 같다. 따라서 규칙에 포함되어 있는 전체의 속성을 사용하지 않고도 코어를 이용하여 효율적인 규칙의 생성이 가능하다.

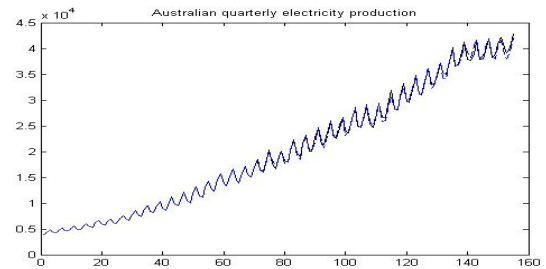
〈표 2〉 확률적 처리와 코어에 의한 규칙 감축

속성 규칙	$\Delta_{m(k)}(1)$	$\Delta_{m(k)}(2)$	$\Delta_{m(k)}(3)$	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)

이렇게 감축된 규칙을 이용하여 입력 데이터의 규칙에 따라 예측 모델이 선택되어 예측을 수행하게 된다.

6. 시뮬레이션 및 결론

시뮬레이션을 위해 분기별 호주 전력생산량 데이터를 이용하였다. 검은색이 원시계열 값이고 파란색이 예측값이다. 비교적 정확한 예측이 수행되었음을 알 수 있으며, 아래의 표 3은 규칙이 감축되는 변화를 비교한 것으로 적은 수의 규칙으로 최적의 모델을 선택할 수 있음을 보인다.



〈그림 3〉 분기별 호주의 전력생산량 예측 결과

〈표 3〉 시뮬레이션 데이터의 규칙 감축 비교

	초기 생성된 규칙	확률적 처리에 의한 규칙감축	코어에 의한 규칙 감축
규칙 수	50개	19개	15개 (속성이 감축된 규칙 생성)

제안된 방법은 비선형 시계열 데이터의 예측에 있어 데이터를 안정화하여 예측성능을 개선할 수 있는 한 방법이 될 것이며, 러프집합을 이용한 모델 선택방법은 제안된 논문의 방식뿐만 아니라 다수의 시스템의 특성을 고려해야 할 경우나, 또는 Hybrid System의 운용에 유용하게 적용될 수 있을 것으로 사료된다.

[참고 문헌]

- [1] 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
- [2] Z. Pawlak, "Rough Sets", Int. J. of Computer and Information Sciences, 11, pp.341-356, 1982.