

병렬구조 퍼지시스템을 이용한 태양흑점 시계열 데이터의 예측

論 文

54D-6-6

Prediction of SunSpot Number Time Series using the Parallel-Structure Fuzzy Systems

金 旻 秀[†] · 鄭 讚 壽^{*}

(Min-Soo Kim · Chan-Soo Chung)

Abstract - Sunspots are dark areas that grow and decay on the lowest level of the sun that is visible from the Earth. Short-term predictions of solar activity are essential to help plan missions and to design satellites that will survive for their useful lifetimes. This paper presents a parallel-structure fuzzy system(PSFS) for prediction of sunspot number time series. The PSFS consists of a multiple number of component fuzzy systems connected in parallel. Each component fuzzy system in the PSFS predicts future data independently based on its past time series data with different embedding dimension and time delay. An embedding dimension determines the number of inputs of each component fuzzy system and a time delay decides the interval of inputs of the time series. According to the embedding dimension and the time delay, the component fuzzy system takes various input-output pairs. The PSFS determines the final predicted value as an average of all the outputs of the component fuzzy systems in order to reduce error accumulation effect.

Key Words : Chaotic Time Series, Prediction, Parallel Structure Fuzzy System, Subtractive Clustering, Sunspot Number

1. 서 론

태양흑점은 태양 표면에 검은 구멍처럼 보이는 것으로 흑점이 나타나면 태양활동이 활발함을 의미한다. 이러한 태양활동은 플레어나 홍염 등의 형태로 표출되어 지구의 자기장을 변동시킴으로써 전력, 통신 시스템의 장애를 유발하게 된다. 따라서 이러한 흑점 데이터를 예측함으로써 사전에 대비할 수 있도록 할 필요가 있다[1][2].

태양흑점에 대한 관측은 1600년대 초기부터 시작되었으며 수집된 태양활동 관련 데이터의 분석을 통해 태양주기(Solar cycles)에 관한 정보를 밝혀냈으며 이에 기초하여 태양흑점 수의 예측에 관련된 많은 연구가 진행되어왔다[1]~[5].

일반적으로 예측은 모델(Model)에 근거한 접근과 비파라미터(Nonparametric) 접근으로 나눌 수 있는데, 비파라미터 접근 방법은 주어진 데이터의 해석을 통해 예측하는 방법으로 시계열 데이터의 예측에 이용하고 있다.

카오스 신호는 자연현상과 비슷한 특성으로 초기조건에 민감하며 장기 예측이 어려운 신호로서 카오스 신호를 정확히 예측한다는 것은 불가능하나 비파라미터 접근 방법에 기초하여 장기 예측 보다는 단기 예측이나 경향 분석에 중점

을 두고 예측이 이루어지고 있다[6]~[9].

퍼지 시스템은 전문가의 경험이나 지식을 IF-THEN 형태의 언어 규칙을 이용하여 적절한 동작을 취하게 하는 시스템으로 퍼지규칙은 전문가의 경험이나 공학적 지식을 토대로 구할 수 있다.

본 논문에서는 병렬구조 퍼지시스템(PSFS; Parallel-Structure Fuzzy System)을 이용하여 카오스 시계열 데이터와 같은 동특성을 보여주는 태양흑점 시계열 데이터를 예측하였다. PSFS의 각 퍼지시스템은 수치정보로 주어진 경우 주로 사용되는 TSK 퍼지모델[11]을 사용하였으며, 퍼지규칙을 생성하기 위한 방법으로 데이터량이 많은 경우 그 특징 추출이 용이한 데이터 클러스터링 방법을 사용하였다. 퍼지시스템의 입력은 지연시간(Time delay) r 에 따른 임베딩차원(Embedding dimension) m 으로 결정되며, m 은 각 r 에 따라 m 을 변경하면서 구한 한단계 이후 예측의 평균제곱오차(MSE; Mean Square Error)와 최대절대오차(MAE; Maximum Absolute Error)가 작은 상수값들로 결정하였다. 이렇게 구성된 퍼지시스템은 PSFS를 구성하게 되며 PSFS의 최종 출력은 각 퍼지시스템의 출력들의 평균값으로 다음 예측을 위한 입력으로 사용된다.

2. 시계열 데이터의 예측

시계열 데이터 예측을 위한 방법 중 비파라미터 접근은 시계열 데이터의 미래 값을 과거 데이터와의 함수관계로 표현할 수 있다는 가정에 기초해 있다. 즉, 비파라미터 접근에 기초한 미래 데이터의 예측은, k 개의 입력 시계열 데이터가 주어진 경우, k 번째 단계에서 r 단계 이후 예측값 $\hat{x}(k+r)$

[†] 교신저자, 正會員 : 世宗-宇宙航空研究所
研究教授 · 工博

E-mail : mskim@sejong.ac.kr

^{*} 正會員 : 崇實大 工大 電氣工學科 正教授 · 工博
接受日字 : 2004年 11月 18日
最終完了 : 2005年 4月 25日

은 식 (1)과 같이 표현된다.

$$\hat{x}(k+\tau) = F[x(k), x(k-1), x(k-2), \dots, x(2), x(1)] \quad (1)$$

단, $F[\cdot]$ 은 비파라미터 접근인 경우의 시계열 데이터 예측을 위한 입출력 관계식을 의미하는 함수이며, τ 는 양의 정수이다. 시계열 데이터의 미래값은 k 개의 입력데이터의 선형 또는 비선형 함수 $F[\cdot]$ 의 출력에 의해 예측되어진다.

본 논문에서는 시계열 데이터의 예측을 위한 여러 방법 중에서 다음 두 가지 예측 방법이 이용되었는데, 첫번째는 한단계 이후 예측(One-step-ahead prediction)이고, 두번째는 단기예측(Short-term prediction)이다. 먼저, 한단계 이후 예측 방법은 식 (2)의 형태로서 k 번째 단계에서 미래 데이터 $\hat{x}(k+\tau)$ 를 예측하기 위한 입력으로 τ 등간격을 갖는 m 개의 데이터가 사용된다.

$$\hat{x}(k+\tau) = F[x(k), x(k-\tau), x(k-2\tau), \dots, x(k-(m-1)\tau)] \quad (2)$$

여기에서 m 과 τ 는 양의 정수이고 임베딩 차원과 시간지연을 각각 의미한다. 시계열 데이터의 미래값은 m 개의 입력데이터의 선형 또는 비선형 함수 $F[\cdot]$ 의 출력에 의해 예측되어진다.

다음으로 본 논문에서 사용된 단기예측 또는 장기예측 방법은 식 (3)으로 표현되며 예측된 결과값이 다음 단계의 예측을 위한 입력으로 다시 사용되는 방법이다.

$$\hat{x}(k+\tau) = F[\hat{x}(k), \hat{x}(k-\tau), \hat{x}(k-2\tau), \dots, \hat{x}(k-(m-1)\tau)] \quad (3)$$

즉, 미래 데이터에 대한 예측값 $\hat{x}(k+\tau)$ 은 과거 예측값들 $\hat{x}(k), \hat{x}(k-\tau), \dots, \hat{x}(k-(m-1)\tau)$ 로 표현되어진다. 그러므로 초기의 작은오차를 갖는 예측결과도 이 값이 재사용됨에 따라 오차가 누적되기 때문에 장기예측은 어려운 문제가 된다.

특정 시간지연 τ 에 대한 적절한 임베딩차원 m 을 결정할 수 있다. 결정 방법은 시간지연 τ 에서 Training 데이터와 Validation 데이터에 오차성능지표(Error performance measure)를 적용해 계산된다. 여기에서 Training 데이터는 예측시스템을 모델링하기 위해 사용되는 데이터를 의미하고, Validation 데이터는 Training 데이터로 사용되지 않은 데이터로서 모델링된 시스템의 최종 성능측정을 위해 사용된다.

오차성능지표는 한단계이후예측 결과와 실제 데이터의 차로 계산된 MSE와 MAE로 정의된다. 따라서 시계열데이터가 주어지면 고정된 τ 에 대해 MSE와 MAE가 모두 작은 적절한 m 을 결정할 수 있다. 이러한 과정은 Training 데이터와 Validation 데이터에 모두 적용되어 모두에서 최소 성능지표를 보여주는 적절한 상수 m 을 결정하게 된다.

3. 병렬구조 퍼지시스템

3.1 병렬구조 퍼지시스템의 구성

병렬구조 퍼지시스템(Parallel-Structure Fuzzy Systems; PSFS)은 여러 개의 부퍼지시스템을 병렬로 연결한 구조로서 각 부퍼지시스템들이 가지는 서로 다른 예측능력을 이용하는 것이 기본 개념이다[12]~[15].

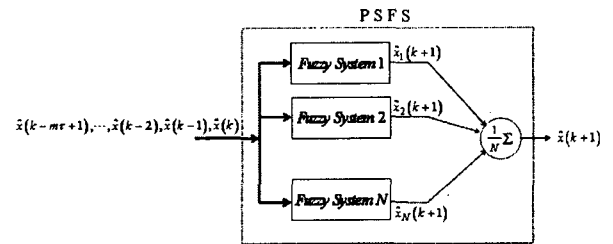


그림 1 병렬구조 퍼지시스템의 구조

Fig. 1 The Structure of the Parallel Structure Fuzzy System

그림 1에는 N개의 부퍼지시스템을 갖는 PSFS의 구조를 나타내고 있다. $(k+1)$ 단계에서 PSFS의 최종 출력 $\hat{x}(k+1)$ 은 식 (4)와 같이 부퍼지시스템의 출력들 $(\hat{x}_1(k+1), \hat{x}_2(k+1), \dots, \hat{x}_N(k+1))$ 의 평균으로 주어진다.

$$\hat{x}(k+1) = \frac{1}{N} \left[\sum_{i=1}^N \hat{x}_i(k+1) \right] \quad (4)$$

PSFS의 입력으로 사용된 데이터는 $\{\hat{x}(k-(m \cdot \tau - 1)), \hat{x}(k-(m \cdot \tau - 2)), \dots, \hat{x}(k-2), \hat{x}(k-1), \hat{x}(k)\}$ 로서 이전단계에서 예측된 $(m \times \tau)$ 개의 값들이 사용되며, 최초 예측값 $\hat{x}(1)$ 은 $(m \times \tau)$ 개의 초기 데이터에 의해 예측된다. PSFS를 이루는 각 부퍼지시스템의 입력은 각 부퍼지시스템이 가지는 (τ, m) 쌍에 의해 결정된다.

만약, PSFS의 부퍼지시스템이 3개로 구성되어 있고 (τ, m) 쌍이 (1,3),(2,4),(3,3)으로 주어지고 $\hat{x}(k+1)$ 을 예측하고자 할 때, 각 부퍼지시스템의 입출력을 살펴보면, 퍼지시스템1은 $\{\hat{x}(k-2), \hat{x}(k-1), \hat{x}(k)\}$ 을, 퍼지시스템2는 $\{\hat{x}(k-7), \hat{x}(k-5), \hat{x}(k-3), \hat{x}(k-1)\}$ 을, 그리고 퍼지시스템3은 $\{\hat{x}(k-8), \hat{x}(k-5), \hat{x}(k-2)\}$ 을 각각 입력으로 사용하게 된다. 각 퍼지시스템의 출력은 각각 $\hat{x}_1(k+1), \hat{x}_2(k+1), \hat{x}_3(k+1)$ 이 되어 최종 예측값 $\hat{x}(k+1)$ 을 결정하게 된다. 이러한 입출력 관계를 그림 2에 나타내었다.

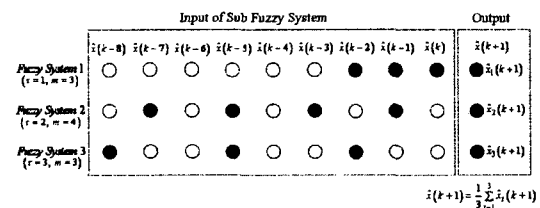


그림 2 각 부퍼지시스템의 입출력

Fig. 2 I/O of Each Sub Fuzzy System

단기예측이나 장기예측의 경우에는 예측된 값은 다음 단계에서 입력으로 사용되기 때문에 작은 오차도 누적되어 몇 단계를 지나면 예측이 어려워진다. 이러한 문제로 인해 여러 시스템 출력들의 평균값을 최종 출력으로 사용하였다.

3.2 각 퍼지시스템 모델링

Takagi와 Sugeno가 제안한 혼합 퍼지추론 방법에서 i번째 퍼지규칙은 식 (5)과 같이 표현된다.

$$R_i : \text{IF } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_n \text{ is } A_{in} \\ \text{THEN } y_i = f_i(x_1, \dots, x_n) \quad (5)$$

여기에서 x_1, x_2, \dots, x_n 은 n개의 입력을, A_{ij} 는 퍼지 입력 변수를, y_i 는 i번째 규칙의 출력을 나타내며 $f_i(\cdot)$ 는 입력 변수들과 상수들의 선형 함수로서 식 (6)와 (7)처럼 표현된다.

$$A_{ij}(x_j) = \exp\left(-\frac{1}{2} \cdot \left(\frac{x_j - c_{ij}}{w_{ij}}\right)^2\right) \quad (6)$$

$$f_i(x_1, \dots, x_n) = a_{0i} + a_{1i}x_1 + \dots + a_{ni}x_n \quad (7)$$

이때 c_{ij} 는 멤버십 함수의 중심을, w_{ij} 는 멤버십 함수의 폭을 나타내는 변수이며 $a_{0i}, a_{1i}, \dots, a_{ni}$ 는 데이터에 의해 결정되어야 하는 상수이다.

n개의 입력과 M개의 규칙, 그리고 max-product 추론 방법을 사용하였을 때, i번째 전건부 규칙의 정규화한 적합도 μ_i 는 식(8)과 같이 표현되며 퍼지시스템의 최종 추론값 y^* 는 식 (9)와 같이 표현된다.

$$\mu_i = \frac{\prod_{j=1}^n A_{ij}(x_j)}{\sum_{k=1}^M \left(\prod_{j=1}^n A_{jk}(x_j)\right)} \quad (8)$$

$$y^* = \sum_{i=1}^M \mu_i \cdot f_i(x_1, \dots, x_n) \quad (9)$$

퍼지규칙 생성 방법인 데이터 클러스터링 알고리즘은 Mountain 클러스터링 방법의 변형된 형태인 Subtractive 클러스터링 알고리즘을 사용하였다[16]~[18]. Subtractive 클러스터링 알고리즘은 데이터간의 거리의 함수로 주어지는 포텐셜 값이 최대인 점을 첫번째 클러스터 중심으로 선정하고, 첫번째 클러스터 중심의 영향을 제거한 상태에서 최대 포텐셜 값을 갖는 데이터가 다음 클러스터 중심이 된다. 입력이 n차원이고 출력이 1차원 공간을 이루고 있는 (n+1) 차원 입출력 공간상에서 N개의 데이터 $\{X_1, X_2, X_3, \dots, X_N\}$ 가 주어졌을 때 데이터 클러스터링은 다음과 같은 순서로 이루어진다.

첫째, 주어진 데이터를 [0, 1]로 정규화 한다.

둘째, 데이터 간의 거리를 구한다. i번째 데이터에서 각 데이터 간의 거리는 식 (10)과 같이 함수 P로 표현되며 P_i 를 포텐셜 값이라 한다.

$$P_i = \sum_{j=1}^N \exp(-a \cdot \|X_i - X_j\|^2) \quad (10)$$

여기에서 a 는 $4/r_a^2$ 로 주어지며, r_a 는 양의 상수로서 r_a 밖의 데이터는 포텐셜값에 영향을 거의 주지 못한다.

셋째, 첫번째 클러스터 중심을 구한다. N개의 포텐셜 값 중 가장 높은 값을 P_1^* 라 놓고 이 때의 데이터가 첫 번째 클러스터 중심 X_1^* 이 된다.

넷째, 첫번째 클러스터 중심의 영향을 제거한다. 첫번째 클러스터 중심 근처에는 많은 데이터가 존재하기 때문에 그 영향을 제거하지 않으면, 두 번째 클러스터 중심 또한 첫번째 클러스터 중심 근처에 존재할 가능성이 높기 때문에 식 (11)과 같이 첫번째 클러스터 중심의 영향을 제거한 포텐셜 값을 구한다.

$$P_i' = P_i - P_1^* \cdot \exp(-\beta \cdot \|X_i - X_1^*\|^2) \quad (11)$$

여기에서 β 는 $4/r_b^2$ 로 주어지며, r_b 는 양의 상수로서 r_a 보다 큰 값을 취해 클러스터 중심 근처에 다음 클러스터 중심이 나타나지 않도록 한다.

다섯째, 두 번째 클러스터 중심을 구한다. 첫번째 클러스터 중심의 영향을 제거한 N개의 포텐셜 값 P_i' 중 가장 높은 값을 P_2^* 라 놓고 이 때의 데이터가 두 번째 클러스터 중심 X_2^* 가 된다. k번째 클러스터 중심 X_k^* 가 구해졌을 때, k번째 클러스터 중심의 영향을 제거한 포텐셜 값은 식 (12)로 표현할 수 있으며 (k+1)번째 클러스터 중심을 구하게 된다.

$$P_i'' = P_i - P_k^* \cdot \exp(-\beta \cdot \|X_i - X_k^*\|^2) \quad (12)$$

여섯째, d_{\min} 을 $(X_1^*, X_2^*, \dots, X_{k-1}^*)$ 과 X_k^* 의 거리 중에서 가장 짧은 거리로 정의하였을 때, $P_k^* \geq \bar{\epsilon} \cdot P_1^*$ 이면 클러스터 중심으로 채택하고 위의 과정을 반복한다.

$P_k^* \geq \epsilon \cdot P_1^*$ 이고 $\frac{d_{\min}}{r_a} + \frac{P_k^*}{P_1^*} \geq 1$ 이면 클러스터 중심으로

채택하고, $P_k^* > \epsilon \cdot P_1^*$ 이고 $\frac{d_{\min}}{r_a} + \frac{P_k^*}{P_1^*} < 1$ 이면

그때의 X_k^* 를 0으로 하고 다음으로 높은 포텐셜 값을 선정

한 후, d_{\min} 을 구하여 $\frac{d_{\min}}{r_a} + \frac{P_k^*}{P_1^*} \geq 1$ 이면 새로운 클러스터

중심으로 채택한다. $P_k^* < \epsilon \cdot P_1^*$ 이면 위의 반복을 중단한다. 이때 $\bar{\epsilon}$ 와 ϵ 은 클러스터 중심 선정의 상하 기준이 되는 상수이다.

$$d_{\min}(k) = \min_i \{ \sqrt{\|X_i^* - X_k^*\|^2} \} \quad i=1, 2, \dots, k-1 \quad (13)$$

위와 같은 방법으로 구한 (n+1) 차원인 M개의 클러스터 중심 $\{X_1^*, X_2^*, \dots, X_M^*\}$ 을 입출력 공간으로 나누었을 때 X_i^* 의 입력공간은 n차원의 x_i^* 로, 출력공간은 1차원의 z_i^* 이 된다.

여기에서 x_i^* 는 퍼지규칙이 되며 퍼지 규칙에 의한 적합도 μ_i 및 최종 출력값 y^* 를 구하기 위해서 z_i^* 를 $G_i \cdot x + h_i$ 로 나누어 식 (7) 형태를 만든 후 선형최소자승법 (linear least-squares estimation) 알고리즘을 이용하여 변수 G_i, h_i 값을 구한다. 적합도 μ_i 는 식 (14)를 통해서 그리고 최종 출력값 y^* 은 식 (15)를 통해 구해지며 이는 식 (8), (9)와 동일한 결과를 얻게 된다.

$$\mu_i = \exp(-\alpha \cdot \|x - x_i^*\|^2) \quad (14)$$

$$y^* = \frac{\sum_{i=1}^M \mu_i \cdot (G_i \cdot x + h_i)}{\sum_{i=1}^M \mu_i} \quad (15)$$

여기에서 x 는 입력 벡터를 나타내며 α 는 식 (10)에서와 동일한 값이다.

4. 시뮬레이션

4.1 태양 흑점 시계열 데이터

본 논문에서 예측에 사용된 태양흑점 시계열 데이터는 1749년 1월부터 2004년 10월까지 수집된 월 평균 데이터로서 여러 지역에서 관측을 통해 식 (16)의 흑점상대수 R 로 표현된다.

$$R = k(10g + s) \quad (16)$$

여기에서 g 는 태양지표에 보이는 흑점 그룹의 수를, s 는 각 지역에서 관측된 흑점의 전체 수를 각각 나타내며, k 는 관측자와 관측소 사이의 차이를 고려한 매개 인자로서 1보다 작은 상수이다.

표 1. 태양 흑점 시계열 데이터

Table 1. Sunspot Time Series

Index	Year/Month	Sunspot Number (R)
1	1749 / 1	58.0
2	1749 / 2	62.6
3	1749 / 3	70.0
4	1749 / 4	55.7
5	1749 / 5	85.0
...
3,065	2004 / 6	43.2
3,067	2004 / 7	51.0
3,068	2004 / 8	40.9
3,069	2004 / 9	27.7
3,070	2004 / 10	48.4

수집된 데이터를 표 1에 나타내었으며, 그림 3에는 태양 흑점 시계열 데이터를 나타내었다. 최대값은 1957년 10월 (2,506번째 시계열 데이터)의 253.8이다.

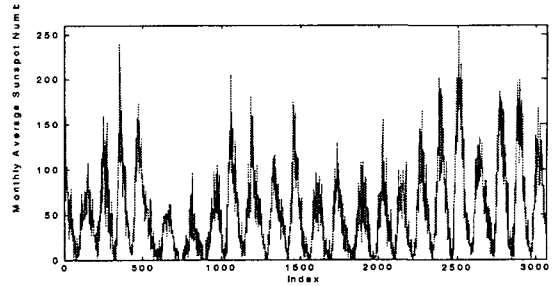


그림 3 태양 흑점 시계열 데이터
Fig. 3 The Sunspot Number Time Series

태양 흑점 시계열 데이터는 Training 데이터, Validation 데이터, Test 데이터로 세분된다. 이 중에서 Training 데이터와 Validation 데이터는 PSFS를 모델링하는데 사용되고, Test 데이터는 모델링된 PSFS의 성능을 검증하는데 사용된다. 즉, 전체 3,070개의 데이터 중에서 PSFS의 성능 검증을 위한 최종 160개의 Test 데이터를 제외한 나머지 데이터 중에서 앞부분 80%는 Training 데이터로 사용되어 PSFS를 모델링 및 최적 m 을 결정하는데 사용하였으며, 나머지 20%의 데이터는 Validation 데이터로서 최적 m 을 결정할 때 사용하였다. 그리고 예측시스템을 모델링하기 위한 Subtractive 클러스터링 알고리즘에 사용된 파라미터는 $r_a = 0.3, r_b = 0.75, \bar{\epsilon} = 0.3, \underline{\epsilon} = 0.1$ 이다.

그림 4는 PSFS 시스템의 m 값들을 결정하기 위한 과정으로서 τ 를 12로 고정하고, m 을 2에서 8까지 변화시켰을 때의 클러스터 중심의 수와 Training 및 Validation 데이터에서 각각 구한 MAE 그리고 MSE를 보여주고 있다. 주어진 시계열 데이터가 월별 평균 데이터이기 때문에 τ 를 가변시키지 않고 12로 고정하였다. 예측해본 결과 τ 를 가변시키면서 구한 m 에 기초한 PSFS보다 τ 를 12로 고정시킨 후 다양한 m 값으로 구성된 PSFS의 예측성능이 보다 우수하였다.

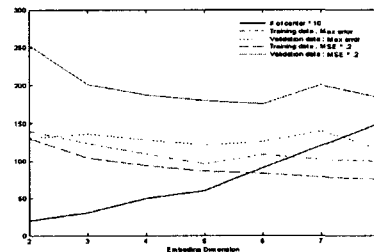


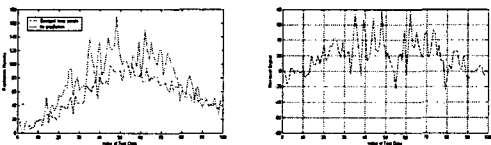
그림 4 m 값의 결정
Fig. 4 Determination of m values

그림 4에서 클러스터 중심의 수는 m 값이 증가함에 따라 비례하여 증가함을 알 수 있다. 이 클러스터 중심의 수는 퍼지시스템의 규칙의 수와 같으며 규칙이 증가하게 되면 시스템의 특성을 그대로 모방하려하기 때문에 카오스특징을 나타내는 시계열 데이터의 예측성능은 떨어지게 된다 그리고 Training 및 Validation 데이터에서 구한 MAE와 MSE는 감소하다 다시 증가하는 모양을 보여주고 있다. 따라서 τ 가 12일 때, $N=3$ 인 PSFS의 m 값을 4,5,6으로 결정하였다.

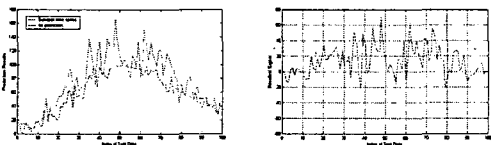
4.2 시뮬레이션 결과

본 논문에서 제안한 PSFS의 성능을 검증하기 위해 각 부퍼지시스템을 단독으로 사용하였을 때와 병렬로 구성하였을 때의 예측성능을 비교하여 보았다. 즉, τ 가 12이고 m 이 4,5, 또는 6일 때, 각 부퍼지시스템이 가지는 예측성능을 그림5와 표2에 각각 나타내었다. 단, 예측에 이용된 데이터는 초기데이터를 포함한 160개의 Test 데이터이며, 표 2에서 예측 성능지표는 식 (17)에서와 같이 잔차신호(Residual signals)의 K 샘플까지의 절대값 평균을 사용하였다.

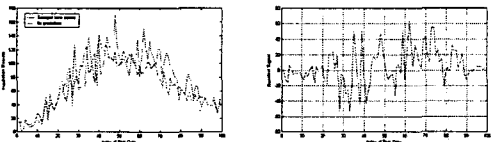
$$\frac{1}{K} \sum_{i=0}^K |x(i) - \hat{x}(i)| \quad (17)$$



(a) $m=4$ 인 부퍼지시스템의 예측결과 및 잔차신호



(b) $m=5$ 인 부퍼지시스템의 예측결과 및 잔차신호



(c) $m=6$ 인 부퍼지시스템의 예측결과 및 잔차신호

그림 5 부퍼지시스템의 예측결과 및 잔차신호
Fig. 5 Prediction of Component Fuzzy Systems

표 2. 부퍼지시스템과 PSFS와의 성능비교

Table 2. Performance Comparison of Component Fuzzy Systems and PSFS

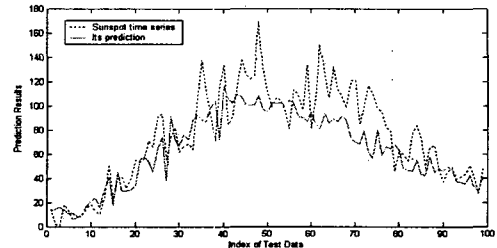
	K=10	K=20	K=30	K=40	K=50
$m=4$	4.74	9.24	15.62	20.78	29.93
$m=5$	5.14	9.31	12.58	16.17	18.78
$m=6$	6.35	6.15	10.34	15.74	17.48
$m=4,5,6$ (PSFS)	4.85	6.89	9.19	12.54	15.01

표 2의 결과에서 알 수 있듯이 PSFS의 성능이 각 부퍼지시스템보다 우수한 예측 결과를 보여주었다.

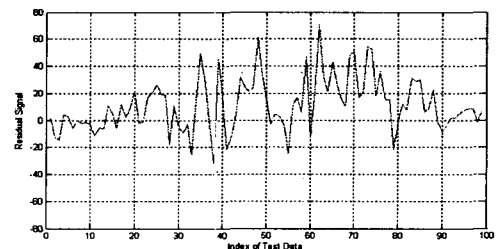
흑점시계열 데이터의 예측에 사용된 PSFS는 τ 가 12로 고정된 값에서 m 이 4,5,6을 갖는 $N=3$ 인 PSFS이다. 각 부퍼지시스템의 (τ, m) 쌍은 (12,4),(12,5),(12,6)을 가지며, 각 부퍼지시스템의 출력들의 평균을 최종 예측값으로 사용하였다.

제안한 PSFS를 이용하여 Test 데이터를 예측하여 예측

결과와 잔차신호를 그림 6에 각각 나타내었다. Test 데이터의 수는 PSFS의 초기데이터를 포함한 160개이다. 예측결과는 그림 6-(b)의 잔차신호를 토대로 알 수 있는데, 30샘플 이상부터는 절대오차가 30이상으로 급격히 증가하는 것을 알 수 있다.



(a) 예측결과



(b) 잔차신호

그림 6. Test 데이터의 예측

Fig. 6 Prediction of the Test Data

다음으로, 제안한 예측시스템에 2004년 11월부터 2013년 4월까지의 미래 데이터를 예측하였고 그 결과를 그림 7에 나타내었다.

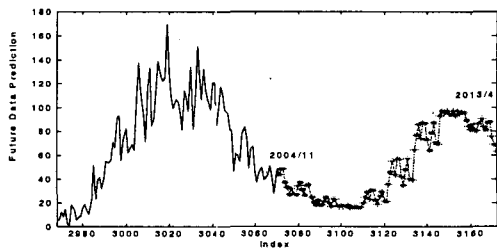


그림 7 미래데이터의 예측

Fig. 7 Prediction of the Future Data

5. 결 론

본 논문에서는 $N=3$ 인 병렬구조 퍼지시스템(PSFS)을 이용하여 태양흑점 시계열 데이터를 예측하였다. 예측에 이용된 PSFS는 여러 개의 퍼지시스템들을 병렬 연결형태로 구성되어 있으며, 각 부퍼지시스템은 입력력 데이터 형태로 재구성된 시계열 데이터를 클러스터링함으로써 모델링되었다. 최종 예측값은 각 퍼지시스템의 출력값들의 평균을 사용하였으며 이 값을 다음 예측을 위해 입력으로 사용되었다.

제안한 PSFS는 시간지연 τ 에 따른 최적 임베딩차원 m 을 입력으로 사용하거나 고정된 τ 에서 적절한 m 을 결정하여 사용할 수 있도록 하여 보다 유연성을 갖도록 하였다.

시뮬레이션에서는 PSFS에 태양흑점 시계열데이터를 적용하여 예측성능을 확인하였으며, 그 결과 비교적 정확한 예측성능을 보여줌을 알 수 있었다.

본 논문의 예측 결과를 토대로 태양 흑점변화를 예견함으로써 전력, 통신 시스템 등의 장애를 사전에 방지하는데 사용할 수 있을 것으로 보여진다. 추후 연구과제는 태양흑점 시계열데이터가 가지는 내부적인 측정오차를 고려하여 이동 평균 방법이나 칼만필터 등을 이용하여 예측 성능을 높이는 연구가 필요할 것으로 보인다.

참 고 문 헌

[1] Thompson, R. J., "A Technique for Predicting the Amplitude of the Solar Cycle," Solar Physics 148, 1993.

[2] Hathaway, D. H., Wilson, R. M., and Reichmann, E. J., "A Synthesis of Solar Cycle Prediction Techniques," Journal Geophys. Res., 104, No. A10, 1999.

[3] K. J. Li, H.S. Yun, H. F. Liang, and X. M. Gu, "Solar activity in Extended Cycles," Journal Geophys. Res., 107, No. A7, 2002.

[4] M. S. Kim, H. S. Lee, C. H. You, and C. S. Chung, "Chaotic Time Series Prediction using PSFS2," 41st Annual Conference on SICE, August 2002.

[5] 김민수, 정찬수, "개선된 PSFS를 이용한 태양흑점 시계열 데이터의 예측," 대한 전기학회 하계학술대회, 2003. 7.

[6] A. S. Weigend and N. A. Gershenfeld, eds. Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley Pub., pp.175-193, 1994.

[7] M. Casdagal, "Nonlinear Prediction of Chaotic Time Series," Physica D, pp.335-356, 1989.

[8] D. Lowe and A. R. Webb, "Time Series Prediction by Adaptive Networks: A Dynamical Systems Perspective," Artificial Neural Networks, Forecasting Time Series, V. R. Vemuri, and R. D. Rogers (ed.), IEEE Computer Society Press, pp. 12-19, 1994.

[9] D. S. Broomhead and D. Lowe, "Multi-variable Functional Interpolation and Adaptive Networks," Complex Systems, pp.262-303, 1988.

[10] E. H. Mamdani, and S. Assilian, "An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller," Int. J. of Man Machine Studies, Vol. 7, No. 1, pp.1-13, 1975.

[11] M. Sugeno, Industrial Applications of Fuzzy Control, Elsevier Science Pub., 1985.

[12] 김민수, 공성곤, "화재사고 경향분석을 위한 비선형 시계열 예측기법," 전기학회 논문지, 47권, 10호, 1774-1779, 1998. 10.

[13] 김민수, 공성곤, "병렬구조 퍼지시스템을 이용한 카오

스 시계열 데이터 예측," 퍼지시스템학회 논문지, 10권, 2호, 3-10, 2000. 4.

[14] M.-S. Kim and S.-G. Kong, "Time Series Prediction using the Parallel-Structure Fuzzy System," 1999 IEEE Int. Fuzzy Systems Conference Proceedings, Vol. 2, 934-938, August 1999.

[15] M. S. Kim and S. G. Kong, "Parallel Structure Fuzzy Systems for Time Series Prediction," Int. Journal of Fuzzy Systems, Vol. 3, No.1, March 2001.

[16] J.-S. R. Jang, and C.-T. Sun, "Neuro-Fuzzy Modeling and Control," Proceedings of the IEEE, March 1995.

[17] S. Chiu, "Fuzzy Model Identification Based on Cluster Estimation," Journal of Intelligent & Fuzzy Systems, Vol. 2, No. 3, Sept. 1994.

[18] R. R. Yager and D. P. Filev, Essentials of Fuzzy Modeling and Control, John Wiley & Sons, pp.246-264, 1994.

저 자 소 개



김민수 (金 旼 秀)

1971년 2월 25일생. 1995년 숭실대 전기공학과 졸업. 2003년 동대학원 졸업(공학박). 2004년~현재 세종-록히드마틴우주항공연구소 연구교수
 Tel : 02-3408-3897
 Fax : 02-3408-3895
 E-mail : mskim@sejong.ac.kr



정찬수 (鄭 讚 壽)

1949년 8월 10일생. 1972년 서울대 전기공학과 졸업. 1987년 동대학원 졸업(공학박). 1981년~현재 숭실대 전기공학과 정교수
 Tel : 02-820-0645
 Fax : 02-817-7961
 E-mail : chung@ssu.ac.kr