변형된 입력을 이용한 퍼지 시계열 예측 방법

A Fuzzy Time Series Prediction method using modified inputs

이성록 . 김인택
Sung-Rock Lee and Intaek Kim
명지대학교 정보제어공학과

요 약

본 논문은 효과적인 시계열 예측을 위한 새로운 퍼지 학습방법을 제안한다. 기존의 학습방법에서는 입력 데이터를 F(y(t),y(t-1),y(t-2)..)의 형태로 주어 예측을 수행했으나 본 논문에서 제안한 방법에서는 입력 데이터를 F(y(t)-y(t-1),y(t-1)-y(t-2)...)로 설정한다. 이것은 각 입력값의 차이를 새로운 입력으로 사용함으로써 유사한 시계열 분포를 좀더 능동적인 퍼지 규칙으로 만들기 때문에 Non-Stationary한 데이터뿐만 아니라 기존의 시계열 데이터 예측방법 보다 나은 결과를 나타낸다. 알고리즘의 수행능력을 살펴보기 위해 Mackey-Glass time series와 Lorenz data를 사용하였다.

1. 서론

시계열(Time series)은 주기적으로 측정된 값들의 나열을 의미한다. 값들의 순서(Order)는 중요한 의미를 가지는데 이는 시계열의 발생원에 대한 정보를 포함하기 때문이다. 시계열중 특정주기를 가지는 값 사이에는 얼마간의 종속성이 발견되고, 이런 성격에 근거하여 시계열의 발생원을 분석할 수 있다. 즉 시계열의 발생원은 어느 정도의 결정론적인 동력학 시스템(Deterministic Dynamic System)에 의해 지배를 받는다고 볼 수 있기 때문에 우리가 얻는 시계열은 예측이나 모델링을 통해 분석이 가능하다[1].

퍼지 논리를 적용한 모델링과 예측은 과거부터 많은 연구가 수행되었던 문제이다. 1980년에 Tong은 실험 데이터로부터 퍼지 모델을 얻고, 이를 위한 평가방법을 제안하였다[2]. 이들은 모델을 복잡성, 정확성, 그리고 불확실성의 측면에서 모델들의 성능을 평가하였다. Pedrycz는 퍼지이론을 사용하여 비결정론적인 시스템의 인식(Identification)에 적용하였다[3]. Jang과 Sun은 그들이 제안했던 ANFIS(Adaptive Network-based Fuzzy Inference System)을 이용하여 시계열의 예측에 적용 하였다[4]. 예로 Mackey-Glass time series를 16개의 퍼지 규칙으로 구현하였으나 실제의 구현은 일반적인 퍼지추론의 그것보다 훨씬 더 복잡한 문제점을 가지고 있다. 본 논문은 이렇게 연구되어온 퍼지 시계열 예측방법에 변형된 입력 방법을 도입함으로써 더 나은 시계열 예측 결과를 보여주고자 한다. 성능을 비교하는 척도로는 RMSE(Root Mean Square Error)를 사용하며 사용된 표준 데이터는 Mackey-Glass time series 와 Lorenz 데이터를 사용하였다.

2. 퍼지 추론 방법

2.1 기존의 퍼지 시계열 예측 시스템

퍼지 추론 시스템은 추론엔진(Inference Engine)으로 아래와 같은 형태의 퍼지규칙을 가진다.

$$R\text{: if } x_1 \text{ is } A \text{ and } x_2 \text{ is } B, \ldots x_n \text{ is } C, \text{ then } x_{n+1} \text{ is } D.$$

여기서 if 이하의 전건부(premise)는 규칙의 조건부에 해당하며, then이하는 후건부(consequence)로 규칙의 작용부를 의미한다. 시계열 $x_1, x_2, \ldots, x_n, x_{n+1}$은 퍼지 언어변수(Fuzzy Linguistic Variables)이며 A, B, C, ... D등은 퍼지 언어값(Fuzzy Linguistic Values)으로 표현한다. 따라서 시계열의 예측은 아래의 식(1)을 만족하는 퍼지 규칙 R을 찾아내는 작업이다.

$$\hat{x}_{n+k} = R(x_1, x_2, \ldots, x_n) \qquad (1)$$

위의 식에서 우리가 가지고 있는 데이터는 $x_1, x_2, \ldots x_n$이며 k시간만큼 후의 값인 x_{n+k}를 예측하고자 한다. 따라서 이는 open loop 예측으로 예측된 결과가 다시 추론 시스템의 입력으로 들어가는 close loop예측과 달리 에러가 적지만 매우 단시간의 예측만 가능하다. 퍼지 규칙을 만드는 작업은 아래와 같다[5][6]:

(1) 시계열 값 x 가 가질 수 있는 범위를 퍼지 언어값에 의해 p개로 구분한다.
(2) 각 시계열 값에 해당하는 퍼지 언어값을 부여하여 식(1)의 퍼지규칙 R을 생성시킨다.

위의 퍼지 추론 시스템(Fuzzy Inference System)은 Stationary한 비선형 시계열에서 잘 동작한다. 하지만 Non-Stationary의 성격이 강한 데이터에서는 좋지 못한 결과를 얻을 수 있음을 알 수 있다. 그러나 이를 보완하기 위해서 본 논문은 다음과 같이 변형된 입력을 통한 퍼지 시계열 예측 시스템을 제안한다. 본 논문에서 제안된 방법은 Non-Stationary한 성격이 강한 데이터 뿐만 아니라 기존의 데이터에서도 나은 결과를 보여 주고 있다.

2.2 변형된 입력을 통한 퍼지 시계열 예측 시스템

변형된 입력을 이용한 퍼지 시계열 예측 방법에서는 다음과 같은 변형된 입력을 전건부로 하는 퍼지규칙을 생성한다.

$$R\text{: if } (x_1 - x_2) \text{ is } A \text{ and } (x_2 - x_3) \text{ is } B, \ldots (x_{n-1} - x_n) \text{ is } C, \text{ then } (x_n - x_{n+1}) \text{ is } D.$$

여기서 변형된 시계열입력 $x_1-x_2, x_2-x_3, \ldots, x_{n-1}-x_n, x_n-x_{n+1}$은 퍼지 언어변수(Fuzzy Linguistic Variables)로 A, B, C, ... D등은 퍼지 언어값(Fuzzy Linguistic Values)으로 표현한다. 따라서 시계열의 예측은 아래의 식(2)을 만족하는 퍼지 규칙 R을 찾아내는 작업이다.

$$\hat{x}_n - \hat{x}_{n+1} = R((x_1 - x_2), (x_2 - x_3), \ldots, (x_{n-1} - x_n)) \qquad (2)$$

여기서 우리가 가지고 있는 데이터는 $x_1, x_2, \ldots x_n$이며 k시간만큼 후의 값인 x_{n+k}를 예측하고자 한다. 이렇게 변형된 입력을 주었을 경우 Non-Stationary 성격이 강한 시계열을 예측하는데 기존 방법보다 좋은 결과를 얻을 수 있다. 다음 그림2와 그림3는 Non-Stationary한 데이터를 기존의 방법과 비교하여 본 것이다. 사용된 데이터는 1000개의 Non-Stationary한 sin함수값으로 700개를

학습에 이용하고 291개의 데이터를 테스트하였다.

위의 결과를 보면 Non-Stationary한 데이터에서 기존의 퍼지 시계열 예측방법이 전혀 예측하지 못하는 반면에 제안된 방법은 거의 정확한 예측을 하는 것을 볼 수 있다. 이것은 직관적으로 유사한 경향을 보이는 구간을 퍼지 규칙으로 삼기 때문에 결과적으로 과거 입력값의 경향에 의해 반응을 보이는 것이 아니라 입력간의 차에 의해 반응을 보이므로 보다 효과적인 결과를 나타내는 것이다.

3. 시계열 예측에의 응용

위의 장에서 설명 한 것과 같이 각 입력간의 차를 퍼지 입력값으로 사용할 경우 기존의 방법보다 개선됨을 알 수 있다. 이 과정에서 각 입력값의 차이를 이용하여 퍼지 소속함수(Fuzzy membership function)을 만들 때 그 범위와 개수가 결과에 영향을 미치게 된다. 소속함수의 개수가 많을 경우 퍼지 규칙의 개수는 많아지나 그 만큼 적은 오차를 가지는 값을 예측 할 수 있다. 그러나 소속함수의 개수가 너무 많아질 경우 퍼지 규칙에서 벗어나는 입력값들이 발생하여 예측을 할 수 없는 경우가 발생한다. 또한 각 입력값의 크기를 가지고 임의로 정한 소속함수의 범위는 소속함수의 개수가 같더라도 더 작은 범위를 가지고 세밀하게 나누었을 경우 더 좋은 결과를 나타내는 것을 볼 수 있다. 본 장에서는 기존의 퍼지 시계열 예측 방법과 비교하여 본 논문에서 제안된 방법의 효율성을 살펴본다. 각 시뮬레이션의 소속함수의 개수는 동일하고 소속함수의 범위는 각 입력값에 따라 달리 부여하였다. 시뮬레이션은 기존의 퍼지 시계열 예측 방법에서는 과거의 9개의 데이터로 한 스텝 앞의 데이터를 예측하며 제안된 방법에서는 똑같이 과거의 9개의 데이터를 사용하여 한 스텝 앞의 데이터를 예측하나 각 데이터들의 차이를 이용하므로 입력값은 8개를 사용한다. 시뮬레이션에 사용되는 데이터로는 Mackey-Glass time series와 Lorenz 데이터를 각각 1000개를 사용하며 700개를 학습에 이용하고 300개를 테스트에 사용한다.

3.1 Mackey-Glass time series

Mackey-Glass series는 다음과 같다.

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \qquad (2)$$

표 1은 각 소속함수의 개수에 따른 시뮬레이션 결과를 보여준다. 그림4, 그림5는 소속함수의 개수가 14개일 때 결과를 비교한 것이다.

3.2 Lorenz 데이터

Lorenz equation 은 다음과 같다.

$$\frac{dx}{dt} = \sigma(y-x) \qquad \frac{dy}{dt} = rx - v - xz \qquad \frac{dz}{dt} = sy - bz \qquad (3)$$

표 2은 각 소속함수의 개수에 따른 시뮬레이션 결과를 보여준다. 그림 6, 그림 7은 소속함수의 개수가 각각 9개일 때 결과를 비교한 것이다.

4. 결론

본 논문에서는 예측을 위한 퍼지 추론 시스템에서 변형된 입력방법을 도입함으로써 얻어지는 개선된 실험결과를 기존의 방법과 비교하여보았다.

시뮬레이션 결과는 본 논문에서 제안한 변형된 입력방법을 도입하여 시계열 예측방법이 기존의 방법보다 나은 결과를 보여주었다. 이것은 각 입력의 차이를 새로운 입력으로 사용함으로써 유사한 시계열 분포를 좀더 능동적인 퍼지 규칙으로 만들기 때문이다.

향후 과제로서는 변형된 입력을 주었을 때 소속함수의 개수와 소속함수의 범위를 결정하는 방법을 들 수 있다. 위 실험결과에서도 알 수 있듯이 소속함수의 개수와 소속함수의 범위는 결과값의 오차와 밀접한 관계가 있기 때문이다. 이러한 문제는 시계열 분석에 국한된 것이 아니라 퍼지 논리가 적용 가능한 일반적인 응용문제에서 퍼지 규칙을 형성하는데 중요한 척도가 되리라 판단된다.

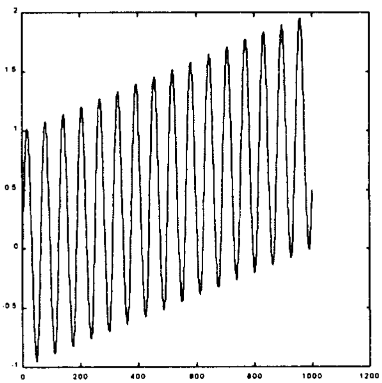

그림 1 실험에 사용된 데이터

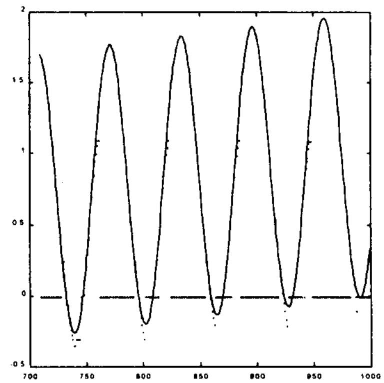

그림 2 기존의 퍼지 시계열 예측 방법

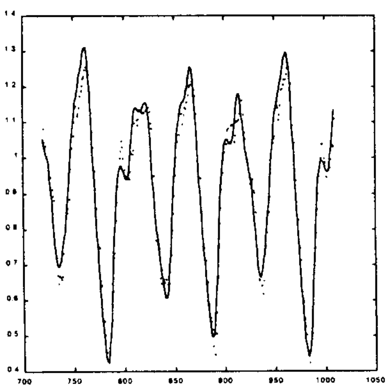

그림 3 소속함수가 14개인 기존 방법에 의한 시계열예측

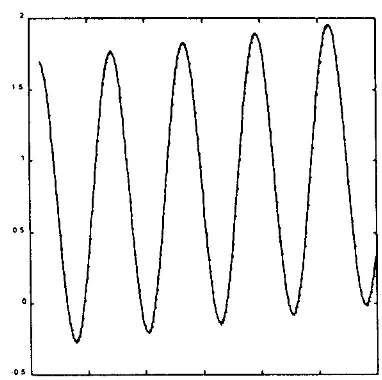

그림 4 제안된 퍼지 시계열 예측 방법

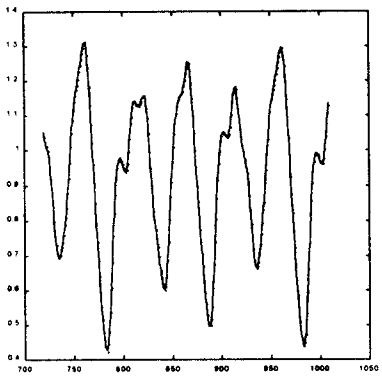

그림 6 소속함수가 14개인 제안된 방법에 의한 시계열예측

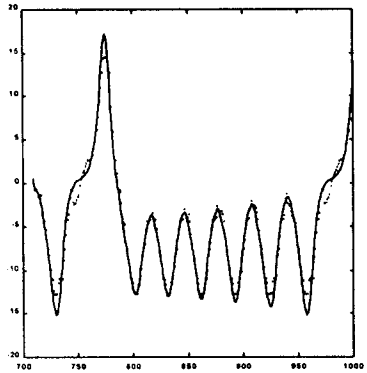

그림 5 소속함수가 9개인 기존 방법에 의한 시계열예측

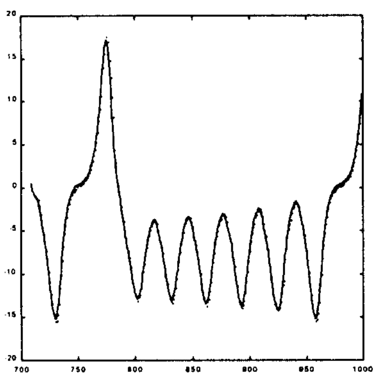

그림 7 소속함수가 9개인 제안된 방법에 의한 시계열예측

표 1 예측오차(RMSE)

입력소속 함수의개수	소속함수의 범위		
	기존 시계열 예측 결과	제안된 시계열 예측 결과	
	[0:1.5]	[-1.05:1.05]	[-0.08:0.06]
6	0.0507	0.0142	0.0080
14	0.0275	0.0103	0.0041
29	0.0155	0.0079	0.0156

표 2 예측오차(RMSE)

입력소속 함수의개수	소속함수의 범위		
	기존 시계열 예측 결과	제안된 시계열 예측 결과	
	[-16:18]	[-34:34]	[-1.8:2.4]
9	0.8694	0.8596	0.2837
16	0.4845	0.3028	0.3663
33	1.1927	0.4960	0.5151

참고 문헌

[1] G. Janacek and L. Swift, TIME SERIES Forecasting, Simulation, Applications, Ellis Horwoed, 1993.

[2] R.M. Tong, "The Evaluation of Fuzzy Models derived from Experimental Data." Fuzzy Sets and Systems, vol.4, 1-12,1980

[3] W. Pedrycz, Fuzzy Control and Fuzzy Systems, John Wiley&Sons Inc.,1989.

[4] J.R. Jang and C.Sun, "Prediction Chaotic Time Series with Fuzzy If-Then Rules." 2nd IEEE Inter.Conf. Fuzzy Systems, San Francisco, pp. 1079-1084, 1993.

[5] 김인택, 공창욱, "시계열 예측을 위한 퍼지 학습알고리즘," 한국퍼지 및 지능시스템학회 논문지 1997 vol. 7, no3.

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