

# 기울기벡터를 이용한 카오스 시계열에 대한 예측

원 석 준<sup>†</sup>

요 약

지금까지 삽입(Embedding)벡터를 이용한 국소적예측방법은 고차미분방정식으로 부터 생성된 카오스 시계열을 예측할 때, 파라메타  $\tau$ 의 추정이 정확하지 않으면 예측성능은 떨어졌다. 지금까지 지연시간( $\tau$ )의 값을 추정하는 방법은 많이 제안되어있지만 실제로 고차원미분방정식부터 생성되어진 수많은 시계열에 모두 적용 가능한 방법은 아직 없다. 이것을 기울기 벡터를 이용한 기울기 선형모델을 도입하는 것에 의해 정확한 지연시간( $\tau$ )의 값을 추정하지 않아도 예측성능에 만족할 수 있는 결과를 표시했다. 이것을 이론뿐만이 아니고 경제시계열에도 적용해서 종래의 예측방법과 비교해서 그 유효성을 표시했다.

## The Prediction of Chaos Time Series Utilizing Inclined Vector

Sek Jun Weon<sup>†</sup>

ABSTRACT

The local prediction method utilizing embedding vector loses the prediction power when the parameter  $\tau$  estimation is not exact for predicting the chaos time series induced from the high order differential equation. In spite of the fact that there have been a lot of suggestions regarding how to estimate the delay time ( $\tau$ ), no specific method is proposed to apply to any time series. The included linear model, which utilizes inclined vector, yields satisfying degree of prediction power without estimating exact delay time ( $\tau$ ). The usefulness of this approach has been indicated not only theoretically but also in practical situation when the method was applied to economical time series analysis.

**키워드 :** 국소(局所)적 모델(Local linear model), 지연시간( $\tau$ )(Delay time( $\tau$ )), 기울기 벡터(Inclined vector), 기울기 끌개(Inclined attractor), 기울기 선형모델(Inclined linear model)

### 1. 서 론

지금까지의 카오스 예측방법은 삽입(Embedding) 벡터를 이용해서 재구성된 상태공간[4]에서 예측점벡터와 부근벡터의 관계를 고려하는 예측방법이 대부분이다[10, 11]. 따라서 3차원 이상의 고차 미분방정식으로부터 생성되는 시계열에서는 필연적으로 부딪히는 지연시간( $\tau$ )의 추정문제가 있다. 지연시간( $\tau$ )을 추정하는 방법으로는 자기상관함수, 상호정보량[12] 등이 있지만 자기상관함수는 로렌츠 시계열의 지연시간을 추정하면 50정도의 큰 값을 표시하고 있다. 이것과 비교하면 상호정보량은 15정도를 표시하고 그래프를 이용해서 판단하면 상호정보량쪽이 보다 적절하게 보인다. 그러나 상호정보량은 수많은 데이터를 필요로 하는점과 3차원 이상의 끌개에 대해서는 궤도의 교차가 생길 가능성이 있기 때문에 현실적이지 않다는 지적이 있다[1].

이 문제는 재구성된 삽입 벡터가 상태공간에서 적당히 퍼지지 않기 때문에 생기는 문제이다. 이것을 해결하기 위해

서는 카오스의 프랙탈성[6, 7]에 착안해서 각 벡터를 관측시계열로 미분하면 각 관측시계열의 움직임이 얻어진다. 새롭게 생성된 기울기 벡터로 재구성된 기울기 끌개로부터 예측점과 거리가 가까운 점을 찾아냄에 따라 상태공간에서 예측점과의 거리가 아니고 움직임이 비슷한 점을 찾아냄으로써 정확한 지연시간( $\tau$ )의 값에 얽매이지 않는 기울기 선형 모델을 제안한다.

이 기울기 선형 모델의 유효성을 표시하기 위해서 이 모델과 파머(Farmer) 등이 제안한 구분선형 모델[9, 5]을 로렌츠 시계열을 사용해서 지연시간( $\tau$ )과 삽입(Embedding) 차원( $m$ )을 변화시키면서 예측이라는 관점으로부터 비교했다. 그 결과 구분 선형모델보다 기울기 선형 모델 쪽이 예측오차가 작고 또 적은 데이터에서도 안정된 값을 표시했다. 그리고 실제 경제시계열에 적용시켜 그 실용성을 표시했다.

본 논문의 구성을 기술한다. 우선 2장에서는 종래의 카오스 예측방법과 문제점을 기술한다. 그리고 3장에서는 미분벡터의 도입을 4장에서는 이 알고리즘의 평가방법으로 예측오차를 조사했다. 5장에서는 경제시계열에 적용시킴으로써 그 유효성을 시험했다. 6장에서는 결론과 다음과제를 기술한다.

<sup>†</sup> 정 회 원 : 일본 지바대학 대학원  
논문접수 : 2001년 8월 18일, 심사완료 : 2002년 8월 22일

2. 종래의 카오스에측

주식시장뿐만 아니라 대부분의 동적 시스템은 복수의 요인이 관여되어있다. 그러나 시스템의 거동(舉動)에 관한 관측결과로써 단지 한 종류의 상태량에 관한 측정값밖에 이용하지 못하는 경우가 많다. 하나의 상태량에 대한 관측데이터로부터 다자유도(多自由度)시스템의 동적 특성이 해석 가능한 것임을 나타낸 정리가 타켄스(Takens)의 삽입(Embedding)정리이다. 삽입 정리에 대한 상세한 것은 문헌에 참고해주시 바란다[1].

어떤 샘플시간 간격으로 측정된 N개의 데이터 점으로 구성되는 시계열  $\{x(n)\}_{n=1}^N (n \in N, x(n) \in R)$ 이 주어진다 하자. 삽입 정리에 따르면 시차(時差)를 표현하는 적당한 정수  $(\tau)$ 간격으로 데이터 점을 선택해서 위상공간을 구성한다.

$$X_n(m) = \{x(n), x(n-\tau), \dots, x(n-(m-1)\tau)\}^T \quad (1)$$

$m$ 는 삽입 차원이라고 불린다. 시스템의 자유도가  $Q$ 라면 동적 특성의 해석을 보증하는 충분조건은  $m > 2Q$ 이다.  $Q$ 를 모를 경우에는 시행착오법에 의해  $Q$ 의 최적 값을 결정할 수 있다. 여러 가지  $m$ 값에 따라서 적당한 통계량을 계산해서 통계량이 정상이 되던가 또는 극치를 가지는  $m$ 값을 발견하면 된다.  $m$ 값보다도 어려운 것이  $\tau$ 의 선택이다. 한정된 길이의 시계열을 취급하는 경우, 다이내믹스의 성질이 잘 반영된  $\tau$ 의 값을 설정하지 않으면 안 된다. 간편한 결정방법은 시계열의 자기상관이 충분히 작게되는 최소의  $\tau$ 를 구하는 것이다.  $x(n)$ 와  $x(n+\tau)$ 간의 선형관계

$$x(n+\tau) - \hat{x} = \gamma(\tau)[x(n) - \hat{x}] \quad (2)$$

$$\hat{x} = \frac{1}{N} \sum_{n=1}^N x(n) \quad (3)$$

가 성립한다고 가정하자. 오차평가범함수(誤差評價汎函數)를

$$E_{acf} = \sum_{n=1}^{N-\tau} \{x(n+\tau) - \hat{x} - \gamma(\tau)[x(n) - \hat{x}]\}^2 \quad (4)$$

로 두고  $E_{acf}$ 가 최소 값을 가지는 계수  $\gamma(\tau)$ 를 추정하면

$$\gamma(\tau) = \frac{\sum_{n=1}^{N-\tau} [x(n) - \hat{x}][x(n+\tau) - \hat{x}]}{\sum_{n=1}^{N-\tau} [x(n) - \hat{x}]^2} \quad (5)$$

가 된다. 식 (5)은 자기상관함수의 정의 그대로이다. 자기상관이 충분히 작게되는  $\tau$ 를 삽입 벡터의 시차로써 선택한다면 벡터  $x(n)$ 의 각성분간의 선형상관은 무시할 수 있다. 각 성분간의 상관이 남아있다고 한다면 그것은 비선형 상관뿐이다. 자기상관에 근거한 시차의 결정은 실제로 잘 이용되

고 있다. 그러나 논리적으로 보다 설득력이 있는 것은 상호 정보량에 의한 결정방법이다.  $x$ 를 확률변수로 보고 시계열의 실제 값의 집합  $X = \{x(n)\}$  및 각 요소에 대해서  $\tau$ 의 시차를 가진 실제 값의 집합  $Y = \{y(n) = x(n+\tau)\}$ 을 생각하자. 각 집합에 있어서 실제 값에 대한 분포의 복잡함은 정보엔트로피  $H$ 로써 측정이 가능하다. 예를 들면  $X$ 에 대해서는

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (6)$$

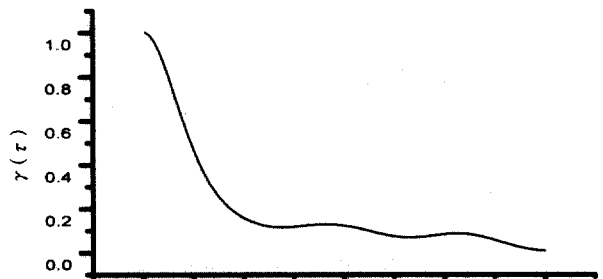
이다.  $p(\cdot)$ 은 실제 값의 확률함수이다. 여기서는 실제 값의 도수분포 즉 히스토그램으로 분포함수를 근사하여  $p(\cdot)$ 을 추정하는 것으로 한다. 같은 모양으로  $x(n), x(n+\tau)$ 의 2차원 히스토그램을 나타내는 결합분포함수로부터  $p(X, Y)$ 을 추정하면

$$H(X, Y) = - \sum_{x \in X, y \in Y} p(x, y) \log_2 p(x, y) \quad (7)$$

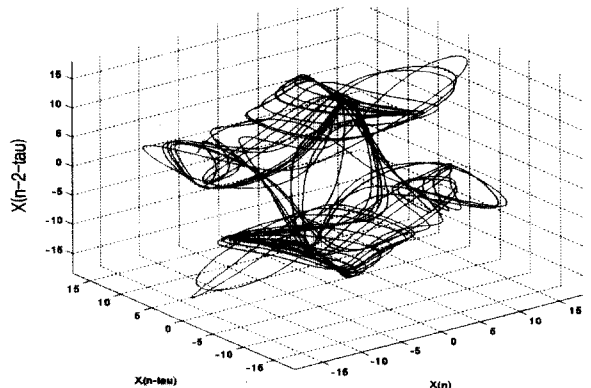
가 된다. Bayes의 정리에 따르면 조건부 확률  $p(Y|X)$ 은

$$p(Y|X) = \frac{p(X, Y)}{p(X)} \quad (8)$$

와 같이 쓸 수 있다.  $X, Y$ 에 관한 상호정보량의 정의는

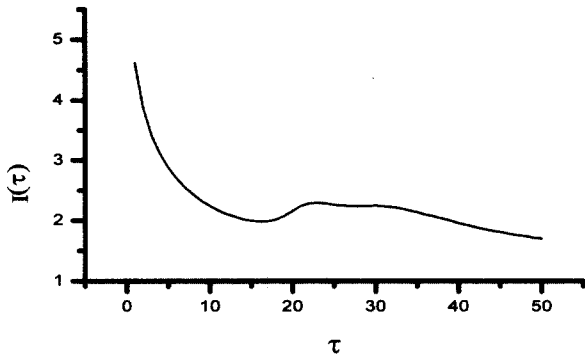


(a)

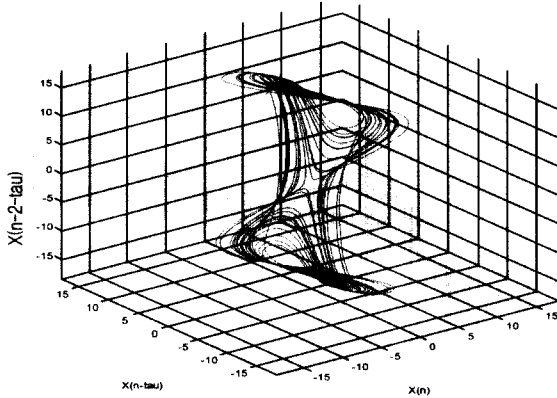


(b)

(그림 1) 로렌츠 모델의 자기상관함수와 지연시간  $(\tau)$  50으로 재구성한 끌개



(a)



(b)

(그림 2) 로렌츠 모델의 상호정보량 과 지연시간 (τ) 15로 재구성한 끌개

$$I(Y; X) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) \tag{9}$$

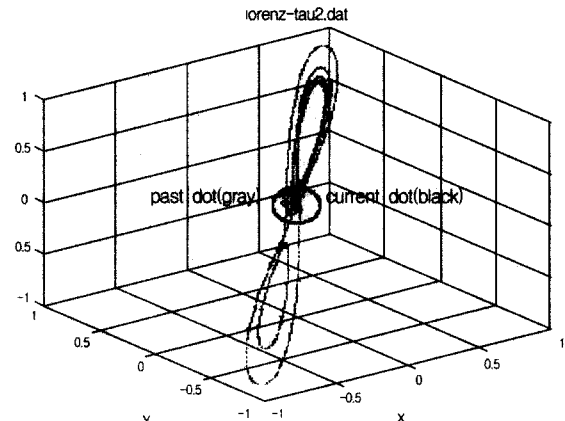
이다.  $I(Y; X)$ 은  $X$ 를 관측했을때  $Y$ 의 분포의 복잡함에 관하여 어느 정도 정보를 끼집어 낼 수 있는가를 측정하는 척도이다. 예를 들면  $X$ 와  $Y$ 가 서로 독립이라면  $p(X, Y) = p(X)p(Y)$ 가 성립하기 때문에  $I(Y; X) = 0$ 가 된다. 이렇게 해서 선형 다이나믹스에 의한 실제 값이든 비선형 다이나믹스에 의한 실제 값이든 상관없이  $X$ 와  $Y$ 에 대한 인과적 상관을 평가할 수 있다. 따라서  $I(Y; X)$ 을  $\tau$ 의 함수로 보고  $I(\tau)$ 가 적당한 레벨에서  $\tau$ 가 작아지는  $\tau$ 를 발견한다면 그  $\tau$ 로써 구성된 위상공간은 각성분간에 중복되는 정보가 적어진다는 의미에서 가장 효율 좋은 삽입(Embedding)이 실현된다고 생각되어진다. 실제의 운용상에서는  $I(\tau)$ 가 극소가 되는 최소의  $\tau$ 를 채용할 때가 많다.  $I(\tau)$ 가 단조 감소하는 경우에는  $I(\tau)$ 가 적당한 값 이하로 감소하는  $\tau$ 를 선택하면 된다.

(그림 1)(a)과 (그림 2)(a)는 로렌츠 모델의  $z$  성분에 대한 자기 상관함수와 상호정보량이다. 각각의 극소 값을 지연시간으로 재구성한 끌개가 (그림 1)(b)과 (그림 2)(b)이다. 여기서 로렌츠 모델 방정식은

$$\begin{aligned} \dot{x}(t) &= -\sigma x + \sigma y \\ \dot{y}(t) &= -\gamma x - y - xz \\ \dot{z}(t) &= xy - bz \end{aligned} \tag{10}$$

이다. 파라메타  $\sigma=10, \gamma=28, b=8/3$ 이다. 수치적분은 4차 룬쿠타법을 이용해서 시간스텝  $\Delta=0.01$ 로 초기조건은  $x(0)=10, y(0)=20, z(0)=30$ 으로  $N=10000$ 을 계산했다.

자기상관함수에 의해 구해진  $\tau \approx 50$ 의 경우, 재구성된 끌개를 보면 데이터사이의 상관이 약해서 마치 난수(random) 데이터와 같이 보인다. 상호정보량에 의해 구해진  $\tau \approx 15$ 의 경우, 재구성된 끌개는 3차원 위상공간에서 적절히 퍼져있다. 이것을 보면 자기상관함수 보다는 상호정보량의 우위성은 명확한 것처럼 보인다. 그러나 상호정보량에 의한 지연시간의 설정은 수많은 데이터를 필요로 한다는 점과  $(x_n, x_{n-\tau})$ 평면을 통과하는 궤도를 고려하기 때문에 3차원 이상의 끌개를  $(x_n, x_{n-\tau})$  평면에 그린다고 하면 궤도의 교차가 생길 가능성이 있어 이 때문에 특히 4차원 이상의 끌개에 대해서는 현실적이지 않다는 지적이 있다[1].



(그림 3) 지연시간 τ=2로 재구성한 로렌츠 끌개

이와 같이  $m$ 와  $\tau$ 가 정해지면 이 삽입 벡터를 이용해서 카오스 시계열을 예측할 수 있다. 삽입 벡터를 이용한 예측 방법을 통상 국소적 예측법이라 한다. 국소적 모델은 파머와 신드로 위치(Sindorowich)가 제안한 국소적 선형모델을 시작으로 1990년에는 스키하라와 메이가 심플렉스 투영법이라고 불리는 수법을 제안했는데 이것은 상태공간을 끌개상의 각 점을 정점으로 하는  $m$ 차원단체로 분할하는 수법이다. 이것에 대해서 1991년에는 미즈에 의한 끌개상의 각 점을 모점(母点)으로 하는 보르노이 분할을 이용한 예측수법들이 제안되어있다[2]. 그러나 이와 같은 예측방법들은 상태공간에서 예측점 벡터와 부근의 벡터와의 관계를 고려하는 예측방법이어서 3차원 이상의 고차미분방정식으로부터 생성되는 시계열에서는 지연시간 (τ)의 추정이라는 까다로운 문제에 부딪히게 된다. 이 삽입 벡터를 사용해서 끌개를 구성해서 예측할 때는 두 가지 문제점이 있다.

(그림 3)은 로렌츠 끌개의 z 성분을 지연시간  $\tau=2, N=550$ 으로 재구성한 것이다. 그림에서 표시하는 바와 같이 지연시간 ( $\tau$ )이 너무 작으면  $x_n$ 와  $x_{n-\tau}$ 의 상관성이 너무 강해져서 궤도가 상태공간에 충분히 퍼지지 않는다. 즉 국소적 예측법을 적용할 경우, 가령 현재점이 흑점이고 다음 점을 예측할 경우 과거의 궤도로부터 공간적으로 가까운 벡터를 선택하면 근방의 점중에 거리는 가까우나 궤도의 방향은 반대인 점(회색점)까지 포함되어 결과적으로 예측성능은 떨어진다. 반대로 지연시간이 너무 커지면  $x(n)$ 와  $x(n-\tau)$ 의 상관성이 소멸되어 벡터들은 무상관적인 데이터 열이 되고 끌개로서의 구조가 보이지 않게되어 또한 예측성능은 떨어진다. 두 번째는 과거의 데이터가 충분하지 않으면 정확한 정보를 가진 근방의 점이 적어져서 예측성능은 저하하는 결점이 있다.

### 3. 미분벡터의 도입

카오스 시계열은 프랙탈성질을 가진다. 프랙탈성질은 전체를 축소한 것이 부분을 구성하는 것과 같은 자기상이성질을 가진다[14]. 여기서는 프랙탈성질을 이용한 기울기 벡터를 제안한다. 프랙탈성질을 가진 시계열은 작은 부분의 궤도의 움직임과 큰 부분의 궤도의 움직임이 비슷하다. 이것을 이용해서 예측을 할 때 현재 점과 비슷한 움직임을 하는 벡터를 찾아서 현재 점을 기준으로써 축적을 조절해서 학습데이터로 사용한다. 구체적으로 설명하면 우선 전체로써 미지의 시스템을 해석하는데 필요한 모든 내부정보는 관측시계열에 삽입 되어있다고 가정한다.

식 (1)의 관측시계열  $x_n$ 을 제외한 새롭게 생성된 벡터  $(x_{n-\tau}, \dots, x_{n-(m-1)\tau})^T$ 을

$$\begin{aligned} x_{n-\tau} &= K^{(1)}(x_n) \\ x_{n-2\tau} &= K^{(2)}(x_n) \\ &\vdots \\ x_{n-(m-1)\tau} &= K^{(m-1)}(x_n) \end{aligned} \quad (11)$$

라고 관측시계열에 대한  $K$ 의 함수라고 생각할 수 있다. 즉

$$X_n(m) = (x_n, K^{(1)}(x_n), \dots, K^{(m-1)}(x_n))^T \quad (12)$$

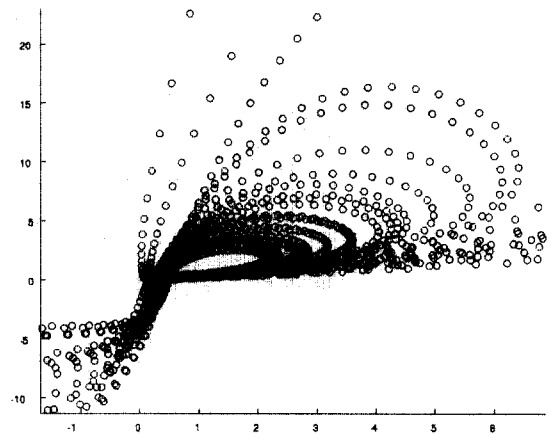
가 된다. 예측할 것이  $X_n(m)$ 의 제1성분이기 때문에 관측시계열  $x_n$ 로써 삽입 벡터를 나누면 관측시계열  $x_n$ 에 대해서 기울기벡터

$$\frac{X_n(m)}{x_n} = (1, w_{n-\tau}, \dots, w_{n-(m-1)\tau})^T \quad (13)$$

가 얻어진다. 이 벡터의 의미는 관측시계열의 움직임을 나타낸다. 관측할 때 시간스텝을  $\lim_{\Delta \rightarrow 0}$  접근하는 개념으로 생각

한다면 미분벡터라고도 말할 수 있다. 여기서

$$\begin{aligned} w_{n-\tau} &= \frac{K^{(1)}(x_n)}{x_n} \\ w_{n-2\tau} &= \frac{K^{(2)}(x_n)}{x_n} \\ &\vdots \\ w_{n-(m-1)\tau} &= \frac{K^{(m-1)}(x_n)}{x_n} \end{aligned} \quad (14)$$



(그림 4) 지연시간  $\tau=18$ 로 재구성한 기울기 끌개

이다. 관측시계열을 제외한  $(m-1)$ 차원의 공간에서  $W$  벡터로써 재구성된 새로운 기울기 끌개가 얻어진다.

1로 된 관측시계열을 제외한 새롭게 생겨난 기울기 벡터의 궤도는 (그림 4)와 같이 된다. (그림 4)는 로렌츠 모델로부터 얻어진 z 성분으로부터  $\tau=18$ 에서 얻어진 기울기벡터의 끌개이다. 시각  $n$ 에서 예측할 때에는 그 기울기 끌개로부터 예측점의 기울기와 공간적으로 가까운 기울기 벡터를  $L$ 개를 골라 예측 점의  $x_n$ 를 곱하면

$$\begin{aligned} \frac{X_{n-j_i}(m) \times x_n}{x_{n-j_i}} &= (x_n, w_{n-j_i-\tau} \times x_n, \dots, \\ &w_{n-j_i-(m-1)\tau} \times x_n) \end{aligned} \quad (15)$$

$(i=1, 2, \dots, L)$

가 얻어진다. 여기서  $j$ 는 기울기 끌개로부터 예측점의 기울기와 공간적으로 가까운 기울기 벡터의 번호를 나타낸다. 즉 이 벡터들을  $U_{n-j_i}(m)$ 라고 한다면

$$U_{n-j_i}(m) = \frac{X_{n-j_i}(m) \times x_n}{x_{n-j_i}} \quad (16)$$

로 된다. 이 시계열이

$$\frac{x_{n-j_i+1} \times x_n}{x_{n-j_i}} = A + B U_{n-j_i}(m) \quad (17)$$

에 따른다고 생각해, 기울기 끝개로부터 예측 점에 공간적으로 가까운 벡터  $L$ 개를 추닝에 이용한다. 단,  $A(m \times 1), B(m \times 1)$ 은 각 예측 점의 위치에 따라서 변화하는 계수이다. 지금부터는 이 알고리즘을 기울기 선형 알고리즘이라 부른다.

여기서 계산(除算)을 사용하므로 알고리즘의 안전성에 의문을 가질 수도 있다. 이것은 우선 예측을 행할때 관측시계열을 -1부터 1로 정규화하는 과정에서 정확히 0이 되는 시계열성분은 거의 없으므로 발산하는 경우는 드물다. 0에 무한히 가까이 가는 값이 있어도 나중에 예측점  $x_n$ 을 곱해서 보상하기 때문에 알고리즘으로써 충분한 안정성을 가진다. 만일 값 중에 정확히 0이 되는 값이 있다면 -0.9부터 1.1로 정규화 하면 된다. 그리고 이 알고리즘은 급격한 변동을 표시하는 1차 또는 2차 카오스 시계열의 예측결과는 이전의 예측방법보다 나쁘다. 이 예측방법은 3차이상의 미분방정식으로부터 얻어지는 시계열중에 관측 시간스텝을 각성분간의 상관이 충분히 존재할 수 있도록 짧게 한 시계열에 적용된다. 그래서 이 방법은 시계열의 자기상관함수를 조사했을 때 벡터간의 상관이 강할수록 적용도는 높다. 이 방법은 끝개의 전체로부터 움직임이 비슷한 점을 찾는 의미에서는 대역적 예측방법이지만 찾아낸 점이 몇 개의 위치에 집중되어 있는 점에서는 국소적 예측방법이라도 할 수 있다.

(그림 5)는 지연시간 ( $\tau$ )이 2와 8일 때 재구성한 로렌츠 끌개이다. 그림의 현재점(예측점)은 로렌츠 시계열의 500점 일 때를 표시하고 과거의 점은 501점을 예측할 때에 기울기공간으로부터 선택된 기울기가 비슷한 점 즉 학습데이터의 위치를 나타낸다. 괄호 안의 숫자는 기울기가 현재점과 가까운 순번을 나타낸다.  $\tau=2$ 에서는 현재점이 좁은 커브를 돌아 나오는 곳에 위치한다. 기울기 벡터로부터 얻어진 학습벡터들의 위치는 끝개위에서 각각 현재점과 같이 좁은 커브를 돌아 나오는 곳에 위치한다.  $\tau=8$ 에서는 좁은 궤도의 진입직전에 현재점이 위치한다. 학습벡터들도 똑같이 비슷한 움직임을 하는 곳에서 골라진다. 여기서 주의해야 할 점은 그림의 학습벡터는 궤도상에서의 벡터를 표시하는 것이 아니고 학습벡터의 위치를 표시하는 것이다. 즉 종래의 예측과의 차이는 학습벡터를 고를 때  $m$ 차원 공간에서 재구성된 점과의 거리에 의해 고르는 것이 아니고 똑같은 움직임을 하는 점을 고르는 것이다. 이 방법은 벡터간의 성분간의 상관이 강해질수록 움직임을 더욱더 정확히 잡을 수 있다. 따라서 지연시간에 따른 상태공간에서의 끌개의 퍼짐에는 관계를 하지 않는다. 즉 정확한 지연시간의 설정을 필요로 하지 않는다.

4. 예측 오차

이 알고리즘의 유효성을 표시하기 위해서 예측오차를 조사했다. 4.1은 지연시간 ( $\tau$ )의 변화에 따른 예측결과를 4.2는 차원  $m$ 의 변화에 따른 예측결과를 표시한다.

4.1 지연시간 ( $\tau$ )의 변화에 의한 예측결과

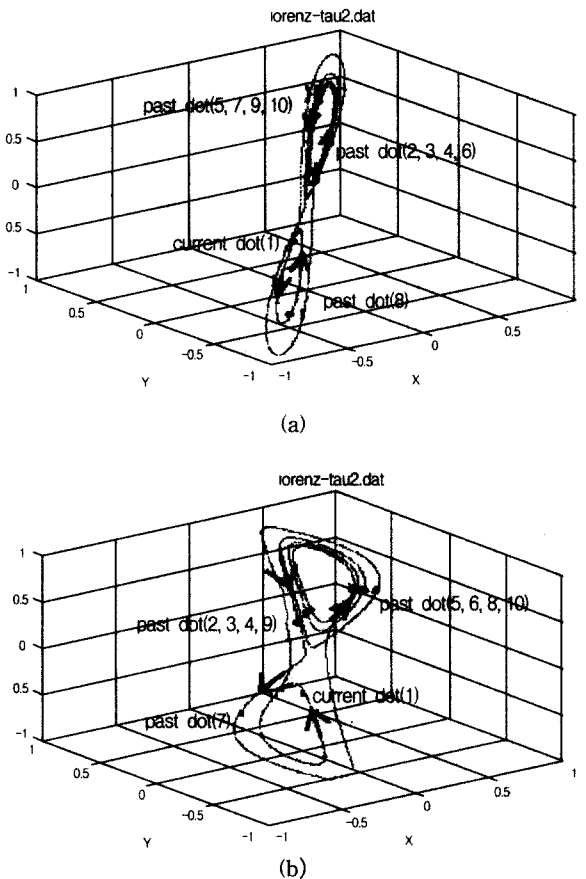
예측성능의 지표로써 일반적으로는 평균2승오차를 이용한다.

$$RMS = \sqrt{\frac{1}{S} \sum_{j=1}^S (\hat{x}_j - x_j)^2} \tag{18}$$

여기서  $\hat{x}_j$ 는 실치,  $x_j$ 는 예측치,  $S$ 는 예측점의 개수이다. 여기서는 이 기울기 선형모델을 RMS로써 평가했을 때 지금까지와 같이 적절한 지연시간 ( $\tau$ )의 값을 고려하지 않고도 좋다는 것을 표시한다. 다음으로 예측을 행할 때의 수법인데 여기서는 국소적 예측의 대표적인 수법인 파머 등이 제안한 구분(區分)선형화 모델과 기울기 선형화 모델의 두 종류를 사용했다. 기울기 선형화 모델을 여기서는 제안수법(Method of a proposal)1이라고 표시한다. 구분선형화 모델은 타겐스의 정리로부터 작성한 지연시간 간격으로 구성된 삼입 벡터를  $m$ 차원 공간 내에 배치했을 때 그때의 시계열이

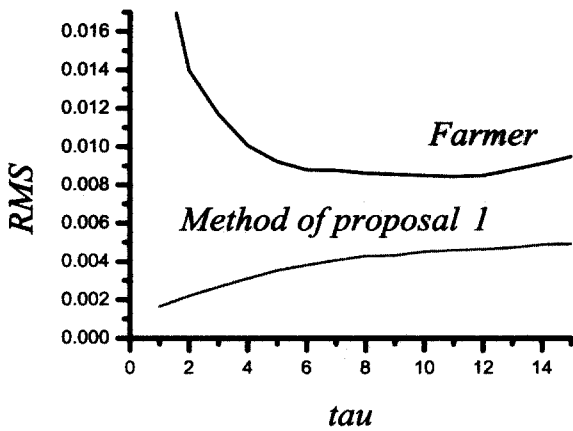
$$X_i(m) = A + BX_{i-1}(m) \tag{19}$$

를 따른다고 생각해, 예측점에 공간적으로 가까운 벡터  $L$ 개

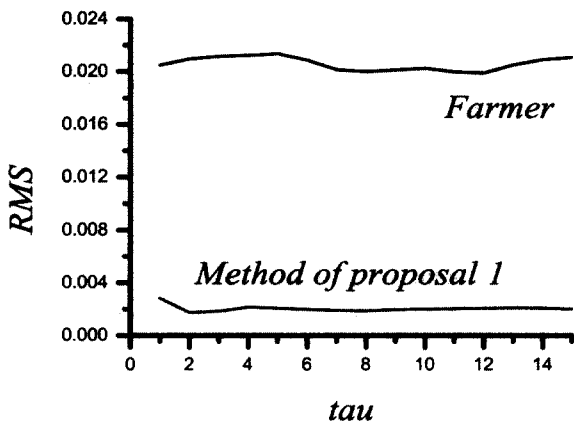


(그림 5)  $\tau=2$ 와  $\tau=8$ 에서의 예측점과 학습 데이터의 위치

를 추닝에 사용하는 모델이다. 단  $A(m \times 1), B(m \times m)$ 은 각 예측점의 위치에 따라서 변화하는 계수이다. (그림 6)은 파머와 제안수법 1에 의해 구해진 RMS를 표시한다. 사용된 데이터는 로렌츠 모델로부터 얻어진 데이터로써 초기치로의 의존성을 고려해서 공회전을 4000번 행한 후에 얻은 4000개의 점으로 1001점부터 4000점까지 총 3000점을 지연시간( $\tau$ )의 값을 변화시키면서 예측한 결과이다. 여기서 차원  $m=3$ , 학습벡터의 수는  $L=15$ 로 했다. 파머의 방법은 지연시간이 적절한 상관을 가지는 값까지 예측오차는 적어진다. 그러나  $x_n$ 와  $x_{n-\tau}$ 사이의 상관성이 없어지기 시작하면 예측오차는 증가한다. 이것에 대해서 제안수법 1은 전체적으로 파머 모델 보다 예측오차는 대폭 향상된 결과를 표시한다. 제안수법 1만을 본다면  $\tau=1$ 일 때, 예측오차는 가장 적은 값을 표시하고 그후에 서서히 증가한다. 그러나 이와 같이 충분한 데이터가 있는 경우는 별로이지만 실생활에서는 약 1000점 정도의 시계열 데이터로써 해석을 해야만 하는 경우가 종종 있다.



(그림 6) 로렌츠 시계열(4000)의 예측결과



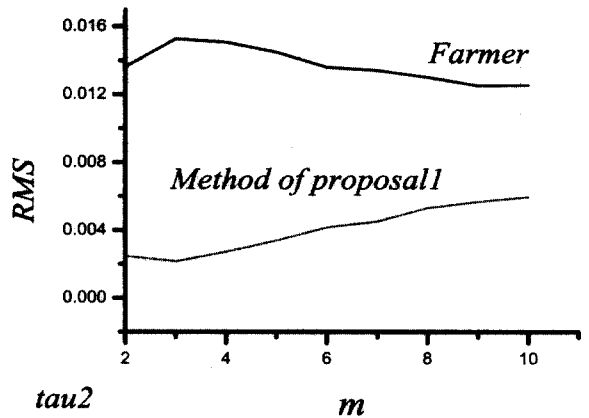
(그림 7) 로렌츠 시계열(1200 점)의 예측결과

(그림 7)은 로렌츠 시계열을 1200점 준비해서 300점부터 1200점까지 총 900점을 지연시간( $\tau$ )의 값마다 예측한 결과

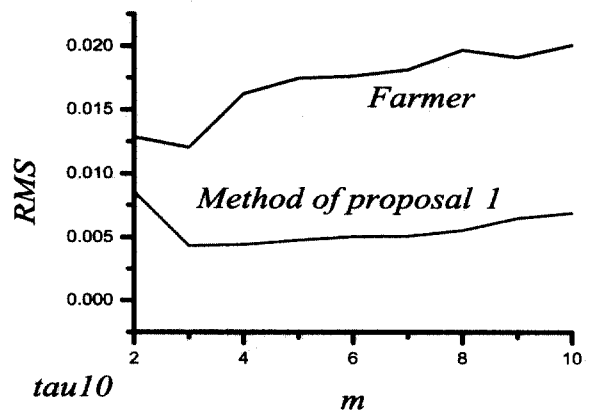
를 표시한다. 데이터의 수가 적어지게 되면 파머의 방법은 예측오차가 적절한 지연시간( $\tau$ )의 값이 되어도 높은 값을 그대로 유지한다.

이것에 대해서 제안수법 1은 적은 데이터에도 불구하고 데이터의 수가 많을 때와 거의 같은 정도(精度)를 표시한다. 이것은 학습데이터를 삽입 벡터로 재구성한 끝개의 전체로부터 찾아냄으로써 국소적 예측방법보다 보다 정확한 정보를 가진 벡터를 학습데이터로서 사용하기 때문이라고 생각되어진다. 그리고 이 방법은 벡터의 각 성분간의 상관성이 가장 강한  $\tau=1$ 의 값에서 최저의 예측오차를 표시한다.

4.2 삽입 차원 ( $m$ )의 변화에 따른 예측오차



(그림 8)  $\tau=2$ 에서의 예측결과.



(그림 9)  $\tau=10$ 에서의 예측결과

다음은 차원  $m$ 의 변화에 따라서 예측오차가 어떻게 표시되는가를 조사했다. (그림 8)과 (그림 9)는 지연시간( $\tau$ )이 2와 10의 경우,  $m$ 의 값에 대해서 RMS의 결과를 표시한다. 사용된 데이터의 수는 1500점으로써 501점부터 1500까지 총 1000점을 예측했다. 학습데이터의 수는 15개로 했다. 지연시간( $\tau$ )이 2의 경우, 파머의 방법은 로렌츠 끝개의 차원  $m=3$ 에서 예측오차가 가장 높게 나타났다. 적절한 지연시간( $\tau$ )의 값이 아니면 정확한 차원으로 예측해도 예측오차

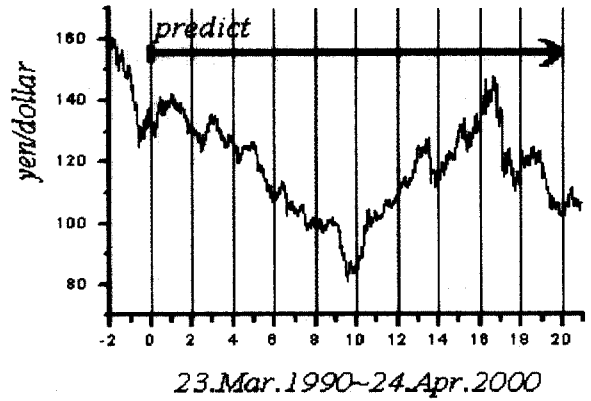
는 오히려 높게 나타났다. 지연시간( $\tau$ )이 10의 경우는 예상대로  $m=3$ 에서 예측오차는 가장 낮아졌다. 그후 예측오차는 서서히 증가한다.

이것에 대해서 제안수법 1은  $m=3$ 에서 지연시간( $\tau$ )이 2와 10일 때 양자 모두 예측오차는 가장 작게 적게 나타났고 그후에 서서히 증가한다.

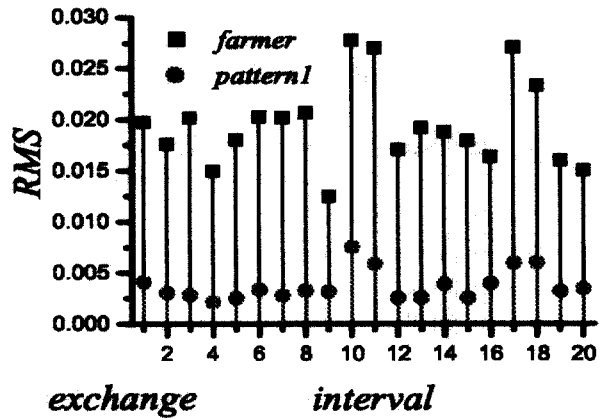
이상의 것을 종합하면 파머의 방법은 적절한 지연시간이 아닐 경우 또는 데이터의 수가 충분하지 않을 경우, 시계열로부터 얻어진 정보는 대단히 불안정한데 반하여 제안수법 1은 벡터의 각 성분간의 상관성이 가장 강한  $\tau=1$ 에서 최저의 예측오차를 표시했으며 적은 데이터 수에서도 좋은 예측결과를 얻었다.

5. 경제시계열로의 응용

실제 경제시계열에 이 알고리즘을 적용했을 때 어떠한 예측결과가 나타나는가를 조사했다. 해석에 사용된 시계열 데이터는 일본경제평균 주가와 엔과 달러의 환율 데이터의 그 날 마지막 값으로 각각 1990년부터 1999년, 1990년부터 2000년

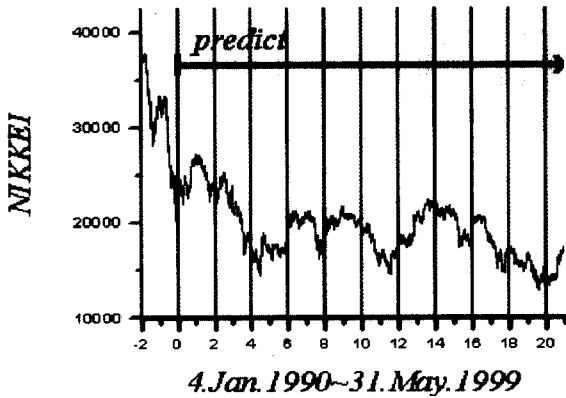


(a)

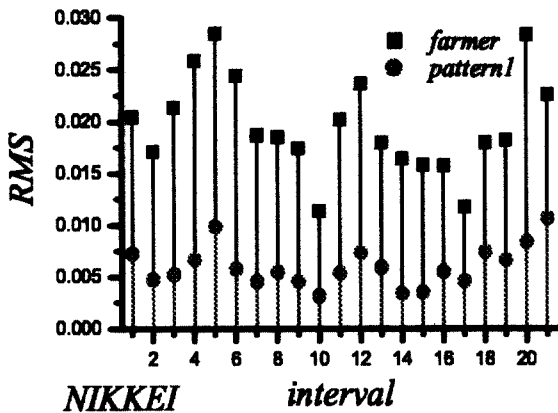


(b)

(그림 11) 환율의 예측결과



(a)



(b)

(그림 10) 일본경제평균의 예측결과

까지의 약 2300점, 2200점의 데이터를 사용했다(그림 10)(a), (그림 11)(a). 노이즈성분을 제거하기 위해 전후 5일간의 이동평균을 구했다. 각각 201점부터 일본경제평균은 2100점을 환율데이터는 2000점을 예측해서 100점마다 블록화해서 각각 21블록, 20블록으로 구분해서 RMS를 구했다. (그림 10)(b)와 (그림 11)(b)는 일본경제평균과 환율시계열의 RMS를 표시했다. 이 데이터는 파머의 방법과 제안수법 1의 방법 양자 모두  $m=3$ 부터 9까지,  $\tau=2$ 부터 10까지 총 63회를 예측해서 평균을 낸 결과이다. 파머의 방법은  $m=3$ ,  $\tau=10$ 에서 가장 낮은 예측결과를 내었고 제안수법 1의 방법은  $m=3$ ,  $\tau=1$ 에서 가장 낮은 예측결과를 나타냈는데 오차의 대소 차는 있지만 대체적으로 예측결과가 (그림 10)과 (그림 11)과 비슷하였다. 여기서는 예측오차의 평균치 그래프를 실었다. 예측결과를 보면 파머의 방법보다 제안수법 1이 RMS에서 좋은 결과를 나타내고 있다는 것을 알 수 있다.

6. 결 론

종래의 삼입 벡터를 이용한 국소적 예측방법은 고차미분 방정식으로부터 생성되는 카오스시계열을 예측할 때, 파라

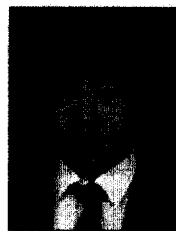
메타중에 지연시간( $\tau$ )의 추정치가 정확하지 않으면 예측성능은 떨어진다. 지금까지 지연시간( $\tau$ )의 값을 추정하는 방법은 많이 제안되어있지만 실제 고차미분방정식으로부터 생성되는 시계열에 모두 적용할 수 있는 방법은 아직 없다. 이것을 기울기벡터를 이용한 기울기 선형 모델을 도입함으로써 정확한 지연시간( $\tau$ )의 값을 추정하지 않고도 예측성능에서 만족할 수 있는 결과를 보였다. 이것을 이론만이 아니라 실제 경제시계열에 적용함으로써 종래의 예측방법과 비교해서 그 유효성을 증명했다.

이후의 과제는 지금 현재 시스템으로부터 카오스 물리량을 측정할 때는 대부분 상태공간에의 삽입 벡터를 이용한다. 이것을 기울기 벡터를 도입한 기울기 끝개의 재구성방법의 개념으로 카오스의 물리량을 측정하는 방법을 생각한다.

### 참 고 문 헌

[1] 松葉 育雄, "非線型時系列解析", 朝倉出版, 2000.  
 [2] 合原 一幸, "카오스의數理と技術", 放送大學教育振興會, 1997.  
 [3] 合原 一幸, "ニューロ.ファジィ.카오스", 오ム社, 1993.  
 [4] F. Takens, "Detecting strange attractors in turbulence," in Dynamical Systems and Turbulence, Lecture Notes in Mathematics, eds. D. A. Rand and L. S. Young, Berlin, Vol.898, pp.366-381, Springer, 1981.  
 [5] J. D. Farmer, "Chaotic attractors of an infinite dimensional system," Physica D, Vol.4, p.366, 1982.  
 [6] B. B. Mandelbrot, "Fractals form Chance and Dimension," Freeman, San Francisco, 1977.

[7] B. B. Mandelbrot, "Fractals and Scaling in Finance," Discontinuity, Concentration, Risk. Springer-Verlag, 1997.  
 [8] 田中賢, 王麟元, 松葉 育雄, "經濟時系列データの構造變化と豫測", 信學論(A), J81-A, No.4, pp658-663, 1998.  
 [9] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," Phys. Rev. Lett, Vol.59, pp.845-848, 1987.  
 [10] 木村 充也, 松葉 育雄, "大域的.局所的手法による時系列の非線型豫測," CAS95-91, NLP95-82, 1996.  
 [11] 松葉 育雄, "카오스と豫測", 數理科學, No.348, pp.64-69, 1992.  
 [12] A. Wolf, J. B. Swift, H. L. Swinney, and J. Vastano, "Determining Lyapunov exponents from a time series," Physica D, Vol.16, pp.285, 1985.  
 [13] I. Matsuba, H. Suyari, S. Weon, and D. Sato, "Practical Chaos Time Series Analysis with Financial Applications," IEEE, Proceedings of ICSP2000, pp.265, 2000.  
 [14] 원석준, "환율데이터의 장기기억성과 정답율", 정보처리학회 논문지, 제7권 제12호, pp.3866-3873, 2000.



### 원 석 준

e-mail : weon@j94.tg.chiba-u.ac.jp

1989년 부산공업대학 전자공학과 졸업 (학사)

1991년 부산 동아대학산업대학원 전산공학과 졸업(전공 제어계측),(석사)

1994년~현재 일본 지바대학 대학원 박사과정

관심분야 : 카오스 비선형해석 뉴럴네트워크 등