

생체 카오스의 최적 어트랙터 재구성에 관한 연구

°장 재 호, 이 병 채, 이 명 호

(A Study on Optimal Attractor Reconstruction of Biological Chaos)

Jaeho Jang, Byungchae Lee, and Myoungcho Lee
Dept. of Electrical Engineering, Yonsei University
Seoul, 120-749 Korea

Abstracts

This paper proposes an fill-factor algorithm that determines embedding parameters which are needed in optimal attractor reconstruction. For reliability test, using this algorithm, we reconstructs the attractor of numerical chaotic data such as Duffing equation, Lorenz equation and Rossler equation whose embedding parameters are known.

Also we reconstructs the attractor of experimental data and evaluates correlation dimension. Experimental data used in this paper are 38 ECG data of AHA(American Heart Association) ECG database.

For numerical chaotic data, correlation dimension and Lyapunov exponent of reconstructed attractor are very close to those of attractor using original coordinate system.

1. 서 론

시계열 데이터로 부터 카오스 어트랙터의 동역학 변수를 추정하기 위해서는 적절한 어트랙터의 재구성이 필요하다. 대부분의 실험환경에서는 단지 하나의 스칼라 시계열 데이터만이 측정 가능하기 때문에 이렇게 얻어진 하나의 스칼라 시계열 데이터로 부터 모든 관련된 요소를 얻어내어야 한다. Packard[1]등은 위상공간 재구성 방법으로 측정 가능한 스칼라 시계열 데이터의 미분치(derivative)를 새로운 좌표계로 사용하는 방법을 제안했다. 하지만 이 방법은 계산상 많은 시간이 필요하고 잡음에도 민감하다.

Takens[2]에 의해 제안이 되어 최근에 많이 사용

되고 있는 방법은 지연시간 좌표계를 사용한다. Takens에 의하면 이러한 재구성법은 동역학 변수에는 영향을 주지 않으면서 위상학적으로 등가의 어트랙터를 제공한다.

이 방법의 경우 기본적인 개념은 명확하지만 재구성을 위한 실제 과정에서는 시계열 데이터의 불충분한 조건 때문에 발생하는 몇가지 문제점을 해결해야 한다. 일반적으로 발생하는 문제점은 정확히 알려지지 않은 자유도(degree of freedom)를 가진 시스템으로부터 얻어진 잡음이 섞인 유한한 갯수의 데이터만을 사용할 수 있음으로 인해 발생한다. 그러므로 실지의 재구성에서는 프랙탈 차원과 Lyapunov 지수를 정확히 계산하고 더불어 계산시간의 감소를 위해 지연시간과 임베딩 차원의 적절한 선택은 매우 중요하다.

본 연구에서는, 최적의 지연시간을 찾아주고 임베딩차원을 추정하는 새로운 방법을 제안하고자 한다. 이 방법은 순수하게 어트랙터의 기하학적인 면만을 고려한 방법으로써 직접적으로 지연시간과 임베딩 차원을 동시에 얻을 수가 있다. 또한 이 방법은 수식적인 알고리즘이 쉽게 구현될 수 있고 고차의 임베딩 차원에 대해서도 소형 computer를 이용하여 계산작업을 할 수 있는 장점이 있다.

2. 어트랙터 재구성방법

Takens의 이론에 의하면 스칼라 시계열 데이터로 부터 위상공간에 어트랙터를 재구성할 수 있다.

지금까지 임베딩 변수를 결정하기 위한 방법들에 관한 많은 연구가 있었다. 우선 자기상관함수 (autocorrelation function)가 최초로 0을 크로싱 하는 순간은 두 좌표계의 선형독립(linearly independent)을

보장한다는 사실에서 지연시간을 결정하는 방법[3]이 소개되었다. 하지만 이 방법은 고차 system의 경우 다른 좌표계들 사이의 correlation으로 인해 부정확한 결과를 야기시킨다.

그 후 위의 ACF가 선형독립을 보장하는데 비해 두 변수사이의 좀 더 일반적인 종속성을 제공하는 상호정보량(mutual information)을 이용한 방법[4]이 제안되었다. 이 방법 또한 4차 이상의 고차원에는 비실용적이라는 문제점을 가지고 있다.

다음으로 일반화된 correlation integral의 로그함수를 취한 값이 최초로 최소값이 되는 순간에서 지연시간을 결정하는 방법[5]도 제안되었지만 이상의 방법들의 문제점은 모두가 다른값의 지연시간을 결정했고 정확한 임베딩 차원에 관한 정보도 제공하지 않는다는 것이다.

반면에 임베딩 차원을 구하기 위해 singular value decomposition(SVD)을 이용한 방법[6]이 제안되긴 했지만 이 방법은 반대로 최적의 지연시간에 관한 기준을 제공하지 않았다.

이상의 방법들이 각각 임베딩 변수의 어느 하나에 관해서만 촛점을 맞추는데 비해 Schuster등[7]이 제안한 방법은 지연시간과 임베딩 차원을 동시에 구할 수 있다는 점에서 주목할만한 것이었다. 이 방법은 Takens의 첫번째 이론에 기초를 둔 것으로 임베딩이란, 위상학자적인 맵핑이란 점을 이용하고, 서로 다른 재구성 어트랙터로부터 이웃한 점들 사이의 관계를 비교, 분석함으로써 임베딩 변수를 동시에 구할 수 있었다. 하지만 이 방법도 복잡한 분류, 탐색 알고리즘으로 인해 computer의 계산시간이 많이 걸릴 뿐 아니라 잡음에 민감하다는 단점을 가지고 있다.

이상의 문제점들을 고려해서 본 연구에서 제안하는 방법은 순수하게 기하학적인 면만을 고려한 fill-factor 알고리즘을 제안한다. 이 방법은 두가지 임베딩 변수를 동시에 결정할 수 있을 뿐 아니라 알고리즘의 구조도 상대적으로 쉽다는 장점을 가지고 있다.

3. Fill-Factor 알고리즘

이 방법의 기본 개념은 어트랙터를 구성하는 궤적상의 점들이 최대한 분리되게 함으로써 임베딩 공간을 폭넓게 사용하는 최적의 이용도를 요구한다.

주어진 임베딩 차원에서 하나 혹은 그 이상의 좌표계의 충돌을 막기 위해서는 모든 좌표계에 똑같은 가중치를 부여한 체적을 정의한다. 어트랙터의 구조를 먼저 알 수는 없기 때문에 많은 수의 임의로 선택된 체적들을 평균내어야만 한다.

아래에서 설명될 알고리즘은, τ 가 변함에 따라 재구성된 어트랙터의 점들에 의해 정의된 평균된 타원체들의 체적을 최대화시킨다. 이 알고리즘에서 사용할 체적은 다음과 같은 방법으로 정의한다.

재구성을 위해 정의된 상태 벡터에 의해 어트랙터를 재구성한 후 (d_E+1) 개의 임의의 index $r_m \in S_0$ 을 선택한다.

$$R_0 = \{ r_m \in S_0 ; m = 0, \dots, d_E \} \quad (1)$$

각각의 r_m 은 타원체의 구석을 나타내고 (d_E+1) 개의 구석은 d_E -차원 공간에서 위의 타원체가 어떤 공간을 차지하도록 한다. 타원체가 임의의 공간을 차지하도록 하는 벡터를 계산하기 위해서는 기준점의 index인 r_0 를 선택한다.

그런 후 다음과 같은 변위 벡터를 정의한다.

$$d_n(t_{r_0}) \equiv \begin{pmatrix} \xi(t_{r_0}) - \xi(t_{r_n}) \\ \xi(t_{r_0} + \tau) - \xi(t_{r_n} + \tau) \\ \dots \\ \xi(t_{r_0} + \tau(d_E-1)) - \xi(t_{r_n} + \tau(d_E-1)) \end{pmatrix} \quad (2)$$

여기서 $n=1, \dots, d_E$ 이다.

위의 변위벡터 d_n 은 타원체를 정의하기 위한 충분한 많은 수의 모서리를 제공한다. 이러한 타원체의 체적을 얻기 위해서는 다음과 같은 $(d_E \times d_E)$ -행렬을 정의해야 한다.

$$M_{d_E, r_0}(\tau) = (d_1, \dots, d_{d_E}) \quad (3)$$

그리고 이 행렬의 determinant를 계산해야 한다.

$$V_{d_E, r_0}(\tau) = | \det [M_{d_E, r_0}(\tau)] | \quad (4)$$

타원체의 평균 체적이 커지면 커질수록 어트랙터에 의해 채워지는 공간의 체적 또한 점점 더 커진다.

우리는 여기서 어트랙터의 체적을 측정하는 방법을 정의한다. 그러기 위해서는 타원체에 대해 충분히 큰 수인 N_v 를 선택하고 이들의 평균 체적을 계산한다.

$j=1, \dots, N_v$ 일때 j 번째 타원체 $V_{d_E, r_0(\tau)}$ 의 체적을 다음과 같이 정의한다.

$$F_{d_E}(\tau) = \frac{\frac{1}{N_v} \sum_{j=1}^{N_v} V_{d_E, r_0(j)}(\tau)}{(\max_{(k \in K_0)} \xi(t_k) - \min_{(k \in K_0)} \xi(t_k))^{d_E}} \quad (5)$$

임베딩 차원이 증가함에 따라 $F_{d_E}(\tau)$ 은 d_E 에 대해 지수함수적으로 감소하기 때문에 식(5)의 로그함수를 사용하는 것이 편리하다.

$$f_{d_E}(\tau) \equiv \log [F_{d_E}(\tau)] \quad (6)$$

이러한 어트랙터의 공간 이용도를 fill-factor라 한다.

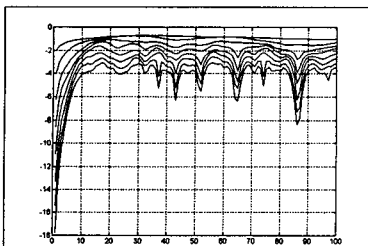
4. 실험결과

* 수치 데이터에 대한 실험

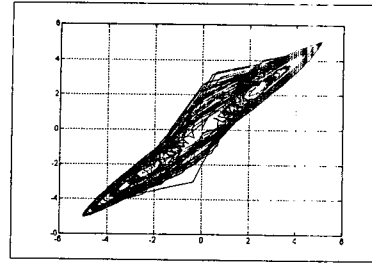
$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= F \cos(x_3) - D x_2 - x_1 - x_1^3 \\ \dot{x}_3 &= w \end{aligned} \quad (7)$$

여기서, $D=0.2$, $F=40$, and $w=2\pi/T_c=1$

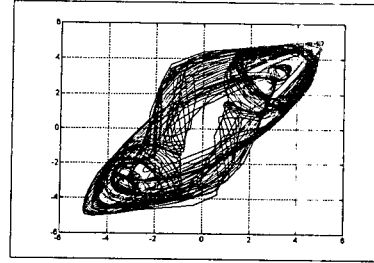
재구성은 x_1 좌표계의 데이터를 사용했다. $f_{d_E}(\tau)$ 를 계산하기 위해 각각의 (τ, d_E) 에 대해 N_v 를 전체 데이터 점들의 수($N_{dat}=32768$)의 약 2%에서 결정했다. 그림 1(a)는 임베딩 차원 $d_E=[2, 7]$ 에 대해 지연시간이 $\tau = 0$ 부터 $\tau = 100$ 까지 변할때의 계산된 fill-factor를 보여준다.



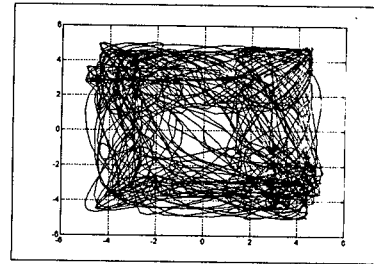
(a)



(b)



(c)



(d)

그림 1 : Duffing 시스템의 x_1 좌표계를 이용한 재구성

(a) Fill-factor(Embedding 차원 : 2~7)

(b) $\tau = 3$ (c) $\tau = 17$ (d) $\tau = 37$

그리고 그림 1의 (b),(c),(d)는 각각의 지연시간에서의 재구성된 어트랙터 구조를 보여준다.

그림 1(a)에서 알 수 있듯이 임베딩 차원은 4이상이어야 하고 지연시간은 17이 적절한 것으로 판단된다.

즉, 임베딩 차원이 4이상인 경우에 대해서는 τ 가 변함에 따라 fill-factor는 어떤 구조를 보여준다.

위에서도 언급했듯이 적절한 재구성은 $f_{d_E}(\tau)$ 이 최대일때 기대할 수 있고 잘못된 재구성은 $f_{d_E}(\tau)$ 이 최소인 곳에서 발생한다.

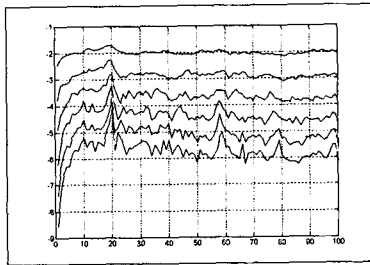
Fill-factor 알고리즘에 의해 수치 데이터의 임베딩 파라미터를 결정한 후 상관차원과 Lyapunov 지수를 구한 결과는 다음과 같다.

표 1 : 수치 데이터의 상관 차원과 Lyapunov 지수

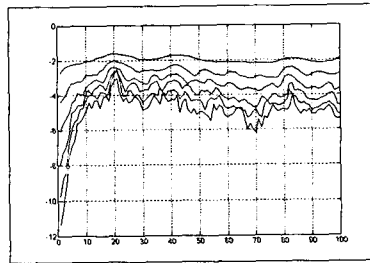
	상관차원		Lyapunov 지수	
	실제값	재구성에 의한 값	실제값	재구성에 의한 값
Duffing equation	1.6259	1.6002		
Lorenz equation	2.4044	2.4421	1.500	1.632
Rosler equation	2.6733	2.7903	0.090	0.095

* 실험 데이터에 대한 실험

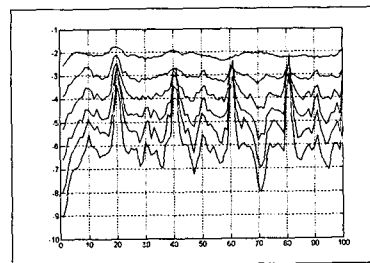
다음은 미국 심장학회(AHA) 권고 부정맥 데이터에 fill-factor 알고리즘을 적용시킨 예이다.



(a)



(b)



(c)

그림 4 : AHA 권고 부정맥 데이터의 fill-factor

Fig 4 : Fill-factor for AHA ECG database

(a) B2202 (b) B2205 (c) B2207

AHA 데이터의 경우 임베딩 차원은 최소한 5이상으로 정해야 하고 지연시간은 역시 20 전후에서 결정되어야 함을 알 수 있다.

다음의 표는 AHA 부정맥 데이터를 fill-factor 알고리즘을 이용하여 재구성을 할 경우 결정된 임베딩 파라미터의 값들이다.

표 2 : AHA 부정맥 데이터의 지연시간 (임베딩 차원은 모두 5이상)

	지연 시간		지연 시간		지연 시간
B2201	20	B3206	20	B5201	23
B2202	19	B3207	21	B5202	22
B2203	20	B3208	16	B5203	19
B2205	20	B3210	18	B5204	20
B2206	20	B4202	19	B5205	18
B2207	17	B4203	19	B5206	17
B2208	19	B4204	19	B5207	20
B2209	19	B4205	21	B5208	18
B2210	20	B4206	21	B5209	20
B3201	22	B4207	20	B5210	20
B3202	21	B4208	20	B6201	20
B3204	18	B4209	19	B6202	19
B3205	20	B4210	18		

다음은 fill-factor 알고리즘을 이용하여 AHA 데이터 베이스의 임베딩 변수를 구한 후 재구성된 어트랙터의 상관차원을 구한 결과이다.

표 3 : AHA 부정맥 데이터의 상관 차원

	상관차원		상관차원		상관차원
B2201	3.1246	B3206	3.1242	B5201	3.2376
B2202	3.6854	B3207	3.5112	B5202	3.7223
B2203	3.3547	B3208	3.3475	B5203	3.7213
B2205	3.2683	B3210	3.5645	B5204	3.2876
B2206	3.3754	B4202	3.2556	B5205	3.3762
B2207	3.5044	B4203	3.2804	B5206	3.2832
B2208	3.2828	B4204	3.4434	B5207	3.2815
B2209	3.4572	B4205	3.2094	B5208	3.1872
B2210	3.1253	B4206	3.7629	B5209	3.2376
B3201	3.6734	B4207	3.8973	B5210	3.2917
B3202	3.1004	B4208	3.3409	B6201	3.2832
B3204	3.5673	B4209	3.7364	B6202	3.2763
B3205	3.5667	B4210	3.3764		

5. 결론 및 토의

최적의 어트랙터 재구성을 위한 임베딩 파라미터의 설정법에 관한 알고리즘으로 Fill-factor 알고리즘을 제안하고 기존의 알려진 수치 데이터인 Duffing 방정식과 Lorenz 방정식 그리고 Rossler 방정식의 임의의 한 좌표계의 값을 이용해 최적의 어트랙터 재구성을 해 본 결과 기존에 알려진 수치에 매우 근접함을 알 수 있었다.

이를 기반으로 본 연구에서 제안된 알고리즘을 AHA 부정맥 데이터에 적용시켜 본 결과 임베딩 차원은 최소한 5이상이고 지연시간은 최소 16에서 최대 23까지로써 데이터에 따라 다른 값을 보였다.

즉, ECG 파형을 만들어내는 시스템의 자유도가 4이고 신호에 특성에 따라 다른 지연시간에 의해 어트랙터가 재구성 되어야 함을 알 수 있었다.

6. 참고 문헌

- [1] N. H. Packard, J. P. Crutchfield J. D. Farmer and R. S. Shaw, "Geometry from a time series", *Phys. Rev. Lett.* 45(1980), pp. 712-716
- [2] F. Takens, "Dynamical Systems and Turbulence", Warwick, 1980 edited by D. A. Rand and L. S. Young, *Lecture Notes in Mathematics* Vol. 898 (Springer-Verlag, Warwick, 1980), pp. 230
- [3] H. G. Schuster, "*Deterministic Chaos*" (VCH-Verlag, Weiheim, 1988)
- [4] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information", *Phys. Rev. A* 33(1986), pp. 1134-1140
- [5] H. G. Schuster and W. Liebert, "Proper choice of the time delay for the analysis of chaotic time series", *Phys. Lett A* 142(1988), pp. 107-111
- [6] D. S. Broomhead and J. P. King, "Extracting qualitative dynamics from experimental data", *Physica D* 20(1986), pp. 217-236
- [7] W. Liebert, K. Pawelzik and H. G. Schuster, "Optimal embeddings of chaotic attractors from topological considerations", *Europhys. Lett.* 14(1991), pp. 521-526