RECONSTRUCTION OF CHAOTIC DYNAMICS USING A NOISE-ROBUST EMBEDDING METHOD

Wako Yoshida †, *Shin Ishii* † ‡ *and Masa-aki Sato* ‡

† Nara Institute of
Science and Technology
8916-5 Takayama Ikoma Nara Japan

‡ ATR Human Information
Processing Research Laboratories
2-2 Hikaridai Seika Soraku Kyoto Japan

ABSTRACT

In this article, we discuss the reconstruction of chaotic dynamics in a partial observation situation. As a function approximator, we employ a Normalized Gaussian network (NGnet), which is trained by an on-line EM algorithm. In order to deal with the partial observation, we propose a new embedding method based on smoothing filters, which is called integral embedding. The NGnet is trained to learn the dynamical system in the integral coordinate space. Experimental results show that the trained NGnet is able to reproduce a chaotic attractor that well approximates the complexity and instability of the original chaotic attractor, even when the data involves large noise. In comparison with our previous method using delay coordinate embedding, this new method is more robust to noise and faster in learning.

1. INTRODUCTION

Time-series prediction for unknown dynamical systems is an important topic in signal processing. When a part of all dynamical variables is available, embedding techniques is often used. In this article, we propose a new embedding method based on smoothing filters for a continuous-time dynamical system. We discuss the reconstruction of a chaotic dynamical system using our new embedding method. As an example, we consider the Lorenz system [1], which is a popular three-dimensional chaotic dynamical system.

In order to approximate a vector field of a dynamical system, function approximators like neural networks have often been used. In this study, we employ a Normalized Gaussian network (NGnet) [5] as a function approximator. The NGnet is a network of local linear regression units. This model softly partitions the input space by using normalized Gaussian functions, and each local unit linearly approximates the output within its partition. The NGnet was trained by an on-line EM algorithm, which we formerly proposed [6]. It has been shown that the NGnet's learning based on the on-line EM algorithm is more accurate and much faster than that of a recurrent neural network trained by the gradient method [4].

In our previous studies [4, 7], we examined partial observation situations, in which delay coordinate embedding [2] was used. The NGnet was trained to approximate the dynamical system in the delay coordinate space. We showed that the trained NGnet was able to reproduce a chaotic attractor. The reproduced attractor well approximated the complexity and the instability of the true embedded chaotic attractor, even under various noise conditions.

In this article, we propose a new learning scheme, in which a new embedding method called integral embedding is used. Since this embedding is based on IIR (infinite impulse response) smoothing filters for the time-series, it is expected to be more robust to noise than the delay coordinate embedding.

In experiments, the NGnet is trained to approximate a vector field of a dynamical system in the integral coordinate space. In order to measure the performance of our new learning method, the dynamical characteristics of chaotic attractors reproduced by the trained NGnet are compared with those of the original chaotic attractor. The results show that our new method is able to well approximate the chaotic dynamics even when the noise level becomes large. Compared to our previous method [4, 7], the new method is more robust to two kinds of noise processes: system noise and observation noise. Moreover, the new learning method is faster than the previous one.

2. NGNET AND ON-LINE EM ALGORITHM

The NGnet [5], which transforms an N-dimensional input vector x to a D-dimensional output vector y, is defined by

$$y = \sum_{i=1}^{M} \left(\frac{G_i(x)}{\sum_{j=1}^{M} G_j(x)} \right) \tilde{W}_i \tilde{x} \tag{1a}$$

$$G_i(x) \equiv (2\pi)^{-N/2} |\Sigma_i|^{-1/2} \times \exp\left[-\frac{1}{2}(x-\mu_i)' \Sigma_i^{-1} (x-\mu_i) \right]. \tag{1b}$$

M denotes the number of units, and the prime($'$) denotes a transpose. $G_i(x)$ is a Gaussian function; its center is an N-dimensional vector μ_i and its covariance matrix is an N-by-N matrix Σ_i. $|\Sigma_i|$ is the determinant of Σ_i. $\tilde{W}_i \equiv (W_i, b_i)$ is a D-by-$(N+1)$ linear regression matrix, where $\tilde{x}' \equiv (x', 1)$ and $\tilde{W}_i \tilde{x} \equiv W_i x + b_i$.

The NGnet can be interpreted as a stochastic model, in which a pair of an input and an output, (x, y), is a stochastic event. For each event (x, y), a unit index $i \in \{1, \cdots, M\}$ is assumed to be selected, which is regarded as a hidden variable. The stochastic model [8] is defined by the probability

distribution for a triplet (x, y, i):

$$P(x,y,i|\theta) = \frac{G_i(x)}{M}(2\pi)^{-\frac{D}{2}}\sigma_i^{-D}\exp\left[\frac{-1}{2\sigma_i^2}(y-\tilde{W}_i\tilde{x})^2\right], \quad (2)$$

where $\theta \equiv \{\mu_i, \Sigma_i, \sigma_i^2, \tilde{W}_i | i = 1, \cdots, M\}$ is a set of model parameters. The expectation value of the output y for a given input x, $E[y|x]$, is identical to equation (1); namely, the probability distribution (2) provides a stochastic model for the NGnet.

Given a data set, the model parameter of the stochastic model (2) can be determined by the EM algorithm [3]. Here, we explain the on-line EM algorithm [6], which repeats the following E- and M- steps. Let $\{(x(t), y(t))|t = 1, \cdots\}$ be the data supplied to the learning system, and $\theta(t)$ be the estimated model parameter after the t-th datum $(x(t), y(t))$.

In an E(Expectation)-step, using the current parameter $\theta(t-1)$, the posterior probability for the t-th datum, $P_i(t) \equiv P(i|x(t), y(t), \theta(t-1))$ is calculated as

$$P_i(t) = \frac{P(x(t), y(t), i|\theta(t-1))}{\sum_{j=1}^{M} P(x(t), y(t), j|\theta(t-1))}. \quad (3)$$

In an M(Maxmization)-step, using the posterior probability (3), the weighted means, i.e., $\ll 1 \gg_i (t)$, $\ll x \gg_i (t)$, $\ll |y|^2 \gg_i (t)$, and $\ll \tilde{x}y' \gg_i (t)$, are calculated according to the following step-wise equations.

$$\ll f(x,y) \gg_i (t) = \ll f(x,y) \gg_i (t-1) + \eta(t)\left[f(x(t), y(t))P_i(t) - \ll f(x,y) \gg_i (t-1)\right] \quad (4a)$$

$$\eta(t) = (1 + \lambda(t)/\eta(t-1))^{-1}. \quad (4b)$$

$\eta(t)$ is an effective learning coefficient. After that, the new parameter $\theta(t)$ is obtained as

$$\mu_i(t) = \ll x \gg_i (t) / \ll 1 \gg_i (t) \quad (5a)$$

$$\tilde{\Lambda}_i(t) = \frac{1}{1-\eta(t)}\left[\tilde{\Lambda}_i(t-1) - \frac{P_i(t)\tilde{\lambda}_i(t-1)\tilde{x}(t)\tilde{x}'(t)\tilde{\Lambda}_i(t-1)}{(1/\eta(t)-1) + P_i(t)\tilde{x}'(t)\tilde{\Lambda}_i(t-1)\tilde{x}(t)}\right] \quad (5b)$$

$$\tilde{W}_i(t) = \tilde{W}_i(t-1) + \eta(t)P_i(t)(y(t) - \tilde{W}_i(t-1)\tilde{x}(t))\tilde{x}'(t)\tilde{\Lambda}_i(t) \quad (5c)$$

$$\sigma_i^2(t) = \frac{\ll |y|^2 \gg_i (t) - \mathrm{Tr}\left(\tilde{W}_i(t) \ll \tilde{x}y' \gg_i (t)\right)}{D \ll 1 \gg_i (t)}, \quad (5d)$$

where $\tilde{\Lambda}_i(t) \equiv [\ll \tilde{x}\tilde{x}' \gg]_i^{-1}(t)$, and $\mathrm{Tr}(\cdot)$ denotes a matrix trace. $\Sigma_i^{-1}(t)$ can be obtained from the following relation with $\tilde{\Lambda}_i(t)$:

$$\tilde{\Lambda}_i(t) \ll 1 \gg_i (t) = \begin{pmatrix} \Sigma_i^{-1}(t) & -\Sigma_i^{-1}(t)\mu_i(t) \\ -\mu_i'(t)\Sigma_i^{-1}(t) & 1 + \mu_i'(t)\Sigma_i^{-1}(t)\mu_i(t) \end{pmatrix}. \quad (6)$$

By decreasing the effective learning coefficient $\eta(t)$ over time, the on-line EM algorithm achieves a stochastic approximation method to obtain a maximum likelihood estimator [6]. We also employ dynamic unit production-deletion mechanisms in order to efficiently allocate the units in the subspace where the data occur. If a given datum is too distant from all of the present units, a new unit is produced to account for the datum. A unit that has rarely been used to account for the data is deleted.

3. INTEGRAL COORDINATE EMBEDDING

In many situations, a part of the dynamical variables is often observed. In order to reconstruct the original dynamical system from such a partial observation, delay coordinate embedding [2] has commonly been used. In our previous studies [4, 7], the NGnet is trained to approximate a vector field of the dynamical system in the delay coordinate space.

In this study, we propose a new embedding scheme. Since it is based on smoothing filters, the scheme is expected to be more robust to noise.

The smoothed data of time-series $x(t)$ is obtained by the following linear differential equations:

$$y_0(t) = x(t) \quad (7a)$$

$$\tau\dot{y}_k(t) = -y_k(t) + y_{k-1}(t) \quad (k \geq 1), \quad (7b)$$

where $\dot{y} \equiv dy/dt$ denotes the time derivative of variable y. $y_k(t)$ is obtained by smoothing y_{k-1} until t with delay time τ:

$$y_k(t) \equiv \frac{1}{\tau}\int_{-\infty}^{t} e^{-(t-s)/\tau} y_{k-1}(s)ds. \quad (8)$$

$y_k(t)$ can also be obtained from $x(t)$ by the following IIR filter:

$$y_k(t) = \int_{-\infty}^{t} G_k(t-s)x(s)ds \quad (9a)$$

$$G_0(t) = \delta(t) \quad (9b)$$

$$G_k(t) = \frac{1}{\tau(k-1)!}\left(\frac{t}{\tau}\right)^{k-1} e^{-t/\tau} \quad (k \geq 1), \quad (9c)$$

where $\delta(t)$ is Dirac's delta function. The filter function $G_k(t)$ has a peak at $t = (k-1)\tau$.

We assume that the time-series is sampled at every sampling time step Δt, namely, $\{x(n\Delta t)|n = 0, 1, ...\}$ is observed. Differential equation (7) is discretized with time step Δt:

$$y_0(n+1) = x((n+1)\Delta t) \quad (10a)$$

$$y_k(n+1) = \lambda y_k(n) + (1-\lambda)y_{k-1}(n), \quad (10b)$$

where $\lambda \equiv e^{-\Delta t/\tau}$.

Using the smoothed data calculated by equation (10), a new coordinate is defined:

$$Y(n) \equiv (y_0(n), y_1(n), \cdots, y_{K-1}(n)), \quad (11)$$

where K is the coordinate dimension. Since the k-th filter function has a peak at $(k-1)\tau$, this coordinate somehow resembles the delay coordinate with lag time τ. However, the smoothing filter functions are expected to reduce noise independently added to the data. The new coordinate (11) is called the integral coordinate, and the embedding in the integral coordinate is called integral embedding.

For the learning system to reconstruct the dynamical system, it should learn the vector field from the observed time-series $\{x(n\Delta t)|n = 0, 1, ...\}$. At a discrete-time n, the input to the learning system is $Y(n)$, and the target output is a discretized velocity of x: $(x((n+1)\Delta t) - x(n\Delta t))/\Delta t$.

After the learning system successfully approximates the discretized vector field by $F_I(Y)$, the predicted observable

variable at the next time step, $\hat{x}((n+1)\Delta t)$, is given by $\hat{x}(n\Delta t)+\Delta t F_I(\hat{Y}(n))$, where $\hat{x}(n\Delta t)$ and $\hat{Y}(n)$ are obtained in the previous time step. From $\hat{x}((n+1)\Delta t)$, $\hat{Y}(n+1)$ can be calculated. By repeating this process, a trajectory of x is estimated by the trained system.

4. EXPERIMENTS

Let us consider the Lorenz system [1], which is defined by

$$\begin{aligned} \dot{x} &= \alpha(y-x) \\ \dot{y} &= -y+(\beta-z)x \\ \dot{z} &= -\gamma z + xy. \end{aligned} \tag{12}$$

The parameters are set at the usual values: $\alpha = 10.0, \beta = 28.0$, and $\gamma = 8/3$. An orbit of the Lorenz equation (12) is obtained by the fourth-order Runge-Kutta method with the integration time step 0.001. From this orbit, the observed time-series is sampled at every $\Delta t = 0.01$. It is assumed that only the time-series of dynamical variable x, $\{x(n\Delta t)|n = 0, 1, \cdots\}$, is supplied to the learning system.

In this experiment, we set the parameters at $K = 3$ and $\tau = 10 \times \Delta t$. Figure 1 shows the phase portrait of $Y(n)$ in the integral coordinate, where a chaotic attractor can be seen. The correlation dimension and the maximum Lyapunov exponent of the embedded attractor are 2.03 and 1.03, respectively. Since these values are very close to those of the original Lorenz system, we consider that the integral embedding preserves the dynamical characteristics of the original dynamical system.

In order to mimic the embedded dynamical system, an NGnet is used as the learning system. Figure 2 shows the attractor reproduced by the NGnet after the learning of 20,000 data points. After the learning, the error (nMSE) becomes 0.1%. The nMSE is the mean squared error of the NGnet's estimation of the vector field from the true value. It is normalized by the variance of the true vector field of the embedded Lorenz attractor. Although the amount of training data is not large, the reproduced attractor in Fig. 2 is similar to that in Fig. 1.

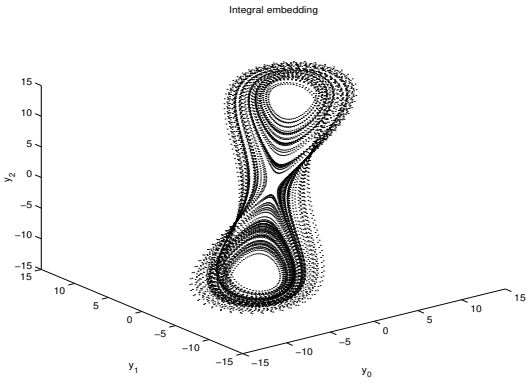

Figure 1: Lorenz attractor embedded in integral coordinate.

In real world problems, observed variables often involve noise. In the next experiment, therefore, two kinds of noise

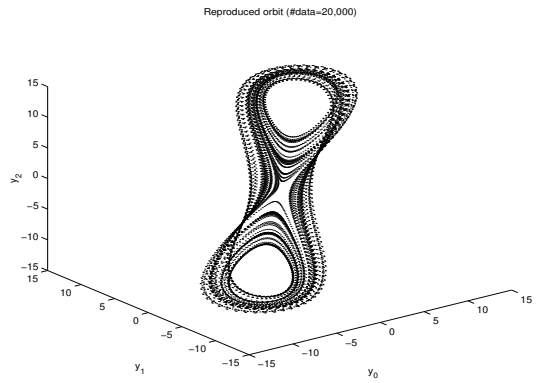

Figure 2: Chaotic attractor reproduced by NGnet.

are added to the observed data: system noise and observation noise [7]. The system noise disturbs the system dynamics itself. Namely, dynamical equation (12) is modified to involve white Gaussian noise. The system noise is assumed to be added to all of the state variables. In contrast, the observation noise does not disturb the system dynamics. It is white Gaussian noise added when the dynamical variable x is observed. The noise level is given in proportion to the variance of the state variable in the original dynamical system.

We investigate the robustness of our method to the two kinds of noise. As the noise level becomes large, the nMSE on the attractor becomes large and the amount of training data needed to approximate the dynamics increases. If a sufficiently large amount of data is supplied, our method is able to reproduce a chaotic attractor that is similar to the original embedded attractor, even when the training data involves fairly large noise.

In order to measure the similarity between the true embedded Lorentz attractor and the attractors reproduced by the trained NGnet, the dynamical characteristics of the attractors are examined. Table I summarizes the results after the learning of 1,000,000 data points under various noise conditions. In Table I, "NU" denotes the number of units after the learning, whose initial value is 300. The correlation dimension (CD) [1] is calculated from four sets, each consisting of 50,000 points. The maximum Lyapunov exponent (MLE) [1] is calculated from Jacobian matrices along a single trajectory that consists of 50,000 points. Each Jacobian matrix is estimated by a second-order polynominal approximation that uses 32 neighboring points. The Gram-Schmidt orthogonalization is also used for the MLE calculation.

As can be seen in Table I, the characteristics of the chaotic dynamics are well reproduced by our method, even when the system noise level increases up to 50% or the observation noise level increases up to 70%. When the data involve further large noise, the trained NGnet cannot reproduce chaotic attractors. For comparison, Table II summarizes the results of our previous study [4], using delay coordinate embedding after the learning of 2,000,000 data points. When embedded in the delay coordinate space, the

		NU	nMSE	CD	MLE
True				2.03±0.02	1.03±0.00
noiseless		288	0.002	2.03±0.02	0.96±0.00
SN	10%	295	0.015	2.01±0.02	1.05±0.02
	30%	283	0.084	1.93±0.03	0.89±0.02
	50%	249	0.218	1.86±0.03	0.79±0.01
ON	10%	280	0.208	2.01±0.02	1.25±0.01
	30%	290	1.103	1.98±0.02	1.82±0.03
	50%	295	2.820	1.86±0.02	1.90±0.09
	70%	284	5.720	1.85±0.03	2.60±0.09

Table I: Integral embedding

		NU	nMSE	CD	MLE
True				2.02±0.01	0.98±0.01
noiseless		387	0.003	1.92±0.06	0.88±0.00
SN	5%	292	0.006	1.97±0.02	0.97±0.00
	15%	813	0.075	1.87±0.03	1.48±0.07
ON	10%	504	0.060	1.73±0.05	4.21±0.13
	20%	1000	0.120	1.60±0.04	4.00±0.20

Table II: Delay embedding.

reproduction of chaotic attractors is successful up to 15% system noise level or up to 20% observation noise level.

By comparing Table I and Table II, the advantage of our new embedding method is obvious. The noise robustness is much larger in our new method than in our previous method. This is probably because the smoothing filters effectively reduce the system or observation noise. Moreover, the CD and MLE values in the new method are much better than those in the previous method. Consequently, the new method is able to approximate the dynamical characteristics of the original system better than the previous method. In addition, the new method is faster in learning, especially when the data includes large noise, because the necessary number of units is smaller than that in the previous method. The number of data points needed for learning is also smaller in our new method. However, even in our new method, the MLE increases as the observation noise level increases; namely, the observation noise increases the instability of the obtained dynamics.

By approximating the dynamics, it is possible to predict the future behavior of the observable state variable $x(t)$ for a certain period of time. Figure 3 shows the prediction error after the learning of 1,000,000 noiseless data points. The prediction error is the squared error between the predicted variable and the true variable and is normalized by the variance of the variable. The line shows the average over 100 trials. The reproduced orbit almost agrees with the true orbit until about $t = 2.3$, i.e., during the future 230 observation steps, on average.

In chaotic dynamical systems, the errors on average grow exponentially as $\exp(\mathrm{MLE} \cdot t)$. Compared to the characteristic time $(1/\mathrm{MLE}) \approx 0.97$ determined by the MLE, the above prediction time is long. This implies that our approach has an excellent prediction ability.

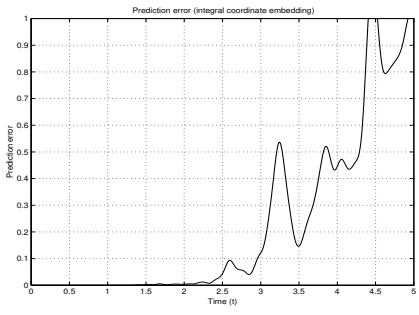

Figure 3: Prediction error.

5. CONCLUSION

In this article, we discussed the reconstruction of chaotic dynamics by using an NGnet trained by our on-line EM algorithm. In order to deal with partial observations, we proposed a new embedding method using smoothing filters, which is called integral embedding. Experimental results showed that our new method was able to well reconstruct the original dynamics.

We also discussed the robustness of our method to two kinds of noise processes, i.e., system noise and observation noise. By investigating the CD and MLE of the reproduced attractor, our new method was able to well approximate the complexity and instability of the original chaotic attractor even with the two kinds of noise.

In comparison with the delay coordinate method, our new method was more robust to noise. Moreover, the learning proceeded faster.

Our approach also showed good performance for predicting the future behavior of the observed variable.

6. REFERENCES

[1] Abarbanel, H. D. I. (1996). *Analysis of observed chaotic data*, Springer-Verlag.

[2] Casdagli, M., Eubank, S., Farmer, J. D., & Gibson, J. (1991). *Physica D*, **51**, 52-98.

[3] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). *Journal of Royal Statistical Society B*, **39**, 1-22.

[4] Ishii, S., & Sato, M. (1998). *Neural Networks for Signal Processing VIII*, (pp. 360-369), IEEE.

[5] Moody, J., & Darken, C. J. (1989). *Neural Computation*, **1**, 281-294.

[6] Sato, M., & Ishii, S. (2000). *Neural Computation*, **12**(2), in press.

[7] Yoshida, W., Ishii, S., & Sato, M. (1999). *IEEE International Conference on Systems, Man and Cybernetics*, **I**, (pp. 414-419), IEEE.

[8] Xu, L., Jordan, M. I., & Hinton, G. E. (1995). *Advances in Neural Information Processing Systems 7* (pp. 633-640), MIT Press.