RECONSTRUCTIONS AND PREDICTIONS OF NONLINEAR DYNAMICAL SYSTEMS BY RAO-BLACKWELLISED SEQUENTIAL MONTE CARLO

T. Soma, K. Yosui and T. Matsumoto

Department of Electrical, Electronics and Computer Engineering
Waseda University, 3-4-1 Ohkubo Shinjuku-ku Tokyo, Japan
Core Research for Evolutional Science and Technology(CREST)
Phone: +81 3 5286 3377
Fax: +81 3 3702 4735
E-mail: takashi@mse.waseda.ac.jp

ABSTRACT

Sequential Monte Carlo (SMC) is a powerful sampling based inference/learning algorithm for Bayesian scheme. The purpose of this paper is two fold. It first attempts to reconstruct and predict nonlinear dynamical systems from one dimensional data which arrives in a sequential manner instead of batch manner. Second purpose is to test the performance of the Rao-Blackwellisation in reconstructing and predicting nonlinear dynamical systems. We demonstrate that Rao-Blackwellised Sequential Monte Carlo (RBSMC) on a chaotic time series prediction problem outperforms generic SMC.

1. INTRODUCTION

Time series prediction problem amounts to making predictions of future values of given data set. This is one of the fundamental problems in science and engineering as well as in other disciplines. A time series prediction algorithm needs to contain a scheme for capturing mechanisms behind which generate given time series data. In other words, it is required to reconstruct the dynamical systems behind the data. There are at least three fundamental issues to be addressed: The dynamics behind the data is unknown. The data contains uncertainties. Observation is often 1-dimensional, while the order of the dynamics is higher. One way of approaching this problem is to prepare model dynamical systems and fit it to given data. Many of the contemporary learning schemes formulate the problem within the probability/statistics frameworks and the Bayesian schemes often work well for such purposes.

This paper proposes an on-line Bayesian scheme [1] with Rao-Blackwellisation [2, 3] and test the scheme against a chaotic time series data. The proposed algorithm appears to be reasonably sound. It outperforms the conventional Sequential Monte Carlo.

2. FORMULATION

Problem : Given time series data $\{x_j\}_{j=0}^{t-1}$, predict $\{x_j\}_{j=t}^{t+k}$, $k = 0, 1, 2, \cdots$.
Architecture : Specification of a family of functions for data fitting, *e. g.*, three layer perceptron with h hidden units.
Model dynamical System :

$$P(x_t \mid \{x_j\}_{j=t-\tau}^{t-1}; \boldsymbol{w}_t) := \frac{1}{Z(\beta_t)} \exp[-\frac{\beta_t (x_t - f(\{x_j\}_{j=t-\tau}^{t-1}; \boldsymbol{w}_t))^2}{2}] \quad (1)$$

where $f(\cdot; \boldsymbol{w}_t)$ denotes the output of perceptron, τ denotes the order of the model dynamical system. β_t stands for the uncertainty level, $\boldsymbol{w}_t$ denotes the perceptron parameters and $Z(\beta_t)$ is normalizing constant.
Weight Update Equation :

$$P(\boldsymbol{w}_t \mid \boldsymbol{w}_{t-1}, \gamma_t) := \frac{1}{Z_{w_t}(\gamma_t)} \exp[-\frac{\gamma_t ||\boldsymbol{w}_t - \boldsymbol{w}_{t-1}||^2}{2}] \quad (2)$$

where $1/\gamma_t$ is the variance and $Z(\gamma_t)$ is normalizing constant.

Remark: The weight update equation often suffers from "over dispersion" when γ fixed. We propose a learning sheme for hyperparameter γ which prevents over dispersion. We also propose a learning scheme for hyperparameter β , the uncertainty level of the dynamics.

Hyperparameter Update Equations :

$$P(\gamma_t \mid \gamma_{t-1}) := \frac{1}{\sqrt{2\pi}\sigma_\gamma \gamma_t} \exp\left[-\frac{(\log \gamma_t - \log \gamma_{t-1})^2}{2\sigma_\gamma^2}\right] \quad (3)$$

$$P(\beta_t \mid \beta_{t-1}) := \frac{1}{\sqrt{2\pi}\sigma_\beta \beta_t} \exp\left[-\frac{(\log \beta_t - \log \beta_{t-1})^2}{2\sigma_\beta^2} \right] \quad (4)$$

where $\sigma_\gamma > 0$ and $\sigma_\beta > 0$ are assumed.

Posterior Distribution : The joint posterior follows from the Bayes formula:

$$\begin{aligned} &P\left(\boldsymbol{w}_t, \beta_t, \gamma_t \mid \{x_j\}_{j=0}^{t-1}\right) \\ &= \int P\left(\boldsymbol{w}_t \mid \boldsymbol{w}_{t-1}, \gamma_t\right) P\left(\beta_t \mid \beta_{t-1}\right) P\left(\gamma_t \mid \gamma_{t-1}\right) \\ &\times P\left(\boldsymbol{w}_{t-1}, \beta_{t-1}, \gamma_{t-1} \mid \{x_j\}_{j=0}^{t-1}\right) d\boldsymbol{w}_{t-1} d\beta_{t-1} d\gamma_{t-1} \end{aligned} \quad (5)$$

$$\begin{aligned} &P\left(\boldsymbol{w}_t, \beta_t, \gamma_t \mid \{x_j\}_{j=0}^{t}, \right) \\ &= \frac{P\left(x_t \mid \{x_j\}_{j=t-\tau}^{t-1}, \boldsymbol{w}_t, \beta_t\right) P\left(\boldsymbol{w}_t, \beta_t, \gamma_t \mid \{x_j\}_{j=0}^{t-1}\right)}{P\left(x_t \mid \{x_j\}_{j=0}^{t-1}\right)} \end{aligned} \quad (6)$$

Predictive Distribution : The predictive distribution is given by:

$$\begin{aligned} &P(x_t \mid \{x_j\}_{j=0}^{t-1}) \\ &= \int P\left(x_t \mid \{x_j\}_{j=t-\tau}^{t-1}, \boldsymbol{w}_t, \beta_t\right) \\ &\times P\left(\boldsymbol{w}_t, \beta_t, \gamma_t \mid \{x_j\}_{j=0}^{t-1}\right) d\boldsymbol{w}_t d\beta_t d\gamma_t \end{aligned} \quad (7)$$

3. SEQUENTIAL MONTE CARLO

Typically, when the model has non-linearity, (5),(6) and (7) cannot be computed analytically since they require complicated high-dimensional integrals.

3.1. Bayesian Importance Sampling

Let $\theta := (\boldsymbol{w}, \beta, \gamma)$, if one has a proposal distribution $Q(\theta)$ and if the importance weight Ω is well-defined,

$$\Omega = \frac{P(D \mid \theta) P(\theta)}{Q(\theta)}$$

the posterior distribution $P(\theta \mid D)$ is described with Ω and $Q(\theta)$:

$$P(\theta \mid D) = \frac{\Omega Q(\theta)}{\int \Omega Q(\theta) d\theta}$$

Given i.i.d. samples $\theta^{(i)} \sim Q(\theta)$, a Monte Carlo approximation of $P(\theta \mid D)$ is given by:

$$P(\theta \mid D) \approx \tilde{\Omega}^{(i)} \delta_i(\theta)$$

where $\tilde{\Omega}$ is the normalized importance weight:

$$\tilde{\Omega}^{(i)} = \frac{\Omega^{(i)}}{\sum_{j=1}^{N} \Omega^{(j)}}$$

3.2. Resampling

Samples with low importance weights are eliminated, samples with high importance weights are multiplied, so that un-weighted samples from the posterior distribution $P(\theta \mid D)$ are obtained.

4. RAO-BLACKWELLIZATION

Since some of the weight parameters of perceptron are linear, marginalizations can be performed without approximations via Kalman Filter or any other linear optimal filters. This leads to dimension reduction in the parameter space which, in turn, leads to prediction accuracy improvements as well as computation time reduction. This scheme is known as Rao-Blackwellization.

4.1. Rao-Blackwellised Sequential Monte Carlo

The posterior distribution of parameter $\boldsymbol{w}_t$ can be decomposed as follows:

$$P\left(\boldsymbol{w}_t \mid \{x_j\}_{j=0}^{t}\right) \quad (8)$$

$$\begin{aligned} &= P(\boldsymbol{w}_t^{(1)}, \boldsymbol{w}_t^{(2)} \mid \{x_j\}_{j=0}^{t}) \\ &= P(\boldsymbol{w}_t^{(2)} \mid \{x_j\}_{j=0}^{t}, \boldsymbol{w}_t^{(1)}) \times P(\boldsymbol{w}_t^{(1)} \mid \{x_j\}_{j=0}^{t}) \end{aligned} \quad (9)$$

where $\boldsymbol{w}_t^{(1)}$ is the parameter between input layer and hidden layer, $\boldsymbol{w}_t^{(2)}$ is the parameter between hidden layer and output. Observe the first term is given by:

$$\begin{aligned} &P(\boldsymbol{w}_t^{(2)} \mid \{x_j\}_{j=0}^{t}, \boldsymbol{w}_t^{(1)}) \\ &= \frac{P(x_t \mid \{x_j\}_{t-\tau}^{t-1}, \boldsymbol{w}_t^{(1)}, \boldsymbol{w}_t^{(2)}) P(\boldsymbol{w}_t^{(2)} \mid \{x_j\}_{j=0}^{t-1}, \boldsymbol{w}_t^{(1)})}{P(x_{t+1} \mid \{x_j\}_{j=0}^{t-1}, \boldsymbol{w}_t^{(1)})} \end{aligned} \quad (10)$$

while the second term is given by:

$$\begin{aligned} &P(\boldsymbol{w}_t^{(1)} \mid \{x_j\}_{j=0}^{t}) \\ &= \frac{P(x_t \mid \{x_j\}_{j=t-\tau}^{t-1}, \boldsymbol{w}_t^{(1)}) P(\boldsymbol{w}_t^{(1)} \mid \{x_j\}_{j=0}^{t-1})}{P(x_t \mid \{x_j\}_{j=0}^{t-1})} \end{aligned} \quad (11)$$

Given $\boldsymbol{w}_t^{(1)}$, the conditional posterior distribution $P(\boldsymbol{w}_t^{(2)} \mid \{x_j\}_{j=0}^{t}, \boldsymbol{w}_t^{(1)})$ and the marginal likelihood $P(x_t \mid \{x_j\}_{j=0}^{t-1})$ in (10), can be computed by the Kalman Filter, since $\boldsymbol{w}_t^{(2)}$ appears linearly . Then, given the marginal likelihood $P(x_t \mid \{x_j\}_{j=0}^{t-1})$, samples can be drawn from (11) with the Bayesian Importance Sampling and Resampling.

5. DEMONSTRATION

5.1. Chaotic time series prediction

Equation (12) is the well-known Rössler system where $(a, b, c) = (0.36, 0.4, 4.5)$.

$$\left\{ \begin{array}{lcl} \dot{x} & = & -y - z \\ \dot{y} & = & x + ay \\ \dot{z} & = & bx - cz + xz \end{array} \right. \qquad (12)$$

We consider a discretized noisy Rössler system:

$$\left\{ \begin{array}{lcl} x_{(t+1)} & = & f(x_t, y_t, z_t) + \nu_t^1 \\ y_{(t+1)} & = & g(x_t, y_t, z_t) + \nu_t^2 \\ z_{(t+1)} & = & h(x_t, y_t, z_t) + \nu_t^3 \end{array} \right. \qquad (13)$$

where f, g and h are Runge-Kutta discretization of (12), and ν_t^i are noise processes: $\nu_t^i \sim i.i.d.\ N(0, \sigma^2),\ i = 1, 2, 3$.

In this demonstration, the data is generated according to (13) with σ = 0.01, and only the x-coordinate of (13) can be observed (Fig.1), while the y- and z-coordinates are not available for learning. The difficulty of this problem is worth noting; in addition to the complex chaotic behavior, noise processes exist, and only single quantity x_t is available.

5.1.1. *Delay Coordinate Embedding*

Even if x_t is one dimensional quantity instead of vector quantity, there is a way of handling the order of the dynamics behind the one dimensional data. This is known as Delay Coordinate Embedding [4].

In this demonstration, $\tau = 4$ is assumed which has been estimated in our previous work [5].

5.2. Prediction of dynamical systems

Given learned posterior distribution $P(\boldsymbol{w}_T, \beta_T, \gamma_T \mid \{x_j\}_{j=0}^{T})$, the prediction of $\{x'_j\}_{j=0}^{t}$ is given by the following:

$$\begin{aligned} & if\ \ t = 0, 1, 2, 3 \\ & \quad \hat{x'}_t = x'_t \\ & if\ \ t \geq 4 \\ & \quad \hat{x'}_t = E[P(x'_t \mid \{x'_j\}_{j=0}^{t-1}] \\ & \qquad = \int x'_t P(x'_t \mid \{x'_j\}_{j=t-\tau}^{t-1}, \boldsymbol{w}_t, \beta_t) \\ & \qquad\quad \times P(\boldsymbol{w}_T, \beta_T, \gamma_T \mid \{x_j\}_{j=0}^{T}) d\boldsymbol{w}_T d\beta_T dx_T \end{aligned}$$

Fig.3 and Fig.4 show predicted x-coordinate trajectory $\{\hat{x'_j}\}_{j=0}^{500}$ by generic SMC (Fig.3) and RBSMC (Fig.4), where $T = 1000$.

Fig.5 shows evolution of cumulative prediction errors:

$$\sum_{k=0}^{t} (x'_j - \hat{x'}_j)^2$$

RBSMC results appears to reconstruct the Rössler system better than SMC. The errors of RBSMC is lower than that of SMC.

6. REFERENCES

[1] A. Doucet, J.F.G.de Freitas, and N.J. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2000.

[2] K. Yosui, T. Kurihara, K. Wada, T. Souma, and T. Matsumoto, "Bayesian on-line learning : A sequential monte carlo with rao-blackwellisation," in *Proc. 12th IEEE Workshop on Neural Networks for Signal Processing*, Martigny, Switzerland, 2002, pp. 99–108.

[3] G. Casella and C.P.Robert, "Rao-blackwellisation of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.

[4] T. Sauer, J. Yorke, and M. Casdagli, "Embeddology," *J. Stat. Phys.*, vol. 65, pp. 579–616, 1991.

[5] T. Matsumoto, Y. Nakajima, M. Saito, J. Sugi, and H. Hamagishi, "Reconstructions and predictions of nonlinear dynamical systems: A hierarchical bayesian approach," *IEEE Trans. On Signal Processing*, vol. 49, no. 9, pp. 2138–2155, 2001.

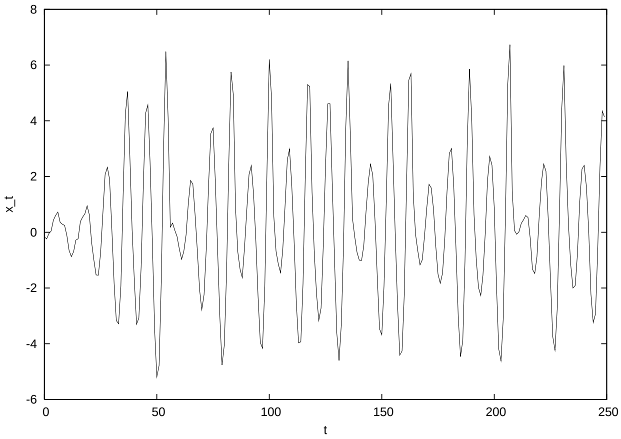

Fig. 1. Training data

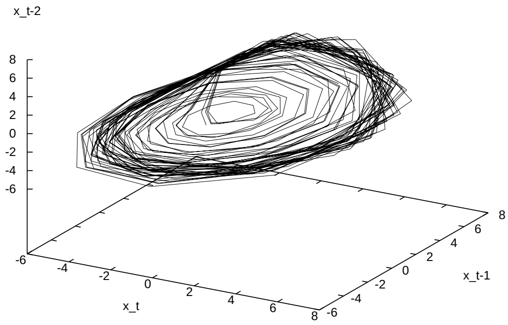

Fig. 4. Prediction on RBSMC

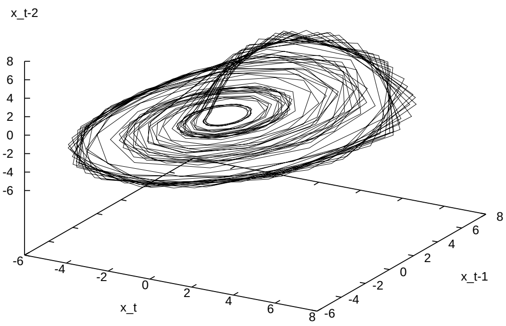

Fig. 2. x-coordinate of true Rössler system

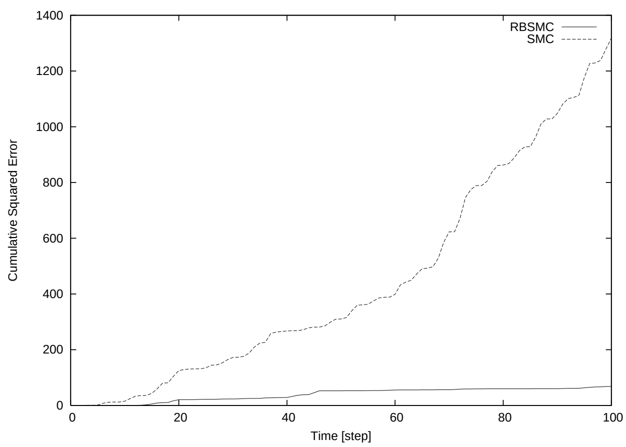

Fig. 5. cumulative error

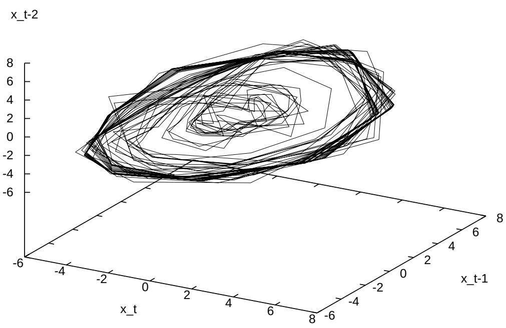

Fig. 3. Prediction on SMC