

中田 洋平, 佐藤 嘉寛, 松本 隆

† 早稲田大学理工学部電気電子情報工学科, 科学技術振興事業団

## 1 まえがき

背後に非線形性が潜む時系列データ  $\{x_t\}_{t=0}^N \subset \mathbb{R}$  が与えられた時, 背後にある非線形ダイナミカルシステムを捕らえるには, なんらかの『非線形モデル』, あるいは『基底関数』が必要である. また, 一般にデータはノイズを含み, 何らかの確率・統計的枠組みが必要となる.

このような観点から筆者らは『階層ベイズ的』ニューラルネットワーク [1][2] を用いた非線形時系列予測手法を提案しており, 具体的問題を含めた時系列予測問題に比較的成功的に適用している [1][2]. 提案手法では“予測分布”が重要な役割を演じている. “予測分布”とは“予測対象となる時系列(予測値)の確率分布”を表し, 一般に予測とは『予測分布の評価』が最終的な目的となる. 小文では, 『予測分布の評価』を

(i) Markov Chain Monte Carlo を用いてパラメータ, ハイパーパラメータの結合事後分布よりサンプルをとる;

(ii) パラメータ, ハイパーパラメータのサンプルを用いて, 時系列の将来値(予測値)のサンプルをとる;

事によりおこなう手法を提案する. サンプルという形で『予測分布の評価』を実現することで, 予測分布の平均やエラーバー(標準偏差)などの時系列予測に有用な統計的な値を求めることが可能である.

## 2 定式化

問題: 時系列データ  $D := \{x_t\}_{t=0}^N \subset \mathbb{R}$  が与えられたとき,  $\{x_t\}_{t=N+1}^T$  を予測せよ.

Hypothesis  $\mathcal{H}$ :

- アーキテクチャ: 用いるモデルの構造を指定する. ここでは中間素子数  $h$ , 中間素子の出力関数  $1/(1+e^{-u})$  の 3 層パーセプトロンを用いる.

2. 尤度 (likelihood):

$$P(\{x_t\}_{t=\tau}^N, \{x_{\tau-1}, \dots, x_0\} | \mathbf{w}, \beta, \mathcal{H}) \\ := \underbrace{\prod_{t=0}^{N-\tau} \frac{1}{Z_D(\beta)} \exp(-\beta E_D(x_{t+\tau} | x_{t+\tau-1}; \mathbf{w}))}_{\text{noisy dynamics}} \\ \times \underbrace{P(x_{\tau-1}, \dots, x_0 | \mathcal{H})}_{\text{initial state uncertainty}} \quad (1)$$

$$E_D(x_{t+\tau} | x_{t+\tau-1}; \mathbf{w}) := \frac{1}{2}(x_{t+\tau} - f(x_{t+\tau-1}; \mathbf{w}))^2 \quad (2)$$

$$\mathbf{x}_{t+\tau-1} := [x_{t+\tau-1}, \dots, x_{t+1}]^T \quad (3)$$

但し  $f(\cdot)$  は weight パラメータ  $\mathbf{w} \in \mathbb{R}^k$  をもつパーセプトロン出力,  $\beta$  は不確実性レベル(未知)を表すもう 1 つのハイパーパラメータ,  $Z_D(\beta)$  は正規化定数, そして  $\tau$  はダイナミカルシステムの次数(未知)である.

3.  $\mathbf{w}$  の prior:

$$P(\mathbf{w} | \alpha, \mathcal{H}) := \prod_{c=1}^C \frac{1}{Z_W(\alpha_c)} \exp(-\alpha_c E_{W_c}(\mathbf{w}_c)) \quad (4)$$

$$E_{W_c}(\mathbf{w}_c) := \frac{1}{2} \|\mathbf{w}_c\|^2 \quad (5)$$

$\mathbf{w}$  は部分ベクトルに分割され,

$$\mathbf{w} := (\mathbf{w}_1, \dots, \mathbf{w}_C), \mathbf{w}_c \in \mathbb{R}^{k_c} \quad (6)$$

対応する  $\alpha$  はベクトル

$$\alpha := (\alpha_1, \dots, \alpha_C), \alpha_c \in \mathbb{R} \quad (7)$$

である. 各  $\exp(-\alpha_c E_{W_c}(\mathbf{w}_c))/Z_W(\alpha_c)$  は, 各  $\mathbf{w}_c$  がいかに分布しているかの先験情報である.

4.  $(\alpha, \beta)$  の prior:  $P(\alpha, \beta | \mathcal{H})$

5.  $\mathcal{H}$  の prior:  $P(\mathcal{H})$

ベイズ公式からパラメータとハイパーパラメータの結合事後分布は,

\* Bayesian prediction for nonlinear dynamical systems: predictive mean and error bar computation via MCMC

† Department of Electrical, Electronics and Computer Engineering, Waseda University

$$P(\mathbf{w}, \alpha, \beta | D, \mathcal{H}) = \frac{P(D | \mathbf{w}, \beta, \mathcal{H})P(\mathbf{w} | \alpha, \mathcal{H})P(\alpha, \beta | \mathcal{H})}{P(D | \mathcal{H})} \quad (8)$$

となる。従ってある特定のモデル  $\mathcal{H}$  における予測分布は次で与えられる：

$$P(\{x_t\}_{t=N+1}^T | D, \mathcal{H}) = \iint P(\{x_t\}_{t=N+1}^T | x_N, \mathbf{w}, \beta, \mathcal{H}) \times P(\mathbf{w}, \alpha, \beta | D, \mathcal{H}) d\mathbf{w} d\alpha d\beta \quad (9)$$

### 3 学習

小文で用いる Markov Chain Monte Carlo によるニューラルネットの学習 [3] では

- (A) Gibbs sampling によるのハイパーパラメータの更新；
  - (B) Hybrid Monte Carlo によるのパラメータの更新；
- を交互に繰り返す。ここで  $j$  回目の操作により得た値を各々  $\mathbf{w}^{(j)}$ ,  $\alpha^{(j)}$ ,  $\beta^{(j)}$  とする。

### 4 予測

小文での目的は予測分布から時系列の将来値  $\{x_t\}_{t=N+1}^T$  のサンプルをとることである。即ち、

$$\{x_t^{(l)}\}_{t=N+1}^T \sim P(\{x_t\}_{t=N+1}^T | D, \mathcal{H}) \quad (10)$$

を行うことである。式 (10) の作業を実現するためには以下の二つの手順を順番に行えばよい。

$$\begin{aligned} (\mathbf{w}^{(l)}, \alpha^{(l)}, \beta^{(l)}) &\sim P(\mathbf{w}, \alpha, \beta | D, \mathcal{H}) \quad (11) \\ \{x_t^{(l)}\}_{t=N+1}^T &\sim P(\{x_t\}_{t=N+1}^T | \mathbf{w}^{(l)}, \beta^{(l)}, \mathcal{H}) \quad (12) \end{aligned}$$

この手法は合成法と呼ばれることもある。ここで  $(\mathbf{w}^{(l)}, \alpha^{(l)}, \beta^{(l)})$  はそれぞれ  $l$  番目のサンプル、 $x_t^{(l)}$  は時刻  $t$  における予測値の  $l$  番目のサンプルをあらわす。(11) は前節で示した学習に対応し、Markov Chain Monte Carlo によってパラメータ  $\mathbf{w}$ , ハイパーパラメータ  $(\alpha, \beta)$  の結合事後分布のサンプルを得る。時系列モデルのマルコフ性 (式 (1)) により (12) のサンプルを得ることができる。

### 5 具体例：カオスの時系列予測

提案手法によりカオスのダイナミカルシステムの代表的なものの一つである Rössler 系に雑音を加わった系への適用を試みる (図 2)。図 4 は中間素子数  $h = 4$  の

モデルによる Rössler 系の予測波形 (平均およびエラーバー) のプロットである。図 4 は中間素子数  $h = 4 \sim 8$  のモデルによる累積二乗誤差の時間変化である。また、比較対照として QAP 法 [1][2] による予測の累積二乗誤差のプロットも描かれている。

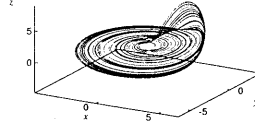


図 1: Rössler 系  
( $a, b, c$ ) = (0.36, 0.4, 4.5)

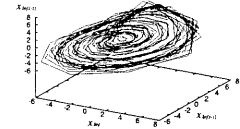


図 2: 学習データ  
(雑音を含む  
Rössler 系)

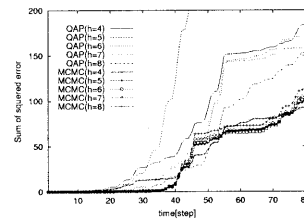


図 3: 累積二乗誤差の時間変化

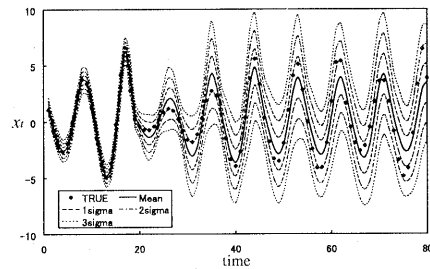


図 4: 予測とエラーバー

### 参考文献

- [1] 松本 隆, 浜岸広明, 杉淳二郎, 斉藤幹貴, 長南吉正, “データから非線形ダイナミクスへ-階層ベイズの時系列予測-,” 信学論 D-II Vol.J82-D-II, No.6, pp.1059-1071, 1999
- [2] 中島芳徳, 斉藤幹貴, 杉淳二郎, 浜岸広明, 松本 隆, “階層ベイズの空調機熱負荷予測- ニューラルネットによる非線形ダイナミカルシステムのアプローチ-,” 信学論 D-II Vol.J82-D-II, No.11, pp.1119-1124, 1999.
- [3] R. M. Neal, “Bayesian Learning for Neural Networks,” Springer-Verlag, New York, 1996.