ブートストラップ法を用いた少数データに対する局所線形近似予測法

上野 佑輔 † 鈴木 智也 ††
† 同志社大学工学部 †† 同志社大学理工学部

1 はじめに

決定論性を有する時系列データを予測する際，ローレンツ類推法などのように，局所的な軌道の流れを線形近似することで未来の振舞いを推定することが可能である [1]．しかし，予測に使用できる過去のデータ (学習データ) が少ない場合，局所線形近似を行うためのサンプルを十分に得られないために近似精度が低下し，その結果，予測精度は悪化する．そこで先攻研究 [2] は，局所線形近似予測法の 1 種であるヤコビ行列推定法にブートストラップ法を融合することで，線形近似に用いるサンプルを擬似的に複製し，近似性能を向上させることに成功している．

しかし，ヤコビ行列推定法では複数の近傍点の推移を元に，状態空間内の各座標に対する偏微分係数 (ヤコビ行列) を推定するので，用いる近傍点数を状態空間の次元以上に設定しなければいけない．さらにブートストラップ法を適用した際に，復元抽出が行列のランク落ちの原因となり，ヤコビ行列を上手く推定できない問題がある [2]．また，線形近似によってある程度の観測ノイズを緩和できるが，ヤコビ行列推定法においては最近傍点に対するノイズは緩和できない．

このようにヤコビ行列推定法には様々な制約があるため，本研究では局所線形近似予測法として最もシンプルなローレンツ類推法にブートストラップ法を融合させ，その有用性について計算機実験を通じて検証を行う．

2 局所線形近似予測法

非線形予測を行うには，まず観測時系列データ $x(t)$ に対して埋め込み定理 [1] を適用し，システムのふるまいとして多次元状態ベクトル $X(t)$ を再構成する必要がある．

$$X(t) = \{x(t), x(t-\tau), \cdots, x(t-\tau(d+1)+1\} \quad (1)$$

ここで，τ は遅れ時間，d は埋め込み次元を表す．これらのパラメータを適切に設定しないとシステムの再現性が低下し，予測精度の低下やシステムの分析誤りを引き起こす．従来においては，様々な推定手法が提案されている [1]．

次に，式 (1) によって再現された軌道を用いて非線形予測を行う．本研究では，非線形予測手法として局所線形近似法に着目する．この手法では，$X(t)$ の将来変動を予測するために，$X(t)$ から半径 ϵ の領域内より近傍点 $X(t_k)$ $(t_k < t)$ を複数探索し，その推移を平均化することで，$X(t+p)$ の予測値 $\tilde{X}(t+p)$ を得る．

$$\tilde{X}(t+p) = \frac{1}{K_t}\sum_{k=1}^{K_t} X(t_k+p) \quad (2)$$

ここで，p は予測ステップ数，K_t は近傍点の総数である．

さらに類似の手法として，平均化する際に各近傍点の距離 d_k に応じて重みを考慮する方法もある [1]．

$$\begin{aligned}\tilde{X}(t+p) &= \frac{\sum_{k=1}^{K_t}\exp(-d_k)X(t_k+p)}{\sum_{k=1}^{K_t}\exp(-d_k)} \quad (3)\\ d_k &= |X(t)-X(t_k)|\end{aligned}$$

本研究では，式 (2) を “単純平均法”，式 (3) を “加重平均法” と呼ぶ．

3 ブートストラップ法

ブートストラップ法 [3] とは，標本集団からリサンプリングを繰り返すことにより母集団の性質を推定する手法である．特に，標本集団が少ないほど母集団は不明確になるため，ブートストラップ法によって標本集団を擬似的に複製することで，データの不足を補うことができる．

局所線形近似法においては，予測に用いる学習データが短いほど，半径 ϵ 内の近傍領域に含まれる近傍点の個数が少なくなり，式 (2) や式 (3) による線形近似の精度が低下し，予測誤差が拡大する．そこで，選ばれた近傍点 $X(t_k)$ $(k = 1 \sim K_t)$ に対してブートストラップ法を適用し，真の将来変動に対する線形近似の精度を向上させる．まず，近傍点 $X(t_k)$ の中からランダムに K_t 個復元抽出する．抽出された近傍点 $X_b(t_k)$ $(k = 1 \sim K_t)$ に対して式 (2) や式 (3) を適用することで，1 つの予測値 (ブートストップ標本)$\tilde{X}_b(t+p)$ を得る．これを B 回繰り返し，得られた予測値を平均することで最終的な予測値 (ブートストップ推定値) を得る．

$$\tilde{X}(t+p) = \frac{1}{B}\sum_{b=1}^{B}\tilde{X}_b(t+p) \quad (4)$$

次章では，単純平均法および加重平均法にブートストラップ法を融合させた場合の効果を検証する．

The Bootstrap Method Applied to the Local Linear Prediction Method

† Yusuke Ueno (btf7083@mail4.doshisha.ac.jp)
†† Tomoya Suzuki (tsuzuki@mail.doshisha.ac.jp)
Department of Information System Design, Faculty of Engineering, Doshisha University (†)
Department of Information System Design, Faculty of Science and Engineering, Doshisha University (††)
1-3 Miyakodani, Tatara, Kyotanabe, Kyoto 610-0321, Japan

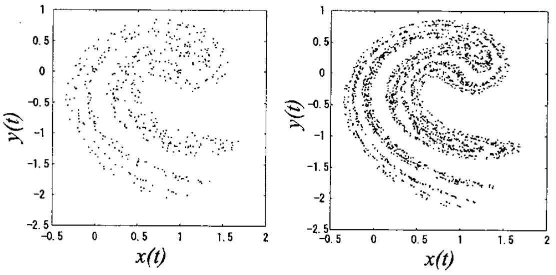

図 1: 池田写像. (左図)$N = 128$, (右図)$N = 4096$.

4 各予測法の比較実験

予測対象データとして，カオス性を有する池田写像 [1] を用いた．

$$\begin{cases} x(t+1) &= a + b\,(x(t)\cos(\theta(t)) - y(t)\sin(\theta(t))) \\ y(t+1) &= b\,(x(t)\sin(\theta(t)) + y(t)\cos(\theta(t))) \end{cases}$$

ただし，$\theta(t) = \kappa - \alpha/(1 + x^2(t) + y^2(t))$，$a = 1, b = 0.9, \kappa = 0.4, \alpha = 6.0$ である．実験用データとして，$x(t)$ および $y(t)$ $(t = 1 \sim N)$ を作成し，前半部を学習データとして軌道 $X(t)$ $(t = 1 \sim N/2)$ を構成し，後半部を予測した．さらに予測においては，池田写像の全変数 x, y を用いるため，式 (1) を $X(t) = \{x(t), y(t)\}$ とした (図 1)．さらに予測精度 E として，次式により正規化二乗誤差を算出した．

$$E = \frac{\sqrt{\sum_{t=1}^{L}[(x(t) - \tilde{x}(t))^2 + (y(t) - \tilde{y}(t))^2]}}{\sqrt{\sum_{t=1}^{L}[(x(t) - \bar{x}(t))^2 + (y(t) - \bar{y}(t))^2]}}$$

ここで L は予測したデータ長である．本研究ではデータの後半部を予測するため $L = N/2$ である．近傍領域 ϵ は，写像の直径に対して近傍率 l[%] の範囲とし，ブートストラップ標本数 $B = 200$，予測ステップ数 $p = 1$ とした．

さらに，池田写像に対して観測ノイズ ξ[dB] を付加した．

$$\xi = 10\log_{10}\frac{\sigma_o^2}{\sigma_n^2}$$

ここで，σ_o^2 は元データの分散値，σ_n^2 は元データに付加する観測ノイズの分散値である．観測ノイズは正規乱数によって生成した．

データ長 N および観測ノイズ量 ξ を変更して予測した結果を，図 2–4 に示す．各状況毎に，(1) 単純平均法, (2) 加重平均法, (3) 単純平均法とブートストラップ法を融合した方法, (4) 加重平均法とブートストラップ法を融合した方法の 4 手法による比較を行った．得られた知見は以下の通りである．

- 単純平均法+ブートストラップ法は，単純平均法とほぼ等しい．
- 加重平均法+ブートストラップ法は，他手法と比較して概ね予測精度が良い．

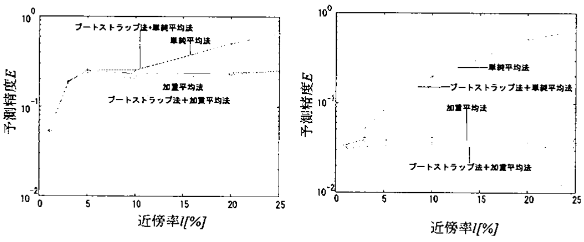

図 2: $\xi = \infty$(ノイズなし) の予測結果. (左図)$N = 128$, (右図)$N = 4096$ の場合.

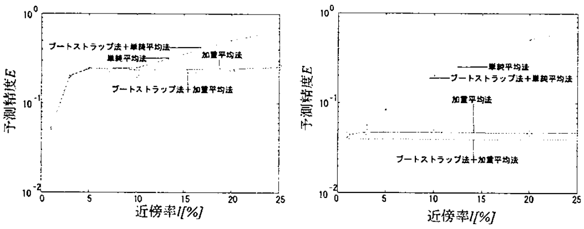

図 3: $\xi = 40$[dB] の予測結果. (左図)$N = 128$, (右図)$N = 4096$ の場合.

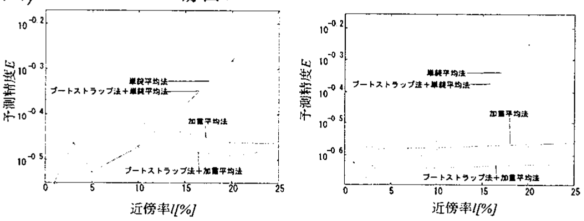

図 4: $\xi = 20$[dB] の予測結果. (左図)$N = 128$, (右図)$N = 4096$ の場合.

- 単純平均法 (+ブートストラップ法) は，近傍率 l[%] の設定を最適化できれば有用であるが，設定ミスに対する悪影響が非常に大きい．
- 加重平均法 (+ブートストラップ法) は，近傍率 l[%] の設定ミスに対して影響が少ない点でも優れている．

5 まとめ

本研究では，局所線形近似予測法 [1] において，近傍情報として使用できる標本が少ない場合に，ブートストラップ法 [3] を応用することで，擬似的に標本を複製し，将来変動の近似性能を向上させる方法を検討した．さらに計算機実験を通じて，本手法の有用性を確認した．

なお本研究の一部は，日本学術振興会科学補助金若手研究 (B)(No.20700217) の援助により行われました．

参考文献

[1] 池口徹, 山田泰司, 小室元政, “カオス時系列解析の基礎と応用,” 合原一幸 編, 産業図書, 2000.

[2] D. Haraki, T. Suzuki, H. Hashiguchi, T. Ikeguchi, ”Bootstrap Nonlinear Prediction,”Physical Review E, vol.76, No.4, 046202, 2007.

[3] B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.