ガウス過程状態空間モデルによる非線形な準周期的現象の予測

玉森　聡　　モデリング研究系　特任研究員

2015年6月19日　統計数理研究所　オープンハウス

研究背景

身の回りの様々な周期的現象
- 捕食ー非捕食関係にある生物の個体数
- 太陽活動（ex.黒点数）や大気海洋相互作用（ex.エルニーニョ現象）
- 生体リズム（睡眠，心拍，脳波，基礎体温など）

⇒非線形な変動かつ周期に揺らぎ（準周期性）

非線形な準周期的現象の予測により得られるメリット
- 生体リズムを考慮した投薬計画や医薬品開発
- 経済指標の予測精度向上に伴う最適な設備投資額の決定
- 基礎体温や月経周期の予測による女性のQOL向上　など

周期変動データのための時系列解析の手法
- SARIMAモデル，季節調整モデルなどの状態空間モデル
- 問題点: 自然現象のモデリングには不適（周期固定など）＆
  周期自体の定量的予測はあまり顧みられてこなかった

ガウス過程状態空間モデルによる準周期的現象のモデル化
- 位相を潜在変数として導入 ⇒ 位相のゆらぎを表現
- 未知の非線形関数をノンパラメトリックに自動選択
- 逐次ベイズフィルタに基づいた位相の逐次予測手法を導出
  ⇒周期の定量的予測手法を実現

準周期的現象を高精度に予測可能な手法の実現

状態空間モデル

$$\boldsymbol{x}_t \sim p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) \quad \text{（システムモデル）}$$
$$\boldsymbol{y}_t \sim p(\boldsymbol{y}_t \mid \boldsymbol{x}_t) \quad \text{（観測モデル）}$$

$\boldsymbol{x}_t$:状態変数
$\boldsymbol{y}_t$:観測変数

- マルコフ性を仮定⇒条件付き分布を簡略化
- 漸化式による効率的な状態推定（一期先予測，フィルタ，平滑化）
- 観測変数が二項分布やポアソン分布に従う場合もモデル化可能

ガウス過程

確率変数の有限個の集まりが結合ガウス分布に従う性質

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$$

平均関数　$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})]$

共分散関数 $k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}(f(\boldsymbol{x}), f(\boldsymbol{x}'))$

⇒平均関数と共分散関数による特徴付け
- 共分散関数により複雑な相関関係を扱うことが可能（周期性など）
- 具体的な関数形をデータ駆動でノンパラメトリックに決定

非線形な準周期的現象のモデル化

モデリングの対象 : 女性の基礎体温と月経周期

第n日目の位相（潜在変数）θ_n
- $0 \le \theta_n \le 1$
- 第n日目が月経開始日のとき $\theta_n = 1$
  その後 $\theta_n = 0$ にリセット

第n日目の体温（観測変数）y_n

状態空間モデルによる定式化

$$\theta_n = \theta_{n-1} + \epsilon_n \mod 1$$
$$y_n = g(\theta_n) + \sigma_n$$

$\epsilon_n \sim \mathrm{ga}(\alpha, \beta)$:ガンマ分布
$\sigma_n \sim \mathcal{N}(0, \sigma^2)$:ガウス分布

$$g(\theta_n) = a_0 + \sum_{m=1}^{M} [a_m \cos(2m\pi\theta_n) + b_m \sin(2m\pi\theta_n)] \quad \text{（次数 } M \text{ の級数近似）}$$

逐次予測手法（右上へ続く）
- 第n日目から第k日目（$n \le k$）までに加えられる位相の増分

$$\Delta(k \mid n) \stackrel{\mathrm{def}}{=} \epsilon_n + \epsilon_{n+1} + \ldots + \epsilon_k$$

$\Delta(k \mid n) \sim \mathrm{ga}(k\alpha, \beta)$ （再生性）
- 第k日目が次回月経開始日となる事象 $\Delta(k \mid n) > 1 - \theta_n$
- 次回月経開始日を迎えるための位相の増分は $1 - \theta_n \mod 1$

逐次予測手法（左下からの続き）

- 事象が実現する確率（＝条件付き累積分布関数）

$$F(k \mid \theta_n) \stackrel{\mathrm{def}}{=} \Pr(\Delta(k \mid n) > 1 - \theta_n) = \int_{1-\theta_n}^{\infty} \mathrm{ga}(x; k\alpha, \beta)\, dx$$
$$= 1 - \mathrm{Ga}(1 - \theta_n; k\alpha, \beta)$$

ガンマ分布の累積分布関数

- 条件付き確率質量関数の計算

$$p(k \mid \theta_n) \stackrel{\mathrm{def}}{=} F(k \mid \theta_n) - F(k-1 \mid \theta_n)$$
$$= \mathrm{Ga}(1 - \theta_n; (k-1)\alpha, \beta) - \mathrm{Ga}(1 - \theta_n; k\alpha, \beta)$$

- 予測分布の計算 $p(k \mid y_{1:n}) \stackrel{\mathrm{def}}{=} \int \underline{p(k \mid \theta_n)} p(\theta_n \mid y_{1:n})\, d\theta_n$
  時刻nにおけるフィルタ分布
- 予測値の計算 $k_{max} = \underset{k}{\mathrm{argmax}}\, p(k \mid y_{1:n})$

ガウス過程状態空間モデルによる定式化

$$f(\theta) \sim \mathcal{GP}_f(m_f(\theta; \boldsymbol{\Phi}_f), k_f(\theta, \theta'; \boldsymbol{\Theta}_f))$$
$$\theta_n = f(\theta_{n-1}) + \epsilon_n \mod 1, \quad \epsilon_n \sim \mathrm{ga}(\alpha, \beta)$$
$$g(\theta) \sim \mathcal{GP}_g(m_g(\theta; \boldsymbol{\Phi}_g), k_g(\theta, \theta'; \boldsymbol{\Theta}_g))$$
$$y_n = g(\theta_n) + \sigma_n, \quad \sigma_n \sim \mathcal{N}(0, \sigma^2)$$

$\boldsymbol{\Phi}_f, \boldsymbol{\Theta}_f$
$\boldsymbol{\Phi}_g, \boldsymbol{\Theta}_g$:ハイパパラメタ

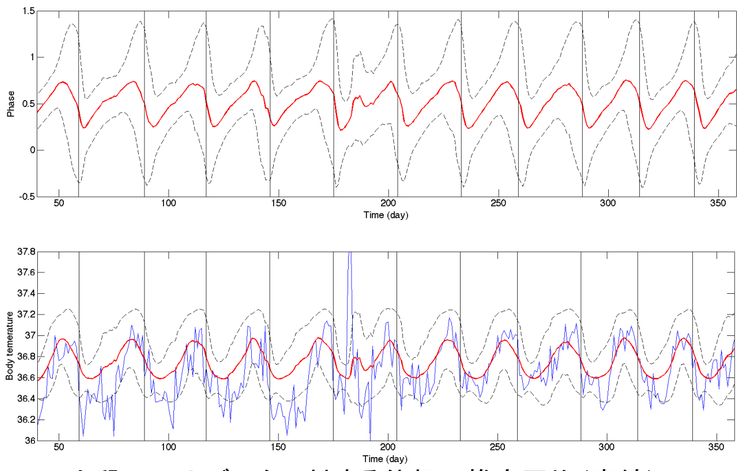
⇒状態/観測関数，ハイパパラメタをParticle MCMC法により推定

逐次予測手法
- 予測分布の計算⇒フィルタ分布のモンテカルロ積分による近似
- 予測値の計算⇒状態空間モデルによる定式化と同様

実験

女性の基礎体温データから次回月経開始日を予測

実験条件
- 学習データ:30代女性の口中計測
  日次体温データ130日分
- テストデータ:同女性の口中計測
  日次体温データ311日分
- Particle MCMCの
  サンプリング回数10000,
  粒子数5000
- GPのカーネル関数:RBFカーネル

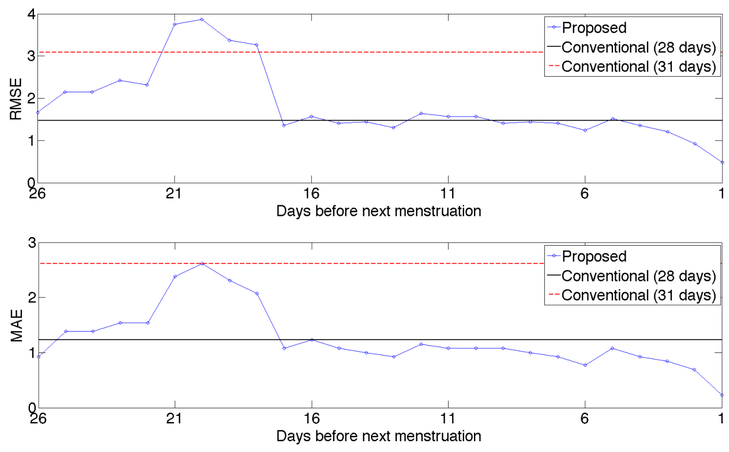


上段:テストデータに対する位相の推定平均（赤線）
下段:テストデータ（青線）と体温の推定平均（赤線）
ただし点線は（平均）＋（標準偏差）×2, 縦線は月経開始日（既知），横軸は時間（日にち）を表す

比較手法:前月の月経開始日から
28日&31日後を予測日とする手法

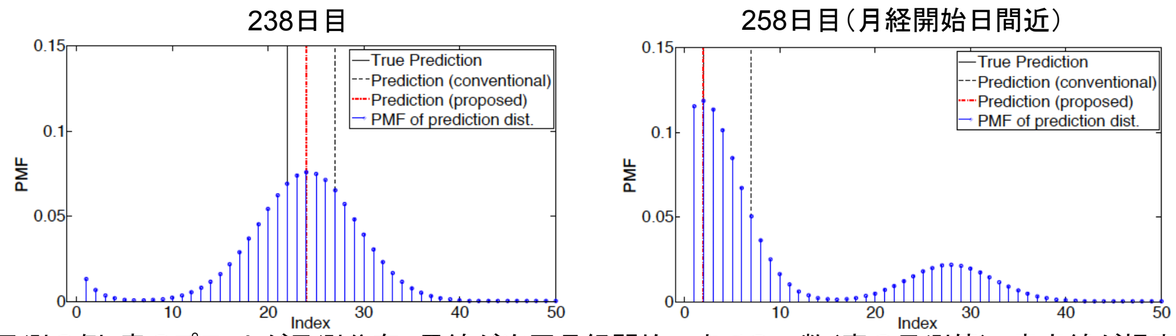
評価基準:最小二乗誤差（RMSE）
および最小絶対誤差（MAE）

実験結果:31日後に予測する
手法からの精度改善を確認



次回月経開始日の予測誤差（上段:RMSE，下段:MSE）:
横軸は次回開始日までの日数を表す

今後の課題
- 観測の異常値に対しても頑健な
  モデリング手法の検討
- SARIMAモデルや
  季節調整モデルとの比較実験
- 周期カーネルを適用した比較実験
- 被験者数を増やした実験



予測の例:青のプロットが予測分布，黒線が次回月経開始日までの日数（真の予測値），赤点線が提案法の予測値，黒点線が比較手法の予測値．横軸は予測分布のインデックス k，縦軸は予測分布の値を表す

大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics