

Predictive Linear Gaussian Models 関連の発表紹介

東大 (航空) 町田・矢入研 博士 2 年 河原 吉伸

email : kawahara@space.rcast.u-tokyo.ac.jp

URL : <http://www.space.rcast.u-tokyo.ac.jp/kawahara/>

2006 年 7 月 26 日

Predictive Linear Gaussian (PLG) Models は, UAI 2005 において Singh らにより提案された動的システムのためのモデルである [1]. ICML2006 では, PLGs に関して, 以下の 2 件の発表が行われている.

- Wingate, D., and Singh, S. "Kernel Predictive Linear Gaussian Models for Nonlinear Stochastic Dynamical Systems"
- Rudary, M., and Singh, S. "Predictive Linear-Gaussian Models of Controlled Stochastic Dynamical Systems"

ここでは, PLGs および上記発表について簡単に紹介する.

1 Predictive Linear Gaussian Models について

動的システムを表現するための方法としては, 自己回帰モデル (AR, History window, Markov model, etc.) や状態空間モデル (HMM, LDS, etc.) が多く用いられる. PLGs は, 状態の予測表現に基づくモデル (PSRs: Predictive state representations) の連続版であり, 従来モデルに比べて主に次のような利点が主張されている (今のところ単次元観測系列のみ).

表現力 自己回帰モデルでは, 一定区間の過去の観測情報を用いて予測を行うため, それ以前の観測情報はシステムに反映されない (ただし状態空間モデルは, 隠れ変数を用いるためにこの問題はない). PSRs は, 状態空間モデル同様, 過去の全ての観測情報を反映したモデルであるため, より正確に動的性質を表現する事ができる.

パラメータ推定の大域性 一般に状態空間モデルのパラメータは, 観測から推定する際の大域性は保証されない. PLGs は, 同次元の状態空間モデルよりも少ないパラメータにより表現され, 十分な観測が得られた際の真値への収束が保証されるアルゴリズムが提案されている.

PLG モデルは, 未来の n ステップの観測系列 $Z_t = [Y_{t+1}, \dots, Y_{t+n}]'$ を用いて, 次式のように表される.

$$Y_{t+n+1} = \langle g, Z_t \rangle + \eta_{t+n+1} \quad (1)$$

η_{t+n+1} は分散が時刻に依らない白色ノイズ $\mathcal{N}(0, \sigma^2)$ であるが, Z_t と相関 $\text{Cov}[Z_t, \eta_{t+n+2}|y_{1:t}] = C$ を持つ. このデータと共に変化するノイズ項が, PLG の表現力の根源となっている. 結局パラメータは, n 次元ベクトル g , および σ, C となる. PLG モデルでは, 全ての観測は多変量ガウシアンにより

生成されるという仮定をおく.

- 状態量が n 次元である LDS は, n 次元の PLG と等価である事が示される (cf.[1] の A.3)
- Z_t の分布を $p(Z_t) = \mathcal{N}(\mu_t, \Sigma_t)$ とすると, これは過去の情報を集約した状態量となっており, その更新式は, カルマンフィルタと同様の形で, [1] 中の (8) および (9) 式のように与えられる.
- PLG モデルのパラメータは, 同次元の LDS よりも少なく, またデータとの直接的な意味を有するものとなっており (e.g. σ^2 は $(Y_{t+n+1} - g'Z_t)|y_{1:t}$ の分散), [1] の (10) ~ (14) のように真値への収束性が保障される推定アルゴリズムが導ける ([1] の Theorem.2).

2 Kernel Predictive Linear Gaussian Models for Nonlinear Stochastic Dynamical Systems

本論文では, カーネルを用いて PLG を非線形化した KPLG (Kernel Predictive Linear Gaussian) モデルを提案し, その推論およびモデル推定の方法が述べられている.

■ 2. The Kernel PLG Model KPLG モデルは, PLG 同様, 未来の観測系列を用いて次式のように与えられる.

$$Y_{t+n+1} = \sum_{j=1}^J \alpha_j k(\xi_j, Z_t) + \eta_{t+n+1} \quad (2)$$

ここで $k(\cdot, \cdot)$ は, 正定値カーネルである.

- ガウシアン・カーネルを用いた場合, [2] 中の (2), (3), (5) ~ (7) 式のように状態量の更新式が解析的に得られる
- 更に, Unscented Transformations を用いた近似 (Fig.2) により, 更新の計算量を改善することが可能となる (V_t の計算量: $O(J^2)$ $O(2(n+1)J)$).

■ 3. Model Estimation 具体的なアルゴリズムは示されていないが, オフラインおよびオンラインでパラメータを学習する方法が述べられている.

オフラインの場合の基本的なストラテジは,

1. システムの次元 n および基関数の数 J を決める (CV).
2. ξ_j およびガウシアン・カーネルの分散 ϕ_j を決める (ランダム, 辞書ベース, EM).

3. 基関数の重み α_j を決める (回帰問題と等価・SVR や KRS などを適用) .

■4.Experiments いくつかの代表的な時系列データ (Santa Fe Laser, Mackey-Glass, K.U.Leuven) に対して, PLG, KPLG, KPLG-SP (UT を用いた近似) および比較のために KAR (Kernel AR) を適用して, 予測性能 (短期・長期) を比較している. 特に特徴的なのは, Fig.5 にも示されている長期予測 (数十~百ステップ) の性能であり, かなり正確にダイナミクスを表すモデルが得られている事がわかる.

3 Predictive Linear-Gaussian Models of Controlled Stochastic Dynamical Systems

本論文では, 制御入力を考慮した PLG モデルである cPLG (controlled Predictive Linear Gaussian) モデルを提案し, その推論およびモデル推定の方法が述べられている.

■2.The cPLG Model cPLG モデルは, 未来の観測および制御系列を用いて次式のように与えられる.

$$Y_{t+n+1} = \langle g, Z_t \rangle + \sum_{i=1}^{n-1} \langle \Gamma_i, u_{t+i+1} \rangle + \eta_{t+n+1} \quad (3)$$

PLG モデルに比べて, 入力 u_{t+i+1} による線形な影響を定義するベクトル Γ_i がパラメータとして加わる.

- 状態量が n 次元である controlled LDS は, n 次元の cPLG と等価である事が示される (cf.[3] の Theorem.1 および A.3)
- 状態量の更新式は, PLG と同様の形で, [3] 中の (2) および (3) 式のように与えられる.
- cPLG モデルのパラメータは, 同次元の controlled LDS よりも $(3n^2 - n)/2$ 少ない.

■3.Optimal Control of the cPLG cPLG モデルでは, 2 次のコスト関数 ([3]p.3 の真ん中あたり) における最適制御, つまりモデルにノイズがなく状態量が最小自乗推定の場合に最適となる制御入力が, 状態量の平均値の線形写像として与えられる事について述べられている ([3] の Lemma.2) .

■4.Parameter Estimation in cPLGs

- PLG 同様, cPLG のパラメータは真値への収束性が保障されるアルゴリズムにより推定する事が可能である ([3] の Theorem.4 および A.5) .
- cPLG に対して CE アルゴリズム, LDS に対して EM アルゴリズムを適用してモデルを学習した際の対数尤度を比較した実験を行っている ([3] の Fig.1) . 訓練に用いた観測系列が長くなる程, cPLG の優位性が大きくなっている.

付録 A

- LDS (1 次元の観測) . $y_t \in \mathbb{R}$ は観測量, $x_t \in \mathbb{R}^n$ は状態量, $v \in \mathbb{R}^n$ および $w \in \mathbb{R}^1$ は白色ノイズ.

$$\begin{aligned} x_{t+1} &= Ax_t + v \\ y_t &= Cx_t + w \end{aligned} \quad (4)$$

- AR モデル (1 次) . $y_t \in \mathbb{R}$ は観測量, $a_t \sim \mathcal{N}(0, \sigma^2)$, $\phi_i \in \mathbb{R}$ は基関数の重み.

$$Y_t = a_t + \sum_{i=1}^n \phi_i y_{t-i} \quad (5)$$

参考文献

- [1] Rudary, M., Singh, S., and Wingate, D. : "Predictive Linear-Gaussian Models of Stochastic Dynamical Systems" in Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence(UAI), 2005
- [2] Wingate, D., and Singh, S. : "Kernel Predictive Linear Gaussian Models for Nonlinear Stochastic Dynamical Systems" in Proceedings of the 23rd International Conference on Machine Learning(ICML), 2006
- [3] Rudary, M., and Singh, S. "Predictive Linear-Gaussian Models of Controlled Stochastic Dynamical Systems" in Proceedings of the 23rd International Conference on Machine Learning(ICML), 2006