

ダイナミクスのベイズ推定に基づくマルチモジュール型適応制御モデル

Modular adaptive control with Bayesian dynamics estimation

大佐賀 智 (PY)[†], 平山 淳一郎[†], 竹之内 高志[‡], 石井 信[†]

Satoshi Osaga(PY)[†], Jun-ichiro Hirayama[†], Takashi Takenouchi[‡], Shin Ishii[†]

[†] 京都大学大学院情報科学研究科, [‡] 奈良先端科学技術大学院大学情報科学研究科

satosh-o@sys.i.kyoto-u.ac.jp

Abstract— Human can achieve accurate and appropriate motor behaviors in various and even uncertain environments. MOSAIC was proposed for describing such a human ability. In this study, we present an online Bayesian learning of MOSAIC, which incorporates uncertainty into parameters' learning.

Keywords— Uncertainty, Modularity, Adaptive control, Variational Bayes

1 導入

ヒトは複雑かつ不確かな環境において適切な行動制御を行うために、自身が置かれた環境のダイナミクスや身体の制御則を経験に基づき学習することができる。しかし、実際の環境は時間とともに変化しうるため、状況ごとに適切な制御則を学習し、それらを状況に応じて切り替える仕組みが必要となる。このような機能を実現する計算論的原理を解明することは、脳の環境適応能力を理解する上で重要である。我々は、先行研究 [3] において、小脳の解剖学的特徴に基づくモジュール型学習制御モデルである MOSAIC [1] を統計的視点から再定式化した。本研究では、このモデルをベイズ学習へと拡張し、運動学習における学習の不確かさを適切に取り扱うことの有効性を計算機シミュレーションを通して検証する。

2 確率的運動制御モデル

時間経過にしたがってダイナミクスが不連続に変化する環境において、前もって与えられた目標軌道に制御対象が追従するように制御することを考える。制御対象の状態 (位置, 速度など) は一定の時間間隔で得られるものとし、制御は状態の値が得られるごとにオンラインで行われるものとする。いま、時刻 t における状態を \tilde{x}_t 、その状態における制御出力を \tilde{u}_t 、 \tilde{x}_t が達成すべき目標状態を x_t^* と表せば、ここでの目的は状態系列 $\{\tilde{x}_1, \dots, \tilde{x}_T\}$ と目標系列 $\{x_1^*, \dots, x_T^*\}$ との誤差をできるだけ小さくする (確率的) 制御器 $q(u_t|x_{t-1}, x_t^*)$ を学習することである。

2.1 制御器

MOSAIC などのモジュール型適応制御モデルでは、不連続に変化するダイナミクスに対処するため、それぞれ異なるダイナミクスに対応する複数の比較的単純な制御器を仮定し、それらを適応的に切り替えることで制御を行う。本研究では、 i 番目の制御器として以下の

線形ガウスモデルを仮定する。

$$u_{i,t} = v_{1i}\tilde{x}_t + v_{2i}x_{t+1}^* + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, \rho^{-1})$$

ここで $\mathcal{N}(\mu, \sigma^2)$ は平均 μ 、分散 σ^2 の正規分布、 $\mathbf{v}_i = (v_{1i}, v_{2i})^T$ は制御器パラメータ、 ρ は各制御器に共通した逆分散である。最終的な制御出力は M 個の制御器の組み合わせにより以下で与える。

$$u_t = \sum_{i=1}^M \lambda_i u_{i,t}$$

ここで、 $\lambda_i \geq 0$ 、 $\sum_{i=1}^M \lambda_i = 1$ とする。上式を確率分布として書けば、以下が得られる。

$$p(u_t|\tilde{x}_t, x_{t+1}^*; \boldsymbol{\lambda}, \mathbf{v}) = \mathcal{N}\left(u_t \mid \sum_{i=1}^M \lambda_i u_{i,t}, \rho^{-1}\right) \\ \propto \prod_{i=1}^M \mathcal{N}(u_t | v_{1i}\tilde{x}_t + v_{2i}x_{t+1}^*, \rho^{-1})^{\lambda_i}$$

ここで $\mathbf{v} = \{\mathbf{v}_1, \dots, \mathbf{v}_M\}$ である。制御出力 \tilde{u}_t はこの分布からサンプリングにより決定される。ただし、状況に応じた適切なモジュール組み合わせを実現するには、各時刻において重み λ を適切に決定する必要がある。

2.2 予測器

MOSAIC やそれに倣った我々の手法では、制御器と一対一に対応する状態予測器を用い、重み λ をそれらの予測誤差に基づいて決定する。本研究では、 i 番目の制御器に対応するものとして以下の線形予測器を仮定する。

$$x_{t+1} = w_{1i}\tilde{x}_t + w_{2i}\tilde{u}_t + \delta_{i,t}, \quad \delta_{i,t} \sim \mathcal{N}(0, \xi_i^{-1})$$

ここで $\mathbf{w}_i = (w_{1i}, w_{2i})^T$ は i 番目の予測器のパラメータ、 ξ_i は各予測器の逆分散である。 M 個の予測器について混合比 λ を仮定すると、以下の混合モデルが得られる。

$$p(x_{t+1}|\tilde{x}_t, \tilde{u}_t, \boldsymbol{\lambda}, \mathbf{w}, \boldsymbol{\xi}) \\ = \sum_{i=1}^M \lambda_i \mathcal{N}(x_{t+1} | w_{1i}\tilde{x}_t + w_{2i}\tilde{u}_t, \xi_i^{-1}) \quad (1)$$

λ をゆるやかに変化するパラメータとみて、この混合モデルに基づいてオンライン最尤推定を行うことにより、近い過去の累積予測誤差に基づいて適切に重み λ を決定することができる [3]。ただし、そのためには個々の予測器パラメータ $\{\mathbf{w}_i, \xi_i\}$ が学習済みである必要がある。

3 予測器のベイズ推定に基づく学習

前節で説明したモジュール型適応制御にはいくつかの問題点があった．第一に，予測器と制御器のオンライン学習を並行して行う場合，予測器の学習が不十分な初期段階において重み入の評価が適切に行えず，結果として適切な制御則の獲得に失敗することがあった．第二に，制御対象のダイナミクスの複雑さに対応するモジュール数を事前に適切に決めておく必要があるが，これは未知の環境では困難である．本研究では，ベイズ学習を用いることで予測器学習に「不確かさ」の効果を取り入れ，上記の問題点の解決を図る．

混合モデル (1) に対するオンライン変分ベイズ学習 [2] を導出する．時刻 t におけるモジュールを表す M 項隠れ変数 $\mathbf{c}_t = (c_{t1}, \dots, c_{tM})$ を導入すると， \mathbf{c}_t とパラメータ $\{\mathbf{w}, \xi, \lambda\}$ の (パッチ型) 事後分布は以下の形で与えられる．

$$q(\mathbf{w}, \xi) \propto p(\mathbf{w}, \xi) \exp \left(\sum_{t=1}^{\tau} \langle \log p(\tilde{x}_{t+1} | \tilde{x}_t, \tilde{u}_t, \mathbf{c}_t, \mathbf{w}, \xi) \rangle_{\mathbf{c}_t} \right) \quad (2a)$$

$$q(\lambda) \propto p(\lambda) \exp \left(\sum_{t=1}^{\tau} \langle \log p(\mathbf{c}_t | \lambda) \rangle_{\mathbf{c}_t} \right) \quad (2b)$$

$$q(\mathbf{c}_t) \propto \exp \langle \log p(\mathbf{c}_t | \lambda) \rangle_{\lambda} \times \exp \langle p(\tilde{x}_{t+1} | \tilde{x}_t, \tilde{u}_t, \mathbf{c}_t, \mathbf{w}, \xi) \rangle_{\mathbf{w}, \xi} \quad (2c)$$

ここで， $\langle \cdot \rangle_{\theta}$ は $q(\theta)$ で期待値をとることを意味しており， $p(\mathbf{w}, \rho)$ ， $p(\lambda)$ ， $p(\mathbf{c}_t | \lambda)$ は事前分布である．特に共役事前分布を用いることで，(2a-c) は以下のように求めることができる¹．

$$q(\mathbf{w}, \xi) = \prod_{i=1}^M \{ \mathcal{G}(\xi_i | \alpha_i, \beta_i) \mathcal{N}(\mathbf{w}_i | \bar{\mathbf{w}}_i, (\xi_i \Omega_i)^{-1}) \}$$

$$q(\lambda) = \mathcal{D} \left(\{ \lambda_i \}_{i=1}^M | \{ \gamma_0 + N_i \}_{i=1}^M \right)$$

$$q(c_{i,t} = 1) \propto \exp(\zeta_{i,t})$$

ここで $\mathcal{G}(\cdot)$ はガンマ分布， $\mathcal{D}(\cdot)$ はディリクレ分布である．十分統計量に対する忘却係数の適切なスケジューリングのもとで，事後分布の (ハイパー) パラメータを逐次的に計算するものとして，オンライン変分ベイズの学習則が得られる．ただし，隠れ変数の推定に関しては，時刻 t において $q(\mathbf{c}_t)$ のみを計算し，過去の $q(\mathbf{c}_1), \dots, q(\mathbf{c}_{t-1})$ については固定しておく．また，ダイナミクスの変動に素早く追従させるため， $\{\mathbf{w}, \xi\}$ と比較して重み入の推定に強い忘却を与えるヒューリスティクスを用いた．

制御器の学習は，[3] と同様にオンライン EM アルゴリズムで行った．

4 シミュレーション

パラメータのオンラインベイズ学習の効果を検証するために計算機シミュレーションを行った．シミュレーションはばね質量ダンパ系を前もって与えられた目標軌道に沿って 12 秒間動かすものである．ただし，2 秒ご

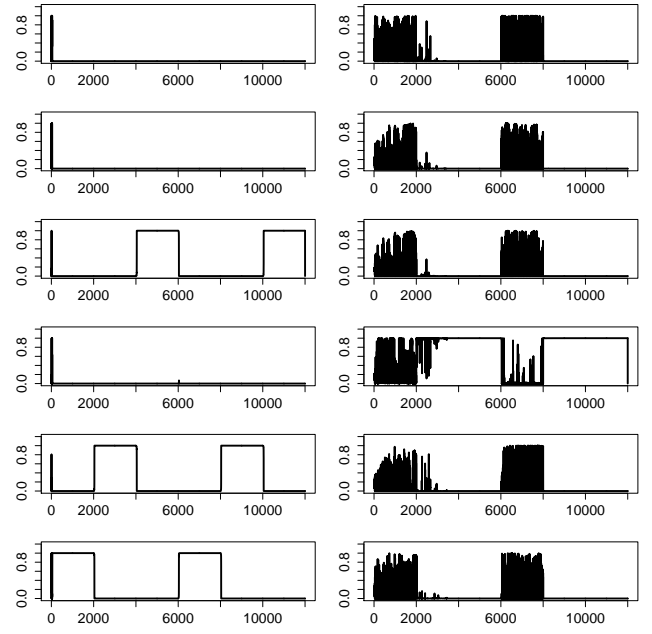


図 1: 6 モジュールの場合の混合比 λ ．横軸は時刻 (ミリ秒)，縦軸は混合比の値を表している．左列はベイズ推定，右列は MAP 推定を用いたもの，行はモジュールのインデックスと対応している．

とに 3 種類の力学系 (それぞれ質量などが異なる) に不連続に切り替わる．パラメータをベイズ推定，MAP 推定した場合について責任信号の比較を行った．

図 1 は，モジュール数が 6 の場合に，責任信号の振る舞いを示している．ベイズ推定によれば，力学系の種類 (3 種) と比べてモジュール数が過剰の場合でも，必要な数のモジュールだけが用いられ，ダイナミクスの変化に追従して制御器の理想的な切り替えがなされていることが分かる．

5 結論

先に提案していた確率的 MOSAIC モデルの予測器パラメータを変分ベイズ法で推定する手法を示した．学習の不確かさをモデルに入れ込むことの有効性を，特にモジュール選択の面について明らかにした．

参考文献

- [1] D. Wolpert, M. Kawato (1998) “Multiple paired forward and inverse models for motor control.” *Neural Networks*, **11**, 1317–1329.
- [2] M. Sato (2001) “Online model selection based on the variational Bayes.” *Neural Computation*, **13**, 1649–1681.
- [3] S. Osaga, J. Hirayama, T. Takashi, S. Ishii (2008) “A probabilistic modeling of MOSAIC learning.” *Journal of Artificial Life and Robotics*, **12**, 1-2, 167–171.

¹ハイパーパラメータの詳細については紙面の都合で省く