

次元削減と動的システムの学習

矢入健久*

Takehisa Yairi

Abstract: 様々な動的なシステムのモデルを蓄積された観測データから推定・学習する問題は、制御工学 (システム同定) と機械学習の両分野に共通する興味対象である。特に近年では、センサー技術の発達等により観測データの高次元化が進んでおり、高次元の観測空間から低次元のシステムの本質的な状態遷移を見つけ出す「次元削減」の重要性が増している。本発表では、この観点からシステム同定・機械学習両分野でのトレンドを概観するとともに、発表者らの取り組みを紹介する。

Keywords: 動的システム学習, システム同定, 状態空間モデル, 次元削減, 部分空間同定法

1 まえがき

時間的に状態を変化させるシステム、すなわち、動的なシステムから発生するデータから、そのシステムのモデル構造・パラメータを推定する問題は、機械学習分野では動的システム学習 (learning of dynamical systems)、制御工学分野ではシステム同定 (system identification) と呼ばれ、それぞれの分野で近年まで互いにあまり干渉することなく独自の発展を遂げてきた。

本発表では、これら2つの分野における近年の研究動向を概観して両者の立場や興味、方法論の相違点を探るとともに、両者の融合による新たな動的システム学習理論の可能性を考察したい。

なお、以下の議論では、次式の (離散時間) 状態空間モデルによって表される動的システムを想定するものとする。

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{v}_t \quad (1)$$

$$\mathbf{y}_t = g(\mathbf{x}_t) + \mathbf{w}_t \quad (2)$$

2 機械学習分野における動的システム学習研究の動向

2.1 スイッチング線形モデルの学習

未知の非線形システムをモデル化する最も基本的かつ現実的な方法は、複数の線形モデルによって局所的に近似することであろう。機械学習分野では、そのようなモデルは、スイッチング線形動的システム (Switching Linear Dynamical System: SLDS)、スイッチングカル

マンフィルタ (SKF) などと呼ばれている。SLDS モデルをデータから学習するには、通常、EM アルゴリズムや、それに類する反復的アルゴリズムが用いられるが、E ステップについては、どの局所モデルを選択するかを表す離散的確率変数 s_t と連続の状態変数 x_t の同時事後分布を厳密に推定することが困難であるため、変分近似を用いる方法 [6] や s_t の推定に Viterbi アルゴリズムを用いる方法 [19] などが提案されている。また、SLDS の学習では、そもそも局所線形モデルの数をいくつにするべきか、という問題が存在するが、最近の研究 [5, 3] では、ディリクレ過程混合モデルや変分ベイズ法を用いることによってモデル数を自動決定する手法が提案されている。

2.2 非線形潜在変数モデルの学習

非線形システムを学習するもう一つのアプローチは、状態空間モデル 1,2 において、状態遷移関数 f および出力関数 g を、非線形回帰や次元削減の手法を用いてデータから推定することである。

まず、カーネル以前の非線形アプローチの例としては、Ghahramani と Roweis による研究 [7] が有名である。これは、状態空間モデルにおける状態方程式および観測方程式を Radial Basis Function Networks (RBFN) によって表現し、そのモデルパラメータを EM アルゴリズムによって反復的に学習するというものである。

Kernel Kalman Filter (KKF) [20] は、「非線形なシステムであっても、カーネルによって写像された特徴空間では線形なモデルで表される」という考えに基づき、線形の状態空間モデルをカーネルトリックによって非線形化したものである。この手法では、特徴空間における状

*東京大学先端科学技術研究センター, 〒153-8904 東京都目黒区駒場 4-6-1, e-mail yairi@space.rcast.u-tokyo.ac.jp, RCAST, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo

態・観測方程式のモデルパラメータはやはり EM アルゴリズムによって学習される。ただし、学習した KKF を実際にフィルタリングや予測に用いるためには、特徴空間から元の観測空間への逆写像 (preimage) を求める必要がある。また、カーネルと後述の部分空間同定法を融合した例として、カーネル正準相関分析 (Kernel Canonical Correlation Analysis: KCCA) を用いた [11] が挙げられる。

また、カーネル行列を分散共分散行列とみなすことによって正規過程 (Gaussian Process) が導かれるが、この立場から動的システムを扱ったものとして、正規過程動的モデル (Gaussian Process Dynamical Models: GPDM)[28] が挙げられる。GPDM は、非線形次元削減法である正規過程潜在変数モデル (Gaussian Process Latent Variable Models: GPLVM)[15] を、隠れ状態のダイナミクスを扱うように拡張したものである。最近、GPDM と類似の手法として、GPIL[24] が提案されているが、こちらでは擬似入力 (pseudo inputs) によるスパース化が図られている。また、これらと同様に、状態遷移関数 f 、出力関数 g を正規過程によってモデル化するものとして、GP-Bayes[12]、GP-ADF[4] などもある。ただし、これらは訓練時において状態変数 $\{x_t\}$ の真値が与えられると仮定しており、その学習は単に 2 つの GP 回帰に帰着される。

グラフベースな非線形次元削減、いわゆる多様体学習 (Manifold Learning) の非線形動的システムへの応用は今のところ多くないが、ラプラス固有写像 (Laplacian Eigenmap: LE)[2] を確率的な潜在変数モデルに拡張したものを非線形な観測モデルとして利用し、モーションキャプチャデータ等のモデル化に応用した例 [17] がある。ただし、この例では状態遷移には単純なランダムウォークモデルを利用している。

2.3 非定常 DBN の構造学習

動的ベイジアンネットワーク (DBN) は、確率的な動的システムを表現・モデル化する手段として機械学習では広く取り入れられているが、特に近年では、時間的にモデル構造やパラメータが変化するような、非定常 DBN の学習についての研究も行われている [21, 23]。

2.4 その他の話題

上で見たように、これまでのところ、機械学習における動的システムの学習では、「確率的な」手法であるガウス過程などが好まれて用いられている。これはフィルタリングや平滑化などの確率的推論との相性の良さに依るところが大きい。

しかし、ごく最近、この常識を変える可能性のある研究が Langford により発表されている [13]。その基本的なアイデアは、確率的な状態変数をそのまま扱う代わりに、その十分統計量を決定論的な状態変数として利用する、という比較的単純なものであるが、これにより、任意の「確率的でない」非線形回帰アルゴリズム (例えば SVR) を、確率的な動的システムの学習に用いることができるようになるという。

3 システム同定分野における動向

80 年代後半以降システム同定分野では、部分空間同定法 (subspace identification) と呼ばれる一連の方法が盛んに研究されている [18, 14, 27]。部分空間同定法では、システムの過去および未来の入出力データが張る部分空間上での、直交射影 (orthogonal projection) や斜交射影 (oblique projection) などの幾何学的演算によって状態ベクトルやモデルパラメータが求められる。

部分空間同定法は、線形動的システムに関する強固な理論に裏打ちされたものであると同時に、「状態とは、未来 (過去) を予測するために必要な過去 (未来) の情報を縮約したもの」[30] という直感的な解釈も可能であるところが魅力的である。その一方、部分空間同定法の非線形化に関しては、(あくまで筆者の印象ではあるが、) まだ限定的と言って良いであろう。また、前述の SLDS に似た考えとして、区分線形 (piecewise linear: PWL) システムに対する部分空間同定法 [26] もあるが、各時刻の観測がどの局所線形モデルから生じているかについて常に情報が得られるという強い仮定に基づいている。

なお、非線形モデルの学習という話題に関しては、[16] によれば、システム同定分野においてもカーネル [8] や多様体学習 (グラフ正則化)[9] など、様々な機械学習の手法や理論を取り入れようとする機運が高まっているようである。また、部分空間同定法以前の主流であった予測誤差法 (Prediction Error Minimization: PEM) は最尤推定と密接に関係しているが、意外にもシステム同定分野で EM アルゴリズムが認知されるようになったのは比較的最近のようである [22, 16]。

4 局所線形モデルの整列による非線形システムの学習法

上では、機械学習およびシステム同定における動的システムの学習の従来研究を概観したが、これら 2 つの方法論を融合する 1 方法として、著者らは局所線形動的モデルの整列による非線形システムの学習法に取り組んでいる [10]。

これは、CCA に基づく (線形の) 部分空間同定法を、CCA の確率的解釈 [1] と局所線形モデルの整列法 [25] によって拡張したものである。詳細については [10] を参考にされたい。

5 おわりに

本稿では、状態空間モデルによって表現される動的システムのデータからの学習法に関して、機械学習分野およびシステム同定分野におけるトレンドを概観した。また、著者らのグループが行っている両分野の要素技術を取り入れた新しい非線形動的システムの学習法を紹介した。今後も、機械学習とシステム同定の両分野が交流を持ち、相互の発展につながることを期待したい。

謝辞

本稿における動的システム学習に関するサーベイの大部分は、井手剛氏 (IBM) との共著解説 [29] に基づいている。また、部分空間同定法と機械学習との関連性については、河原吉伸氏 (現・大阪大学) の博士論文研究から得た情報に基づいている。なお、著者らのグループの取り組みとして紹介した「局所線形モデルの整列による非線形システムの学習法」は、河原氏および東大院修士課程の上甲昌郎氏との共同研究である。これらの関係者の協力を厚くお礼を申し上げる次第である。

参考文献

- [1] Francis R. Bach and Michael I. Jordan. A probabilistic interpretation of canonical correlation analysis. Technical Report 688, Department of Statistics, University of California, Berkeley, 2005.
- [2] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems 14*, pages 585–591. MIT Press, 2001.
- [3] Silvia Chiappa, Jens Kober, and Jan Peters. Using bayesian dynamical systems for motion template libraries. In *Advances in Neural Information Processing Systems 21*. MIT Press, 2009.
- [4] Marc Peter Deisenroth, Marco F. Huber, and Uwe D. Hanebeck. Analytic moment-based gaussian process filtering. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 225–232, 2009.
- [5] Emily Fox, Erik Sudderth, Michael Jordan, and Alan Willsky. Nonparametric bayesian learning of switching linear dynamical systems. In *Advances in Neural Information Processing Systems 21*. MIT Press, 2009.
- [6] Zoubin Ghahramani and Geoffrey E. Hinton. Variational learning for switching state-space models. *Neural Computation*, 12:963–996, 1998.
- [7] Zoubin Ghahramani and Sam Roweis. Learning nonlinear dynamical systems using an em algorithm. In *Advances in Neural Information Processing Systems 11*, pages 431–437. MIT Press, 1999.
- [8] Ivan Goethals, Kristiaan Pelckmans, Johan A. K. Suykens, and Bart De Moor. Subspace identification of hammerstein systems using least squares support vector machines. *IEEE Trans. on Automatic Control*, 50:1509–1519, 2005.
- [9] Lennart Ljung Henrik Ohlsson, Jacob Roll. Manifold-constrained regressors in system identification. In *Proc. 47th IEEE Conference on Decision and Control*, pages 1364–1369, 2008.
- [10] Masao Joko, Yoshinobu Kawahara, and Takehisa Yairi. Learning non-linear dynamical systems by alignment of local linear models. In *20th International Conference on Pattern Recognition (ICPR)*, page (to appear), August 2010.
- [11] Yoshinobu Kawahara, Takehisa Yairi, and Kazuo Machida. A kernel subspace method by stochastic realization for learning nonlinear dynamical systems. In *Advances in Neural Information Processing Systems 19*, pages 665–672. MIT Press, 2007.
- [12] Jonathan Ko and Dieter Fox. Gp-bayesfilters: Bayesian filtering using gaussian process prediction and observation models. *Autonomous Robots*, 27(1):75–90, 2009.
- [13] John Langford, Ruslan Salakhutdinov, and Tong Zhang. Learning nonlinear dynamic models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 593–600, 2009.

- [14] Wallace E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *Proceedings of the 29th IEEE Conference on Decision and Control*, pages 596–604, 1990.
- [15] Neil D. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.
- [16] Lennart Ljung. Perspectives on system identification. In *the IFAC Congress*, 2008.
- [17] Zhengdong Lu, Miguel Carreira-Perpinan, and Cristian Sminchisescu. People tracking with the laplacian eigenmaps latent variable model. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1705–1712. MIT Press, Cambridge, MA, 2008.
- [18] Peter Van Overschee and Bart De Moor. N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 31:75–93, 1994.
- [19] Vladimir Pavlovic, James M. Rehg, and John Maccormick. Learning switching linear models of human motion. In *Advances in Neural Information Processing Systems 13*, pages 981–987. The MIT Press, 2001.
- [20] L. Ralaivola and F. d’Alche Buc. Time series filtering, smoothing and learning using the kernel kalman filter. In *Proc. of IEEE Int. Joint Conference on Neural Networks*, pages 1449–1454, 2005.
- [21] Joshua W Robinson and Alexander J Hartemink. Non-stationary dynamic bayesian networks. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1369–1376. MIT Press, 2009.
- [22] Thomas B. Schn, Adrian Wills, and Brett Ninness. Maximum likelihood nonlinear system estimation. In *In Proceedings of the 14th IFAC Symposium on System Identification*, pages 1003–1008, 2006.
- [23] Le Song, Mladen Kolar, and Eric Xing. Time-varying dynamic bayesian networks. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 1732–1740. MIT Press, 2009.
- [24] Ryan Turner, Marc Deisenroth, and Carl Rasmussen. State-space inference and learning with gaussian processes. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 868–875, 2010.
- [25] Jakob Verbeek. Learning nonlinear image manifolds by global alignment of local linear models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(8):1236–1250, 2006.
- [26] Vincent Verdult and Michel Verhaegen. Subspace identification of piecewise linear systems. In *Proceedings of 43rd IEEE Conference on Decision and Control*, pages 3838–3843, 2004.
- [27] Michel Verhaegen. Identification of the deterministic part of mimo state space models in innovations form from inputoutput data. *Automatica*, 30(1):61–74, 1994.
- [28] Jack Wang, David Fleet, and Aaron Hertzmann. Gaussian process dynamical models. In *Advances in Neural Information Processing Systems 18*, pages 1441–1448. MIT Press, 2006.
- [29] 井手 剛 and 矢入 健久. 機械学習技術の最近の発展とシステムモデリングへの応用. *計測と制御*, 49(7):(to appear), 2010.
- [30] 片山 徹. システム同定: 部分空間法からのアプローチ. 朝倉書店, 2004.