解説：特集
ダイナミクスと機械学習の融合に挑む

機械学習とシステム同定：動的システム学習研究の動向

矢　入　健　久[*]

＊東京大学大学院工学系研究科　東京都文京区本郷 7–3–1
＊School of Engineering, The University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo, Japan
＊E-mail: yairi@g.ecc.u-tokyo.ac.jp

キーワード：機械学習 (machine learning)，動的システム学習 (learning dynamical systems).
JL 0003/19/5803–0176 ©2019 SICE

1. まえがき

近年の人工知能・機械学習ブームに伴って，教師あり分類学習，回帰学習，深層学習，強化学習，などの「メジャー」な機械学習問題･手法は広く知られるようになったが，**動的システム学習** (learning dynamical systems) というジャンルは本誌の読者にとってあまり聞き慣れないものかもしれない．それどころか，機械学習のコミュニティにおいてさえも，必ずしも**動的システム学習**という分野が認知され確固たる地位を築いているとはいい難い．しかしながら,「システム同定問題への機械学習的なアプローチ」と表現すれば，本学会の方々には大体の雰囲気を掴んでもらえるのではないかと思う．

多くの機械学習の入門書では，「データサンプル $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ は互いに独立かつ同一の分布に従う (independent and identically distributed; i.i.d.) と仮定する」とことわった上で，教師あり学習（分類学習，回帰学習）や教師なし学習（クラスタリング，次元削減，等）の説明が展開されている．つまり，機械学習では，サンプル同士が独立であることが基本であり，そうでない場合，つまり，サンプル同士の順序関係や時系列性を明示的に扱うのはやや特殊なものという位置づけになっている．もちろん，適用対象によっては順序データや時系列データを扱わないといけないが，そういう場合，まず，固定長窓を使ってベクトル化したり，タスクに適した特徴量を抽出することによって，i.i.d. なデータサンプルに変換し，静的なシステムを対象とした既存の学習手法を利用するというのが一般的な機械学習の「作法」といえる．そのような中，時系列データに特化し，その観測データの背後にあるダイナミクスの生成モデル (generative model) を推定（学習）しようとするのが**動的システム学習**である．

ところで，本誌の読者であれば,「**動的システム学習**は制御学における**システム同定**と何が違うのか？」という疑問をもつのは当然であろう．どちらも，時系列の計測データから動的なシステムのモデルを獲得することを目的としている点は共通している．一般論としては，システム同定のほうがよりモデルを重視しシステムの制御を最終的な目標としているのに対して，動的システム学習はデータを重視し将来の予測や異常検知などを目的とすることが多いといえる．また，システム同定では事前に対象システムの特性や利用する観測データについて注意深く分析するのが一般的であるのに対して，動的システム学習では「まずデータありき」という立場でそのデータの生成を尤もらしく説明できるモデルを推定してみるという「データドリブン」な指向性が強い．もっとも，両者を目的や傾向の違いによって明確に区別することは困難になっており，今後，制御学と機械学習の交流がさらに深まれば方法論・手段も融合が図られて，ますます境界が曖昧になると予想される．

以下，本稿では，機械学習の文脈で，動的システム学習に関する諸研究を，(1) 最尤法アプローチ，(2) スペクトルアプローチ，(3) ニューラルネットワークアプローチの 3 つに分けて解説する．なお，この 3 カテゴリーは著者が本稿の執筆にあたって便宜上名付けたものであり，分野全体の同意を得ているわけではないこと，また，各カテゴリーが必ずしも互いに排他的になっているわけではないことをことわっておく．

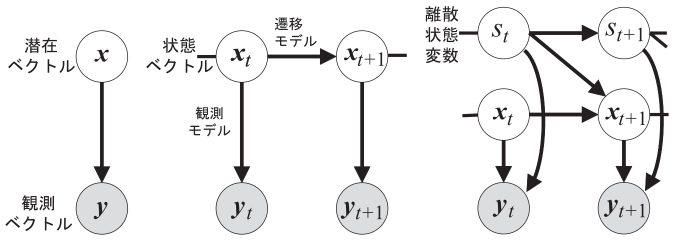
2. 最尤法に基づく動的システム学習

2.1 潜在変数モデルと EM アルゴリズム

動的システム学習の最も基本的な方法は，システムモデルを確率的な潜在変数モデルによって表現し，そのモデルパラメータを最尤法などによって推定することである．すなわち，モデルパラメータを $\boldsymbol{\theta}$，計測データを $\boldsymbol{y}_{1:T}$，システムの潜在変数（状態変数）列を $\boldsymbol{x}_{1:T}$ としたとき，

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \int p(\boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}|\boldsymbol{\theta}) d\boldsymbol{x}_{1:T} \quad (1)$$

として，モデルパラメータを推定する．潜在変数モデル (latent variable model; LVM) は，機械学習では，クラスタリングや次元削減などの教師なし学習で頻繁に使われるが，状態空間モデル (state space model; SSM) も潜在変数モデルの一種と解釈することができ（**図 1**），潜在変数モデルの最尤推定手段としての定番である **EM アルゴリズム** (expectation-maximization algorithm) を利用することができる．また，モデルパラメータ $\boldsymbol{\theta}$ についての事前分布が得られる場合には，最大事後確率 (maximum a posteriori; MAP) 推定，



(a)（最も単純な）潜在変数モデル　(b) 状態空間モデル（自律系）　(c) スイッチング動的システム

図１　潜在変数モデル・状態空間モデル・スイッチング動的システムのグラフ表現

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{y}_{1:T}) = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{2}$$

さらには，点推定の代わりに，パラメータについての事後分布を求めるベイズ推定 (Bayesian estimation)

$$p(\boldsymbol{\theta}|\boldsymbol{y}_{1:T}) = \frac{p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{3}$$

へと拡張することが可能である．なお，本稿では便宜上，MAP 推定，ベイズ推定による動的システム学習も広義の最尤推定アプローチと称することに注意されたい．このアプローチの最大の特徴は，(1) モデルパラメータを固定して潜在変数（状態ベクトル）の分布を推定する E ステップと，(2) E ステップで求めた状態変数に関する分布に関しての尤度期待値を最大化することによってモデルパラメータを更新する M ステップ，の反復によってモデルの学習を行う点である．

2.2　線形状態空間モデルの EM アルゴリズム

隠れマルコフモデル (hidden Markov model; HMM) も潜在変数モデルの一種であり，状態空間モデルにおいて状態変数が離散的な値を取る場合に対応すると解釈できる．そして，HMM のための EM アルゴリズムは，1970 年代にすでにバウム-ウェルチアルゴリズムとして発明され，音声認識などに応用され機械学習分野でも早くから知られていた．

一方，意外なことに，線形ガウス状態空間モデルのための EM アルゴリズムは 1996 年に Ghahramani と Hinton によって書かれた技術ノート[1] が最初とされており，2000 年代に Bishop によって統計的機械学習のバイブルともいわれる PRML[2] で紹介されることによって機械学習コミュニティに広く知られるようになった．線形ガウスシステムの EM アルゴリズムでは，E ステップが Kalman フィルタ・スムーザ（Rauch-Tung-Striebel スムーザ）に一致している．EM アルゴリズムは，一見すると交互最小二乗法 (ALS) と大きな違いがないようにも思えるが，E ステップにおいて各時刻の状態 $\boldsymbol{x}_t$ を点推定ではなく多次元ガウス分布として求め，M ステップにおいて尤度の状態に関する期待値を最大化している点で一手間掛かっている．

EM アルゴリズムによる線形状態空間モデルの最尤推定は，モデルパラメータ $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{Q}, \boldsymbol{R}$ に対して（共役な）事前分布を導入することによって容易に最大事後確率 (MAP) 推定に拡張することができる．また，変分近似やマルコフ連鎖モンテカルロ法を用いることによって，モデルパラメータの事後分布をベイズ推論することも可能である[3]．また，これらの学習法は非ガウス的な観測モデルをもつ線形状態空間モデルの学習にも拡張可能である．たとえば，文献 4)〜6) などでは，神経細胞の発火回数遷移をポワソン分布線形状態空間モデルでモデル化し，EM アルゴリズムや変分ベイズ EM アルゴリズム（VB-EM アルゴリズム）でモデルパラメータを学習している．

2.3　スイッチング動的システムの学習

線形状態空間モデルの拡張方法として，複数（有限個）の異なる局所線形モデル間を確率的に遷移するスイッチング線形動的システム (switching linear dynamical systems; SLDS) がある．これは，グラフ表現で表わすと，図 1 (c) のように，通常の SSM のような連続的な潜在状態ベクトル $\boldsymbol{x}_t$ と，HMM のような離散的な値を取る潜在変数 s_t をもつ形となっており，HMM と SSM のハイブリッドともいえる[注1]．

SLDS のモデル学習についても，EM アルゴリズムに基づく最尤法の考え方が適用できる[8),9)]．すなわち，E ステップにおいてはモデルパラメータを固定して潜在変数である $\boldsymbol{x}_t$ と s_t の同時分布 $p(\boldsymbol{x}_{1:T}, s_{1:T})$ を求め，M ステップにおいては求めた潜在変数の分布に関するモデル尤度の期待値を最大化すれば良い．しかし，実際には，K 個の局所線形モデルをもつ SLDS では状態に関するフィルタ分布 $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ が K^t 個のコンポーネントをもつ混合ガウス分布になるため，厳密に求めることは困難である．そこで，一般化擬似ベイズ法 (generalized pseudo Bayesian; GPB)[8]，ガウシアンサムフィルタ (Gaussian sum filter; GSF)，変分近似[7),10),11)]，MCMC などによって近似的に潜在変数 $\boldsymbol{x}_t$ と s_t の同時分布を求める方法が用いられている．

SLDS は，複数の動作モード（レジーム）を遷移するような動的システムのモデル化や人間の動作・行動認識[7),11)] など応用範囲が広いが，局所モデルの数をいかにして決定すべきかという問題がある．これは，一般的なクラスタリング問題においてクラスタ数を決定する問題と似ており，階層的なディリクレ事前分布を用いたノンパラメ

[注1] 図 1 (c) は，状態遷移モデルと観測モデルの両方が離散状態変数 s_t の値によって切り替わるケースを表わしているが，片方のモデルのみ切り替わるもの（たとえば文献 7)）や，s_t の遷移が $\boldsymbol{x}_t$ に依存するもの（augmented SLDS などと呼ばれる）などさまざまな亜種が考えられる．

トリックベイズによるモデル数の決定法[12] などが提案されている．また，図1からも明らかなように，スイッチング動的システムはより複雑な構造をもった潜在変数・状態空間モデル（たとえば，潜在変数が複数のグループに分かれるものなど）に拡張が可能である．実際，これらのモデルは動的ベイジアンネットワーク (dynamic Bayesian networks; DBN) として一般化され，その推論・学習法が研究されている[13] (注2)．

2.4 非線形・ガウス過程状態空間モデルの学習

最尤法とEMアルゴリズムによる状態空間モデルの学習は，状態・観測モデルが非線形な場合にも拡張が容易である．すなわち，Eステップで拡張カルマンスムーザ (EKS) や連続モンテカルロ・粒子フィルタを用いて状態ベクトル列 $\boldsymbol{x}_{1:T}$ の事後分布を近似的に求め，Mステップでは尤度期待値を非線形モデルに含まれるパラメータについて最大化すれば良い．そのような考えに基づく初期の例としては，Ghahramani と Roweis による研究[14] が挙げられる．この研究では，状態・観測モデルともに放射基底関数ネットワーク (radial basis function networks; RBFN) が採用され，Eステップでは拡張カルマンスムーザが用いられている．また，2000年代に多くの線形学習手法が**正定値カーネル** (positive definite kernel) によって非線形化されたように，元の空間では非線形な動的システムを正定値カーネルによって無限次元の非線形特徴空間に写像し，特徴空間においてEMアルゴリズムによって線形動的システムを学習する方法も提案されている[15],[16]．

ほかにもさまざまな非線形関数近似器を状態・観測モデルに利用することが可能であるが，現在では特に，ガウス過程回帰 (Gaussian process; GP) が頻繁に用いられている．これは，教師あり非線形回帰学習手法としてのGPの性能の高さに加えて，GPが確率的モデルであること，すなわち，入力を $\boldsymbol{x}$，出力を y としたときに条件付き確率分布 $p(y|\boldsymbol{x})$ を自然に求められる長所をもつためと考えられる．最初は，潜在状態列 $\boldsymbol{x}_{1:T}$ が直接観測可能であるという強い仮定のもと，すなわち，教師あり学習の設定下で，GP-Bayesfilter[17] やGP-ADF[18] が発表されている．これらでは，状態モデル f と観測モデル g の推定をそれぞれ教師あり回帰学習問題として通常のガウス過程回帰問題で解いている．その後，状態列も観測から推定しなければならないという教師なし学習の条件下で，Wang らによってガウス過程動的モデル (Gaussian process dynamical model; GPDM) が提案されている[19]．これは，その前に静的な非線形潜在変数モデル・確率的非線形次元削減法として提案されていたガウス過程潜在変数モデル (Gaussian process latent variable model; GPLVM)[20] を動的システム学習に発展させたものである．ただし，GPDMでは通常のガウス過程回帰と同様，訓練データサンプルをすべて保存しておく必要があった．その後，これを少数の仮想サンプルによって近似する方法[21] や変分近似する方法[22],[23] が提案されている．

このように，最尤法とEMアルゴリズムに基づく動的システム学習は，連続的・離散的な状態ベクトルが混在するシステムにも利用でき，ガウス過程回帰のような非線形モデルとも組み合わせることができるという点で汎用性・拡張性が非常に高い．その一方，最も基本的な線形状態空間モデルですら尤度関数すなわち目的関数は多峰性を有しており，適切な初期解を与えないと不適切な局所解に陥ってしまうという問題がある．

3. スペクトラル動的システム学習

3.1 時系列データの次元削減

動的システム学習そのものではないが，機械学習・データマイニングで盛んに研究されてきたテーマの1つに（多次元）時系列データの変化点・異常検知問題がある．この分野では，井手ら[24] が特異スペクトル解析を発展させた手法を提案しており広く知られている．その基本的な考え方は，多次元時系列データから固定長のスライド窓によって部分時系列を切り出して高次元ベクトルに変換し，特異値分解あるいは主成分分析を適用することによって低次元の部分空間を同定するというものであり，明らかに，部分空間同定法や動的PCAなどとも関連している．

3.2 正準相関分析による動的システム学習

制御学ではよく知られているように，確率過程の過去と未来との正準相関分析によって状態列を得る方法は，赤池によって研究[25] され，その後，正準相関分析による部分空間同定法[26],[27] として発展してきた．しかし，意外にも機械学習のコミュニティでは部分空間同定法自体が2000年代後半まであまり知られていなかった．非反復的なアルゴリズムによって大局解を求められる部分空間同定法は，動的システム学習にとっても非常に魅力的であり，河原[28] は，正準相関分析を用いた部分空間同定法を，正定値カーネルで定まる再生核ヒルベルト空間上で展開することで，非線形状態空間モデルを推定可能なアルゴリズムを導出した．また，河原は，前節の時系列変化点検知問題においても，過去と未来の部分時系列の正準相関分析によって得られた部分空間を用いる方法[29] を提案している．その後，上甲ら[30] はこの考え方を応用し，混合確率的正準相関分析 (Mixture of Probabilisic CCA; MPCCA) を用いることによって，線形動的システムの確率混合モデルを学習する方法を提案している．筆者の知る限り，これらの研究が部分空間同定法を機械学習の立場から発展させた先駆的研究であり，現在知られているスペクトラル動的システム学習の元祖であると認識している．しかし，動的システム学習の研究者らからは一

(注2) 名前からも明らかであるが，DBN もまたベイジアンネットワークの特別な場合である．

定の評価を受けたものの，機械学習全体から注目を受けるには至らなかった．

3.3 隠れマルコフモデルのスペクトラル学習

動的システム学習においてスペクトラル学習という言葉が使われ始めたのは，Hsu ら[31] が隠れマルコフモデル (HMM) のための非反復的なアルゴリズムを発表してからである．Hsu らの方法は，過去の観測と未来の観測との正準相関分析によって離散的な状態列を求めるという部分空間同定法のアイデアを利用している．この手法は「遷移確率行列と観測確率行列のランクが隠れ状態数に等しく即時出力から隠れ状態を推定することが可能 (1-step observability).」という強い仮定のもとでの HMM の学習アルゴリズムを示したものではあったが，従来，ほとんどの機械学習研究者は，「HMM は EM アルゴリズム (Baum-Welch アルゴリズム）によって学習するもの」と信じ込んでいたので，非反復的なスペクトラル学習アルゴリズムによって大局解を得られるという発見は十分に衝撃的であった．その後，多くの研究者たちによって，前提条件の緩和[32] や一般化[33]，より複雑なモデルへの拡張[34)~37] が行われている．

また,「車輪の再発明」的な現象にも思えるが，HMM のスペクトラル学習の成功に触発されて，動的システム学習では，連続的な状態空間モデルのスペクトラル学習，すなわち，部分空間同定法にも再び注目が寄せられる状況になっている．たとえば，文献 38) はポワソン分布出力モデルをもつ連続的な状態空間モデルのスペクトラル学習を提案している．

3.4 予測状態表現

機械学習の主要分野の 1 つである強化学習では，意思決定主体（エージェント）の離散的状態空間および遷移と観測をモデル化するのに部分観測マルコフ決定過程 (partially observable Markov decision process; POMDP) を用いるのが主流であるが，POMDP とは違うアプローチとして予測状態表現 (predictive state representation; PSR) があり，スペクトラル動的システム学習および部分空間同定法との関連性が知られている．PSR は，HMM の一般化として Jaeger により提案された観測オペレータモデル (observable operator models; OOM)[39] を Littman[40] や Singh[41] らがさらに発展させたものであり，動的システムへの入力と出力からシステムの潜在状態についての確率分布すなわち**信念** (belief) を陽に求める代わりに，将来の入力に対する出力（「テスト」と呼ぶ）の予測の集合によってシステムの状態を表現するという考えに基づいている．特に，Rosencrantz ら[42] は，大量のテスト集合を特異値分解によってコンパクトな集合に変換する変形予測状態表現 (transformed PSR; TPSR) を提案しており，線形動的システムに対する部分空間同定法との関連性が指摘されている[43]．

4. 深層ニューラルネットワークと動的システム学習

4.1 ニューラルネットワークと動的システム学習

2.4 節において RBFN を用いた非線形状態空間モデルを EM アルゴリズムによって学習する Roweis らの研究を紹介したが，その後は正定値カーネルやガウス過程回帰を用いた非線形化が主流となり，人工ニューラルネットワークと状態空間モデルを組み合わせた動的システム学習法は，多層パーセプトロン (multilayer perceptron; MLP) を用いた[44] など少数の例に限られていた．大規模なニューラルネットワークの学習の難しさ，および，潜在状態事後分布 $p(\boldsymbol{x}|\boldsymbol{y})$ の推定の困難さがその最大の理由である．しかし，2010 年代になって深層学習がまず教師あり学習問題で華々しい成果を上げると，教師なし学習，さらには動的システム学習への応用もつぎつぎと報告されるようになった．

4.2 変分オートエンコーダの動的システム学習への応用

動的システム学習にも深層ニューラルネットワークが用いられるようになった大きなきっかけは変分オートエンコーダ (variational auto-encoder; VAE)[45] の出現である．VAE の本来の目的は，与えられた高次元観測データ $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N$ に対して潜在変数モデル，すなわち，$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$ を深層ニューラルネットワークの高い表現力を利用して学習することである．2.1 節で述べたように，伝統的な機械学習では，このような潜在変数モデルは，ベイズ定理を用いて潜在変数 $\boldsymbol{x}$ についての事後分布を求め，その分布を用いて対数尤度の期待値を最大化する EM アルゴリズム（あるいはその変種）を用いるのが定石であった．しかし，潜在変数 $\boldsymbol{x}$ から観測変数 $\boldsymbol{y}$ が生成される過程を複雑な深層ニューラルネットワークでモデル化した場合，事後分布 $p(\boldsymbol{x}|\boldsymbol{y})$ を求めることはきわめて困難である．そこで VAE では潜在変数事後分布 $p(\boldsymbol{x}|\boldsymbol{y})$ を別のニューラルネットワーク（エンコーダネットワーク）によって近似し，対数尤度の変分下界を誤差逆伝搬法と確率勾配法によって最大化することによって，2 つのネットワークを同時に学習することでこの問題を解決している．

前述のように，状態空間モデルは潜在変数モデルにダイナミクスを追加したものであるから，深層ニューラルネットワークによって状態遷移･観測モデルを表現した状態空間モデルに対しても同様の技法を利用するのは自然な考えといえる．Krishnan ら[46] はこのような考えに基づき，非線形状態空間モデルと観測列から状態列の事後分布を求める "推論ネットワーク" (inference network) を同時に学習する枠組みを提案している[注3]．この手法

[注3] Krishnan らは，当初この手法を deep Kalman filter と呼んでいたが，その後，deep Markov model と改称している．

で特筆すべきなのは，推論ネットワークは言わば状態推定器（フィルタ／スムーザ）であり，動的システムの生成モデルと同時に学習される点である．Karl ら[48] もこれと非常に似た方法で VAE を拡張した非線形状態空間モデルを提案しているが，潜在変数列 $\boldsymbol{x}_{1:T}$ の代わりに遷移パラメータ列 $\boldsymbol{\beta}_{1:T}$ をエンコーダネットワークによって推定している点が異なる．ほかにも Watter ら[49] は局所線形な状態遷移モデルと VAE による非線形観測モデルを組み合わせた状態空間モデル，Fraccaro ら[50] は時変の線形遷移モデルと VAE に基づく非線形の観測モデルを組み合わせた状態空間モデルを提案している．

深層学習を非線形状態空間モデルに組み合わせた研究は出現してからまだ日が浅いが，たとえば動画像から動的システムのモデルを学習し，新たな動画像データを生成するなど，魅力的な応用が期待される．

5. むすび

本稿では，機械学習分野における動的システム学習の概要と動向を3つのカテゴリーに分けて紹介した．近年，制御学においても機械学習への関心が高まっていると聞いている．また，機械学習においても動的なシステムを陽に扱う必要性が増しており，制御学のさまざまな方法論や技法を学ぶ重要性を感じている．今後，両分野の交流がますます活発になり，相互の発展に寄与し合うことを願う次第である．

（2018 年 12 月 22 日受付）

参 考 文 献

1） Z. Ghahramani and G. Hinton: Parameter Estimation for Linear Dynamical Systems, Technical Report CRG-TR-96-2, Dept. Comp. Sci., Univ. Toronto (1996)

2） C.M. Bishop: *Pattern Recognition and Machine Learning*, Springer (2006)

3） D. Barber and S. Chiappa: Unified Inference for Variational Bayesian Linear Gaussian State-Space Models, *NIPS 19* (2006)

4） J. Macke, et al.: Empirical Models of Spiking in Neural Populations, *NIPS* (2011)

5） M. Park, et al.: Unlocking Neural Population Non-Stationarities Using Hierarchical Dynamics Models, *NIPS* (2015)

6） Y. Gao, et al.: High-Dimensional Neural Spike Train Analysis with Generalized Count Linear Dynamical Systems, *NIPS* (2015)

7） V. P. Pavlovic, et al.: Learning Switching Linear Models of Human Motion, *NIPS* (2000)

8） K. P. Murphy: Switching Kalman Filters, Technical Report, DEC/Compaq Cambridge Research Labs (1998)

9） Z. Ghahramani and G. E. Hinton: Variational Learning for Switching State-Space Models, *Neural Computation*, **12**–4, 831/864 (2000)

10） S. M. Oh, et al.: A Variational inference method for Switching Linear Dynamic Systems, Technical Report, GIT-GVU-05-16 (2005)

11） S. Chiappa, et al.: Using Bayesian Dynamical Systems for Motion Template Libraries, *NIPS* (2008)

12） E. Fox, et al.: Nonparametric Bayesian learning of switching linear dynamical systems, *NIPS* (2008)

13） D. Koller and N. Friedman: *Probabilistic Graphical Models*, MIT Press (2009)

14） Z. Ghahramani and S. T. Roweis: Learning Nonlinear Dynamical Systems Using an EM Algorithm, *NIPS* (1999)

15） L. Ralaivola and F. D'alche-Buc: Time Series Filtering, Smoothing and Learning Using the Kernel Kalman Filter, *Proceedings of the IEEE International Joint Conference on Neural Networks*, **3**, 1449/1454 (2005)

16） P. Zhu, et al.: Learning Nonlinear Generative Models of Time Series With a Kalman Filter in RKHS, *IEEE Transactions on Signal Processing*, **62**–1, 141/155 (2014)

17） J. Ko and D. Fox: GP-BayesFilters: Bayesian Filtering Using Gaussian Process Prediction and Observation Models, *IROS* (2008)

18） M. P. Deisenroth, et al.: Analytic Moment-Based Gaussian Process Filtering, *ICML* (2009)

19） J. M. Wang, et al.: Gaussian Process Dynamical Models for Human Motion, *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 283/298 (2008)

20） N. D. Lawrence: Gaussian Process Latent Variable Models for Visualisation of High Dimensional Data, *NIPS* (2003)

21） R. Turner, et al.: State-Space Inference and Learning with Gaussian Processes, *AISTATS* (2010)

22） A. Damianou, et al.: Variational Gaussian Process Dynamical Systems, *NIPS* (2011)

23） R. Frigola, et al.: Variational Gaussian Process State-Space Models, *NIPS* (2013)

24） T. Idé and K. Inoue: Knowledge Discovery from Heterogeneous Dynamic Systems using Change-Point Correlations, *Proceedings of the 2005 SIAM International Conference on Data Mining*, 571/575 (2005)

25） H. Akaike: Markovian Representation of Stochastic Processes by Canonical Variables, *SIAM Journal on Control*, **13**–1, 162/173 (1975)

26） W. E. Larimore: Canonical Variate Analysis for System Identification, Filtering, and Adaptive Control, *Proc. 29th IEEE Conference on Decision and Control*, **2**, 596/604 (1990)

27） T. Katayama and G. Picci: Realization of Stochastic Systems with Exogenous Inputs and Subspace System Identification Methods, *Automatica*, **35**–10, 1635/1652 (1999)

28） Y. Kawahara, et al.: A Kernel Subspace Method by Stochastic Realization for Learning Nonlinear Dynamical Systems, *NIPS* (2006)

29） Y. Kawahara, et al.: Change-Point Detection in Time-Series Data Based on Subspace Identification, *ICDM*, 559/564 (2007)

30） M. Joko, et al.: Learning Non-linear Dynamical Systems by Alignment of Local Linear Models, *20th International Conference on Pattern Recognition* (*ICPR*), 1084/1087 (2010)

31） D. J. Hsu, et. al.: A Spectral Algorithm for Learning Hidden Markov Models, *COLT* (2009)

32） S. M. Siddiqi, et al.: Reduced-Rank Hidden Markov Models, *AISTATS* (2010)

33） A. Anandkumar, et al.: A Method of Moments for Mixture Models and Hidden Markov Models, *COLT* (2012)

34） L. Song, et al.: Hilbert Space Embeddings of Hidden Markov Models, *ICML* (2010)

35） C. Subakan, et al.: Spectral Learning of Mixture of Hidden Markov Models, *NIPS* (2014)

36） C. Zhang, et al.: Spectral Learning of Large Structured HMMs for Comparative Epigenomics, *NIPS* (2015)

37） K. Kandasamy, et al.: Learning HMMs with Nonparametric Emissions via Spectral Decompositions of Continuous Matrices, *NIPS* (2015)

38) L. Buesing, et al.: Spectral Learning of Linear Dynamics from Generalised-Linear Observations with Application to Neural Population Data, *NIPS* (2012)
39) H. Jaeger: Observable Operator Models for Discrete Stochastic Time Series, *Neural Comput.*, **12**–6, 1371/1398 (2000)
40) M. L. Littman et al.: Predictive Representations of State, *NIPS* (2001)
41) S. Singh, et al.: Predictive State Representations: A New Theory for Modeling Dynamical Systems, *UAI* (2004)
42) M. Rosencrantz and G. Gordon: Learning Low Dimensional Predictive Representations, *ICML* (2004)
43) B. Boots, et al.: Closing the Learning-Planning Loop with Predictive State Representations, *Robotics: Science and Systems VI* (2009)
44) H. Valpola and J. Karhunen: An Unsupervised Ensemble Learning Method for Nonlinear Dynamic State-Space Models, *Neural Computation*, **14**–11, 2647/2692 (2002)
45) D. P. Kingma and M. Welling: Auto-Encoding Variational Bayes, *ICLR* (2014)
46) R. G. Krishnan, et al.: Deep Kalman Filters, In: arXiv preprint arXiv:1511.05121 (2015)
47) R. G. Krishnan, et al.: Structured Inference Networks for Nonlinear State Space Models, *AAAI* (2017)
48) M. Karl, et al.: Deep Variational Bayes Filters: Unsupervised Learning of State Space Models from Raw Data, *ICLR* (2017)
49) M. Watter, et al.: Embed to Control: A Locally Linear Latent Dynamics Model for Control from Raw Images, *NIPS* (2015)
50) M. Fraccaro, et al.: A Disentangled Recognition and Nonlinear Dynamics Model for Unsupervised Learning, *NIPS* (2017)

［著　者　紹　介］

矢入（やいり）健久（たけひさ）君

1971 年生．1994 年東京大学工学部航空学科卒．1996 年，1999 年同大学大学院工学系研究科修士課程，博士課程修了．博士（工学）．同大学先端科学技術研究センター 助手，講師，准教授を経て，現在は同大学工学系研究科准教授．専門は機械学習・確率的推論とその航空宇宙分野等への応用．人工知能学会（2014〜2015 年理事），航空宇宙学会，ロボット学会，IEEE の各会員．