システム同定のための
機械学習技術の最近の発展について

IBM東京基礎研究所 数理科学担当
担当部長 井手 剛

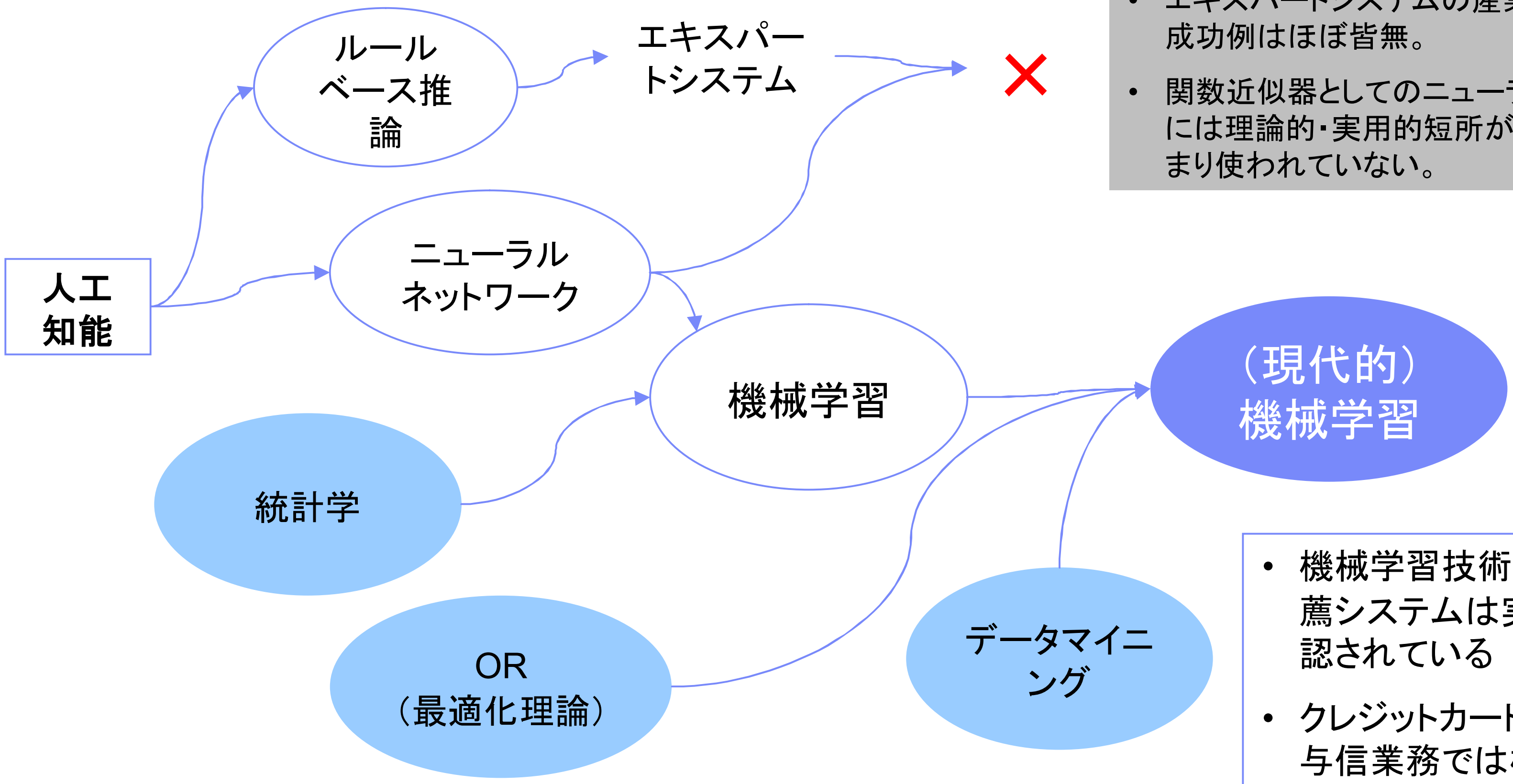
目次

- **最近の機械学習の進歩**
- **Volterra展開とカーネルトリック**
- **その他の関連する話題**
- **IBM東京基礎研究所での取り組み**
  – 無線位置推定
  – 変化点検出

目次

- **最近の機械学習の進歩**
- **Volterra展開とカーネルトリック**
- **その他の関連する話題**
- **IBM東京基礎研究所での取り組み**
  - 無線位置推定
  - 変化点検出

ここ10年の機械学習技術の急速な発展に伴い、その産業応用が進展しています



- エキスパートシステムの産業上での成功例はほぼ皆無。
- 関数近似器としてのニューラルネットには理論的･実用的短所があり、あまり使われていない。

- 機械学習技術に基づくWeb検索や推薦システムは実用化され、有用性が確認されている
- クレジットカードの不正検知や銀行の与信業務では機械学習は必須の技術
- 統計的プロセス制御にも機械学習の応用が進出している

自動車産業でデータ解析（機械学習）技術がお役に立てる領域

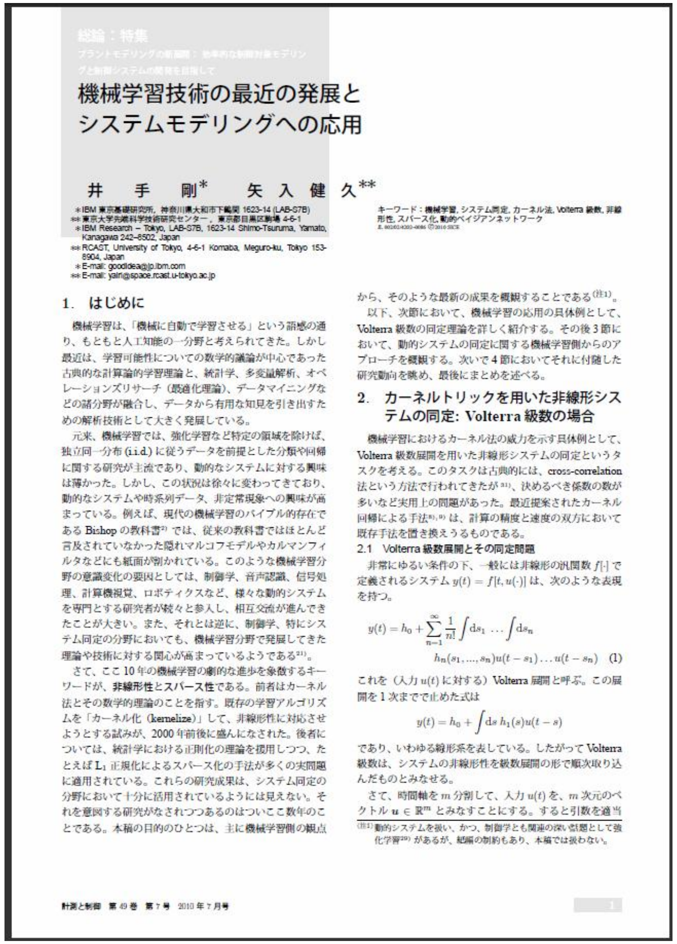
設計

異常解析

適合

とりわけ「適合」の分野では、機械学習技術の適用に期待が持たれています: 計測自動制御学会誌・特集企画「プラントモデリングの新展開」より

- **プラントモデリングの新展開 — 効率的な制御対象モデリングと制御システムの開発を目指して**
  - 解説 No.9「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」
    - IBM東京基礎研究所 井手剛, 東京大学先端科学技術研究センター 矢入健久
      - ✓ 機械学習は、最近の情報技術の発展を象徴する研究分野のひとつである。本稿では、この10年の機械学習技術の急速な進歩を象徴するキーワードとして、「非線形性」と「スパース性」を取り上げ、これらを軸に関連する手法を概観する。同時に、機械学習技術のシステムモニタリングへの応用について最近の研究事例を紹介する。



総論：特集
プラントモデリングの新展開：効率的な制御対象モデリングと制御システムの開発を目指して

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

井手 剛* 矢入健久**

*IBM 東京基礎研究所，神奈川県大和市下鶴間 1623-14 (LAB-S7B)
**東京大学先端科学技術研究センター，東京都目黒区駒場 4-6-1
*IBM Research – Tokyo, LAB-S7B, 1623-14 Shimo-Tsuruma, Yamato, Kanagawa 242–8502, Japan
**RCAST, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan
*E-mail: goodidea@jp.ibm.com
**E-mail: yairi@space.rcast.u-tokyo.ac.jp

キーワード：機械学習, システム同定, カーネル法, Volterra 級数, 非線形性, スパース化, 動的ベイジアンネットワーク

1. はじめに

機械学習は、「機械に自動で学習させる」という語感の通り、もともと人工知能の一分野と考えられてきた。しかし最近は、学習可能性についての数学的議論が中心であった古典的な計算論的学習理論と、統計学、多変量解析、オペレーションズリサーチ（最適化理論）、データマイニングなどの諸分野が融合し、データから有用な知見を引き出すための解析技術として大きく発展している。

元来、機械学習では、強化学習など特定の領域を除けば、独立同一分布 (i.i.d) に従うデータを前提とした分類や回帰に関する研究が主流であり、動的なシステムに対する興味は薄かった。しかし、この状況は徐々に変わってきており、動的なシステムや時系列データ、非定常現象への興味が高まっている。例えば、現代の機械学習のバイブル的存在である Bishop の教科書[2] では、従来の教科書ではほとんど言及されていなかった隠れマルコフモデルやカルマンフィルタなどにも紙面が割かれている。このような機械学習分野の意識変化の要因としては、制御学、音声認識、信号処理、計算機視覚、ロボティクスなど、様々な動的システムを専門とする研究者が続々と参入し、相互交流が進んできたことが大きい。また、それとは逆に、制御学、特にシステム同定の分野においても、機械学習分野で発展してきた理論や技術に対する関心が高まっているようである[21]。

さて、ここ 10 年の機械学習の劇的な進歩を象徴するキーワードが、**非線形性**と**スパース性**である。前者はカーネル法とその数学的理論のことを指す。既存の学習アルゴリズムを「カーネル化 (kernelize)」して、非線形性に対応させようとする試みが、2000 年前後に盛んになされた。後者については、統計学における正則化の理論を援用しつつ、たとえば L_1 正規化によるスパース化の手法が多くの実問題に適用されている。これらの研究成果は、システム同定の分野において十分に活用されているようには見えない。それを意図する研究がなされつつあるのはついここ数年のことである。本稿の目的のひとつは、主に機械学習側の観点から、そのような最新の成果を概観することである(注1)。

以下、次節において、機械学習の応用の具体例として、Volterra 級数の同定理論を詳しく紹介する。その後 3 節において、動的システムの同定に関する機械学習側からのアプローチを概観する。次いで 4 節においてそれに付随した研究動向を眺め、最後にまとめを述べる。

2. カーネルトリックを用いた非線形システムの同定: Volterra 級数の場合

機械学習におけるカーネル法の威力を示す具体例として、Volterra 級数展開を用いた非線形システムの同定というタスクを考える。このタスクは古典的には、cross-correlation 法という方法で行われてきたが[21]、決めるべき係数の数が多いなど実用上の問題があった。最近提案されたカーネル回帰による手法[8],[9] は、計算の精度と速度の双方において既存手法を置き換えうるものである。

2.1 Volterra 級数展開とその同定問題

非常にゆるい条件の下、一般には非線形の汎関数 $f[\cdot]$ で定義されるシステム $y(t) = f[t, u(\cdot)]$ は、次のような表現を持つ。

$$y(t) = h_0 + \sum_{n=1}^{\infty} \frac{1}{n!} \int \mathrm{d}s_1 \ldots \int \mathrm{d}s_n \; h_n(s_1, \ldots, s_n) u(t - s_1) \ldots u(t - s_n) \quad (1)$$

これを（入力 $u(t)$ に対する）Volterra 展開と呼ぶ。この展開を 1 次までで止めた式は

$$y(t) = h_0 + \int \mathrm{d}s \, h_1(s) u(t - s)$$

であり、いわゆる線形系を表している。したがって Volterra 級数は、システムの非線形性を級数展開の形で順次取り込んだものとみなせる。

さて、時間軸を m 分割して、入力 $u(t)$ を、m 次元のベクトル $\boldsymbol{u} \in \mathbb{R}^m$ とみなすことにする。すると引数を適当

(注1) 動的システムを扱い、かつ、制御学とも関連の深い話題として強化学習[20] があるが、紙幅の制約もあり、本稿では扱わない。

計測と制御 第 49 巻 第 7 号 2010 年 7 月号

（ご参考）現代の機械学習の2つのキーワード: 非線形性とスパース性

今日の話の中心

非線形性

- **入力と出力の間の非線形的な関係を柔軟かつ簡単に取り込む**
- **技術的キーワード**
  – カーネル法

スパース性

- **入力と出力の間の関係を、より少ないパラメターや変数で表す**
  – 例えば、100個変数があっても、本当は予測には8個しか寄与しないかもしれない
  – なるべく疎なモデルを作りたい場合は実用上多い
- **技術的キーワード**
  – L1正則化、Lasso
  – 関連度自動決定（ARD: automated relevance determination）

目次

- 最近の機械学習の進歩
- Volterra展開とカーネルトリック
- その他の関連する話題
- IBM東京基礎研究所での取り組み
  – 無線位置推定
  – 変化点検出

出発点: 動的システムをボルテラ級数で表す

- 非常に一般的な条件で、非線形システムは次の表現を持つ（ボルテラ展開）
  –
$$y(t) = h_0 + \sum_{n=1}^{\infty} \frac{1}{n!} \int \mathrm{d}s_1 \ \ldots \int \mathrm{d}s_n \ h_n(s_1, ..., s_n) u(t - s_1) \ldots u(t - s_n)$$
  – これは線形系を特別な場合として含む（展開を1次で打ち切ればよい）
$$y(t) = h_0 + \int \mathrm{d}s \ h_1(s) u(t - s)$$
- ここでは次の離散ボルテラ展開を想定し、パラメター η の推定方法を考える
$$y = h^0 + \sum_{n=1}^{\infty} \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} h^n_{i_1,\ldots,i_n} u_{i_1} \cdots u_{i_n}$$

モデルの可読性は著しく低いが、
極めて一般的な表現。

ボルテラ表現におけるシステム同定問題: 基本的に困難な問題である

- 次のデータDが与えられた時にボルテラ係数 $h^0, \{h^1_{i_1}\}, \ldots, \{h^n_{i_1,\ldots,i_n}\}, \ldots$ を求めたい

$$\mathcal{D} \equiv \left\{ \left(\boldsymbol{u}^{(t)}, y^{(t)}\right) \middle| \boldsymbol{u}^{(t)} \in \mathbb{R}^m, y^{(t)} \in \mathbb{R}, t = 1, 2, \ldots, N \right\}$$

$$y = h^0 + \sum_{n=1}^{\infty} \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} h^n_{i_1,\ldots,i_n} u_{i_1} \cdots u_{i_n}$$

無限個のパラメターを定める必要がある ＝ 基本的に不可能

IBM

ボルテラ表現におけるシステム同定問題：
それでも解くにはどうするか

従来の考え方	機械学習の考え方
無限個は扱えないので、有限個の項のみをがんばって同定する。 通常、精度は今ひとつで、「どこまで項を取り込めばいいのか」と疑心暗鬼になりやすい	無限個は扱えないので、双対な問題に飛ぶ。双対問題において、無限が出てこない！ （いわゆるカーネルトリック） 仮に精度がよくなくても、それは「今のデータでできる最善のこと」なので、あきらめがつく

ボルテラ表現におけるシステム同定問題：カーネル法による解

- **異なる時刻の入力ベクトル同士に「カーネル関数」kを定義しておく**
  - これは内積の一般化と考えてよい
  - 与えられたN個のサンプルに対して、N×N行列が定義でき、それをカーネル行列と呼ぶ

$$\mathsf{K} \equiv \left(k(\boldsymbol{u}^{(i)}, \boldsymbol{u}^{(j)})\right)$$

- **このカーネル行列を使うと、ボルテラ級数が（有限項近似に訴えることなく）、次のように同定できる**

$$y = \boldsymbol{k}(\boldsymbol{u})^\top \left[\mathsf{K} + \lambda N \mathsf{I}_N\right]^{-1} \boldsymbol{y}_N$$

$$\boldsymbol{k}(\boldsymbol{u}) \equiv \left[k(\boldsymbol{u}, \boldsymbol{u}^{(1)}), k(\boldsymbol{u}, \boldsymbol{u}^{(2)}), ..., k(\boldsymbol{u}, \boldsymbol{u}^{(N)})\right]^\top \in \mathbb{R}^N$$

$$\boldsymbol{y}_N \equiv \left[y^{(1)}, y^{(2)}, ..., y^{(N)}\right]^\top$$

- N はサンプルの数
- I_N はN次元単位行列

最終結果のどこにも無限次元が現れていない

- **システム同定が、できた**
  –
  $$y = k(u)^{\top}\left[\mathsf{K} + \lambda N \mathsf{I}_N\right]^{-1} y_N$$
  – データの数Nのサイズの行列計算が必要だが、無限次元はどこにも出てきていない
  – Nの長さの量を扱うことで、無限個のボルテラ級数が暗に定まったことになる

- **無限次元をカーネル行列の中に繰り込む手法を「カーネルトリック」と呼ぶ**

- **これにより一般性は失われない ――「レプレゼンター定理」**

- **無限を繰り込んだ代償＝同定されたモデルは完全なブラックボックスに**
  – ボルテラ級数のどの項がどう効いているかはまったく分からない

目次

- **最近の機械学習の進歩**
- **Volterra展開とカーネルトリック**
- **その他の関連する話題**
- **IBM東京基礎研究所での取り組み**
  - 無線位置推定
  - 変化点検出

機械学習と制御工学のスタンスの相違

	機械学習 （動的システム学習）	システム同定 （部分空間同定法）
共通問題	観測列 $Y_{1:T}$ と入力列 $U_{1:T}$ から、状態列 $X_{1:T}$ および、状態遷移モデル f , 観測モデル g を推定	
対象系	非線形一般、高次元出力、入力無し	線形が基本（非線形拡張あり）、入力あり
興味	観測の次元削減（可視化）、学習アルゴリズム	ダイナミクス、制御、系自体の性質
モデル化	離散・連続変数区別無し（DBN）	基本的に連続変数（Hybrid系もあり）
手法	EM algorithm による状態・モデル同時推定	SVD, QR等で状態、モデルの一方を先に推定

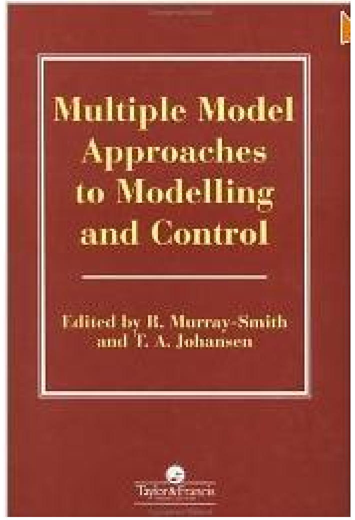
矢入健久, "次元削減と動的システムの学習について,**" Latent Dynamics Workshop, 2010.**

[非線形システム同定]
スイッチング線形動的モデル

- **制御工学でも広く知られたモデル**
  – 局所的に線形モデルをフィット。局所モデルの貼り合わせで全体を表現

- **EM法でシステム同定を行うのが定石**
  – Expectation ステップ
    - 次の二つの変数についての事後分布を求める
      - ✓状態変数
      - ✓「どのモデルを使うか」という変数
  – Maximizationステップ
    - 選択された局所モデルの尤度を最大化

- **どのくらいの数の局所モデルを使うのがベストかを知るのが極めて難しい**
  – 最先端の研究課題



Multiple Model Approaches to Modeling and Control, R Murray-Smith (Editor), T. Johansen (Editor), CRC Press (1997)

[非線形システム同定]
基底展開モデル

- **観測モデル、遷移モデル共に非線形関数を想定するが、その関数形を、基底展開の形に表しておく**

$$
\begin{aligned}
\mathbf{x}_t &= f(\mathbf{x}_{t-1};\mathbf{A}) + \mathbf{n}_{x,t} \\
\mathbf{y}_t &= g(\mathbf{x}_t;\mathbf{B}) + \mathbf{n}_{y,t}
\end{aligned}
\quad \longrightarrow \quad
\begin{aligned}
f(\mathbf{x};\mathbf{A}) &= \sum_i \mathbf{a}_i \, \phi_i(\mathbf{x}) \\
g(\mathbf{x};\mathbf{B}) &= \sum_j \mathbf{b}_j \, \psi_j(\mathbf{x})
\end{aligned}
$$

- **いくつかのアプローチがある**
  - 明示的に基底関数を与えるもの
  - 既定関数をカーネル関数の中で陰に処理するもの
    - 右: Wang et al., Gaussian Process Dynamical Models, NIPS 2005

- **システム同定のための計算が、やや複雑な最適化問題になり、計算量的な問題が生じがちである**

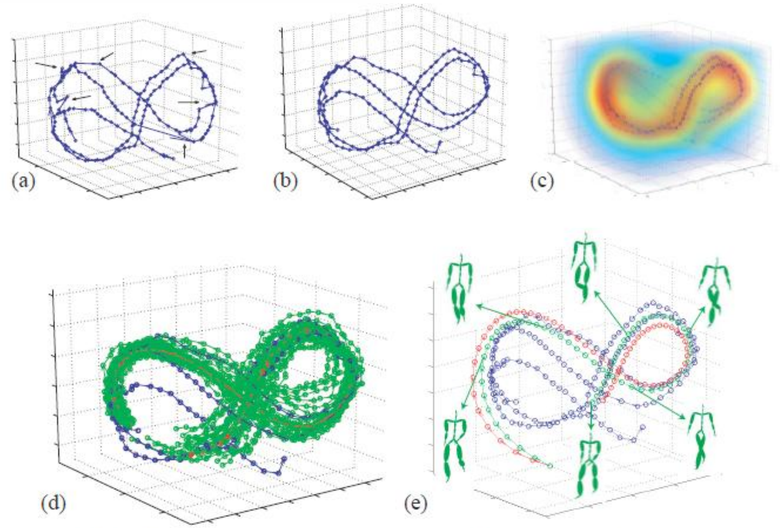


Figure 2: Models learned from a walking sequence of 2.5 gait cycles. The latent positions learned with a GPLVM (a) and a GPDM (b) are shown in blue. Vectors depict the temporal sequence. (c) - log variance for reconstruction shows regions of latent space that are reconstructed with high confidence. (d) Random trajectories drawn from the model using HMC (green), and their mean (red). (e) A GPDM of walk data learned with RBF+linear kernel dynamics. The simulation (red) was started far from the training data, and then optimized (green). The poses were reconstructed from points on the optimized trajectory.

[計算の高速化]
データの「間引き」技術が盛んに研究されている

- **カーネル法では、一般にデータ点数Nの2乗から3乗の計算量がかかり、データ点数が多いと計算が難しい**

- **そのため、効率よくデータを間引く工夫についてよく調べられている**
  - 代表的手法: Nyström法
    - Fixed-size LS-SVM applied to the Wiener-Hammerstein benchmark.
      - ✓ Brabanter et al, in *Proceedings of the 15th IFAC Symposium on System Identification*, 2009.
    - Identification of Wiener-Hammerstein systems using LS-SVMs.
      - ✓ Falck et al., in *Proceedings of the 15th IFAC Symposium on System Identification*, pages 820–825, 2009.
  - その他さまざまな技術がある
    - “Large-Scale Kernel Machines”
      - ✓ Léon Bottou, Olivier Chapelle, Dennis DeCoste and Jason Weston, ed., MIT Press, 2007
    - "Recent Advances and Trends in Large-scale Kernel Methods,”
      - ✓ Hisashi Kashima, Tsuyoshi Ide, Tsuyoshi Kato, and Masashi Sugiyama, IEICE Transactions on Information and Systems, Vol.E92-D, No.7 pp.1338-1353, 2009

[スパース化技術]
最近、モデルの簡素化に適用した例があるが、まだ研究途上

- **モデルパラメターをできるだけ削減して「小さい」モデルを作りたい**
  – そのためには、L1正則化や、ARDなどの技術が有効のはず

- **システムの自由度を明示的に推定する試みがいくつかある**
  – リッジ回帰の自由度の推定
    - 統計学の「ハット行列」の理論

- **L1正則化技術（Lasso回帰）を用いた自由度の削減**
  – Kukreja et al. A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification. In *Proceedings of the14th IFAC Symposium on System Identification*, volume 14.

目次

- **最近の機械学習の進歩**
- **Volterra展開とカーネルトリック**
- **その他の関連する話題**
- **IBM東京基礎研究所での取り組み**
  – 無線位置推定
  – 変化点検出

IBM東京基礎研究所での取り組み

- **無線位置推定**
  – 粒子フィルタおよび拡張カルマンフィルタによるトラッキング [Takahashi 08]

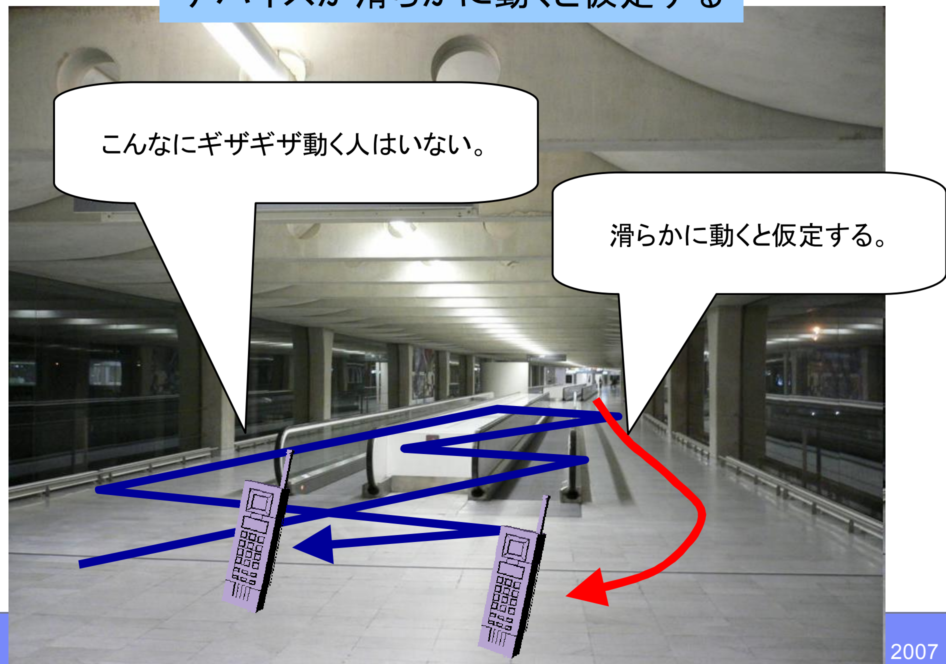
- **変化点検出**
  – 部分空間同定法に基づく変化点検出技術 [Idé-Inoue 05, Idé-Tsuda 07]

無線LANアクセスポイントの電界強度の情報を用いて、店舗内で顧客の動きをトラッキングしたい

- **顧客のトラッキングは応用上重要**
  - スーパーマーケット内でのお客様の動きを解析してマーケティングに生かす、など
- **無線LANアクセスポイントを用いたソリューションの特徴**
  - インフラが非常に安価
  - 精度が悪い
- **非線形・非ガウスの状態空間モデルを用いて、位置推定精度向上を試みる**

事前キャリブレーション

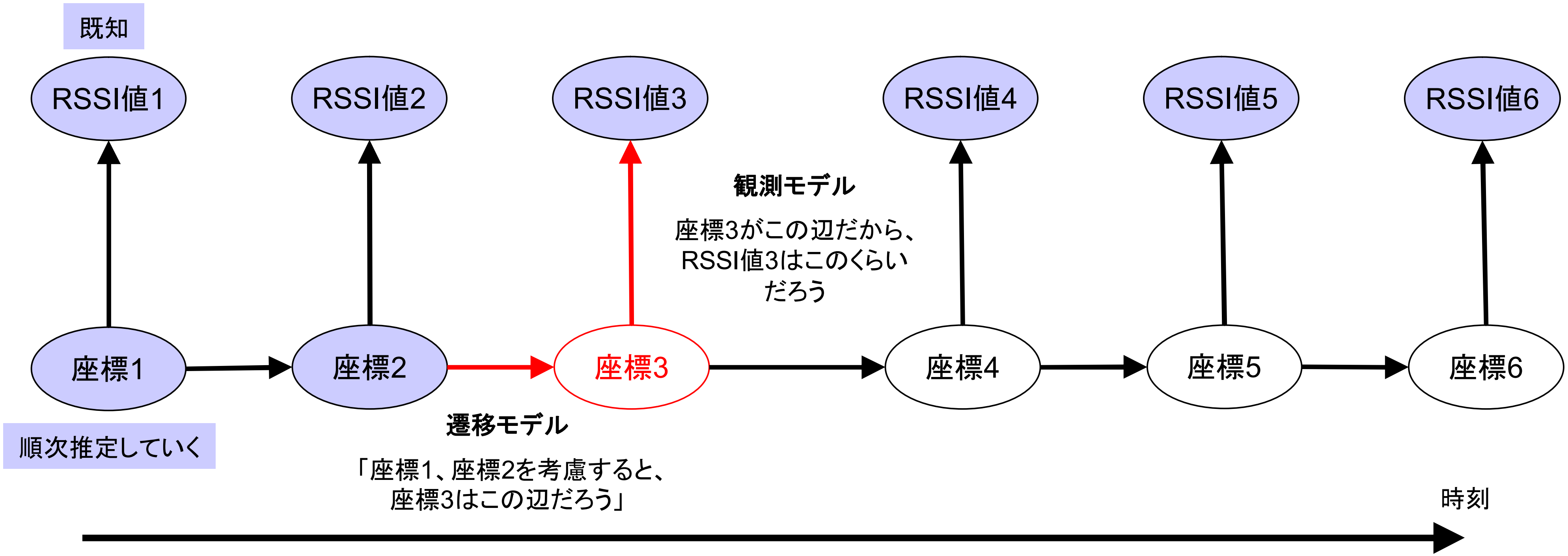
デバイスが滑らかに動くと仮定する



非線形な状態空間モデルで経路を推定する問題として定式化する。

- 観測モデル　→　カーネル回帰モデル＋ガウスノイズ
- 遷移モデル　→　等速直線運動（線形）＋非ガウスノイズ
  - （ノイズが小さければ、あまりギザギザ動かない）

$$\begin{pmatrix} x_{t+1} \\ s_{t+1} \end{pmatrix} = \begin{pmatrix} I & \Delta t \\ 0 & I \end{pmatrix} \begin{pmatrix} x_t \\ s_t \end{pmatrix} + \begin{pmatrix} \sigma_x \\ \sigma_s \end{pmatrix}$$

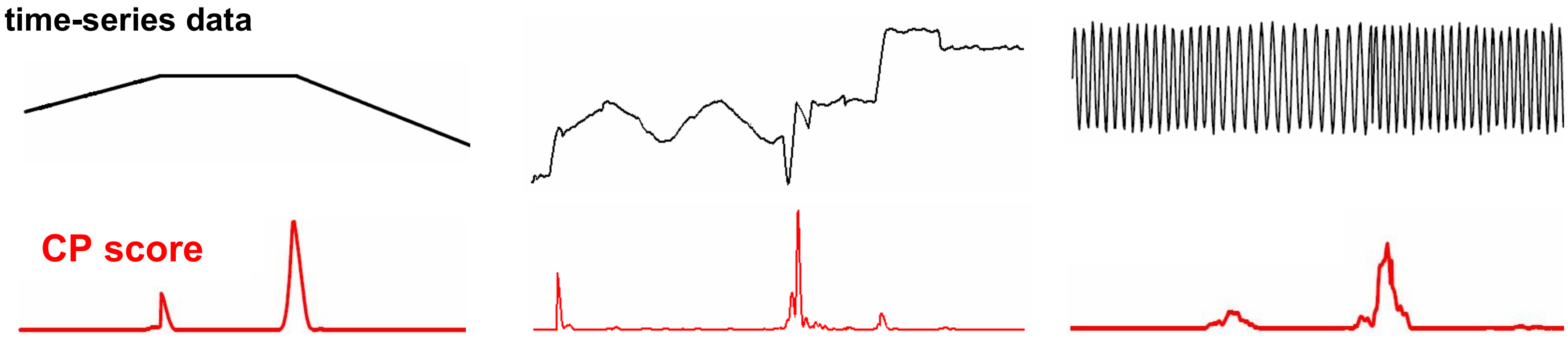


IBM東京基礎研究所での取り組み

- **無線位置推定**
  – 粒子フィルタおよび拡張カルマンフィルタによるトラッキング [Takahashi 08]

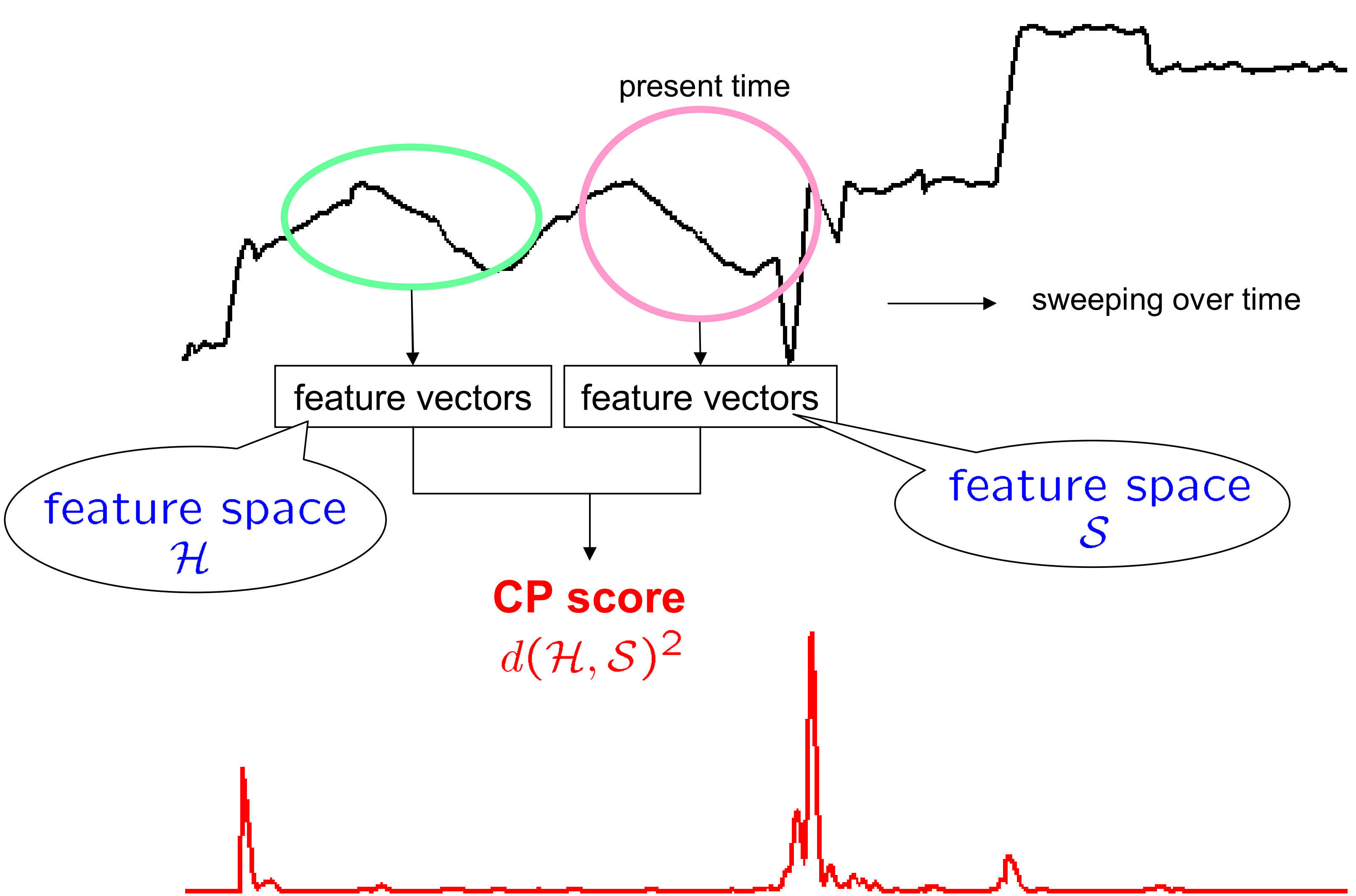
- **変化点検出**
  – 部分空間同定法に基づく変化点検出技術 [Idé-Inoue 05, Idé-Tsuda 07]

変化点検出問題は実用的に重要



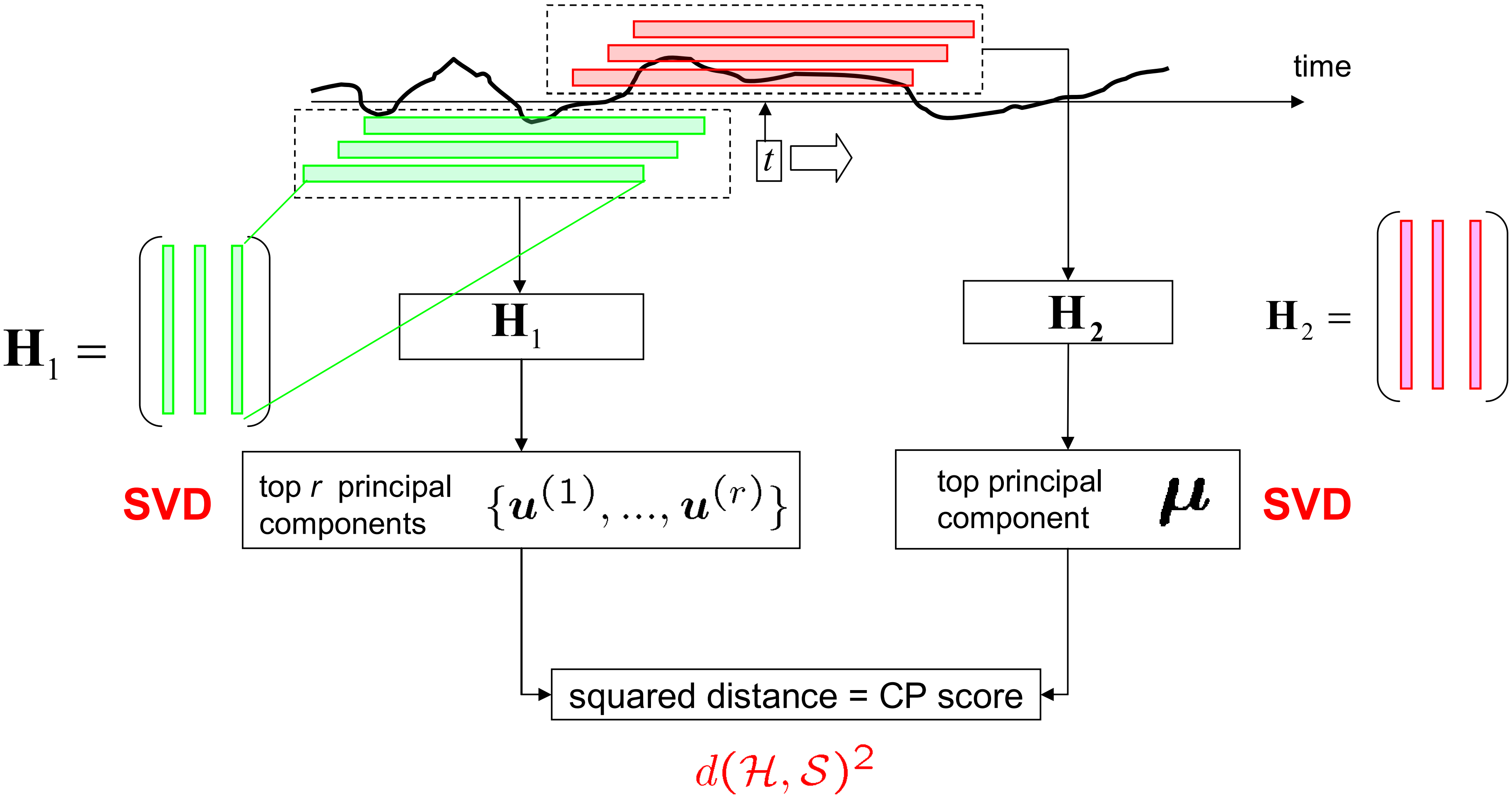
- **CP detection ≒ knowledge discovery**
- **Need to handle the variety of CPs**
  - “model-free” methods are preferable

今と、直前を比べることで変化点を見つける



SST — 特異値分解を用いて変化度を計算

SST = singular spectrum transformation



Krylov 部分空間学習という手法を使って劇的な計算高速化が可能

- **この変化点検出手法は、ほぼ部分空間同定法と等価であることが示されている**
  - [Kawahara 08]参照
  - すなわち、状態空間モデルのパラメター推定を逐次行い、そのパラメターの変化をもって変化点を検出することに対応している
- **Krylov 部分空間学習という手法を使って劇的な計算高速化が可能**
  - 本手法では、変化度の計算は、異なるハンケル行列の特異ベクトル同士の内積の計算に帰着される
  - この内積は、Krylov部分空間学習という手法で劇的に高速化できる [Idé-Tsuda 07]
  - システム同定の文脈で再検討してみるのも面白いかもしれない

ありがとうございました