

ベイズモデリングと歩んだ30年

樋口知之(中央大・理工学部)

新エネルギー・産業技術総合開発機構(NEDO)
ロボット・AI部 PM

産業技術研究所 人工知能研究センター 顧問

(株)ブレインパッド 顧問

(株)ブリヂストン 特別技術顧問



一般社団法人

データサイエンティスト協会 理事

私の研究者人生はちょうど30年

1989年3月（平成元年）

東京大学理学系研究科地球物理学専攻 博士課程修了
同時に 理学博士号取得

1989年4月（平成元年）

文部省統計数理研究所 予測制御研究系 予測理論研究部門助手に採用

2002年7月： 予測制御研究系 システム解析研究部門教授

2004年4月～2010年3月： 副所長(研究企画担当)

2011年4月～2019年3月（平成31年）： 統計数理研究所 所長

2019年4月～ 中央大学理工学部

広尾(港区南麻布)に20年、立川に10年

モデルとモデリング

- 常にモデルを改良する姿勢
- ものの見方、捉え方を柔軟に変化
- 思考のプロセス自体を科学する

モデルを比較し選択するための羅針盤は必須
→ 情報量規準 赤池情報量規準AIC

アウトライン

- i. 1980年代: ベイズ統計
- ii. 1990年代: 粒子フィルタ
- iii. 2000年代: データ同化
- iv. 2010年代: 機械学習とAI
- v. 2020年代: そしてこれから
- vi. さいごに

アウトライン

- i. 1980年代: **ベイズ統計**
- ii. 1990年代: 粒子フィルタ
- iii. 2000年代: データ同化
- iv. 2010年代: 機械学習とAI
- v. 2020年代: そしてこれから
- vi. さいごに

第八代所長 赤池先生との出会い

- 1983年(大学4年時)に「数理科学」の雑誌や、音声圧縮の技術解説書で業績を知る。

数理科学 1981年3月号 No.213

統計モデル

モデル構成の新しい波

<https://www.saiensu.co.jp/search/?isbn=4910054690316&y=1981>

- 大学院時代に宇宙科学研究所(現在、JAXA/ISAS)で故小田稔先生(当時所長)が主催された小さな談話会で直接お話を拝聴する。

統数研が文部省直轄研→[国立]大学共同利用機関化(1985年)

学生の質問にも丁寧にお答え頂き、お優しく、非常にクールでスマートという印象でした。

生物学、天文学、国語学、遺伝学などの現代諸科学の中で、統計学はどのような役割を果たし、どのように利用されているのか。その問いに答えて、国立研究所の各所長が、自己の体験をも通して各分野での現状と展望を興味深く解説する。あわせて、統計的な物の見方の本質と手法の一端を具体的な事例を通して明らかにする。



1987年06月

データ駆動モデリング & 制御の先駆者

巨大なシステムの制御 (第八代所長 赤池先生らの業績)

多変量自己回帰モデル ※ y_n は観測ベクトルデータ

$$y_n = \sum_{j=1}^m A_j y_{n-j} + w_n$$



【写真提供】太平洋セメント(株)

セメント生産プロセスへの統計的方法の適用は、秩父セメント株式会社で初めて実現され、複雑なシステムに対しても制御の可能性を明らかにしました。

(Nakamura and Akaike, 1981)



【写真提供】九州電力(株)

統計的方法を用いた火力発電所のボイラー温度制御は、九州電力株式会社で世界に先駆けて実用化され、優れた制御性を示しています。

(Nakamura-Akaike, 1981)

データ駆動モデリング & 制御の先駆者

船舶のオートパイロット

船の制御

(第十代所長 北川先生らの業績)



【写真提供】東京商船大学(現 東京海洋大)

統計的な方法によって設計されたオート・パイロットの研究
研究成果が生かされ、更に新しい研究が進められています。
(Ohtsu-Horigome-Kitagawa, 1978)

$$\mathbf{x}_n = F\mathbf{x}_{n-1} + G\mathbf{v}_n, \quad \mathbf{y}_n = H\mathbf{x}_n$$

状態空間モデル

$$J(\mathbf{x}_n) = \sum_{j=1}^L \left(\mathbf{x}_{n+j}^T Q \mathbf{x}_{n+j} + r_{n+j-1}^T R r_{n+j-1} \right),$$

評価基準

$$\mathbf{r}_n = \mathbf{y}_n - H\mathbf{x}_{n|n-1}$$

横河電子機器が大型船舶用
主システムPT500Aに採用
(2008年) 川崎汽船は換装

オートパイロットシステムは、業界トップシェアを誇っています。

国内シェア: 50%

海外シェア: 38%

日本郵船、商船三井、川崎汽船をはじめとする船会社が
当社のお客様です。

PT500 オートパイロット

直進中も連続的に船体運動特性を同定
BNAACが省エネ性を実現します

BNAAC: Batch Noise Adaptive Autopilot Controller

制御型自己同定モデルで省エネモード

PT500Aにプリント基板モブドオン

BNAACの頭脳
新プリント基板

PT500Aアダプティブ制御

YOKOGAWA
横河電子機器株式会社
Yokogawa Denki K.K. Co., Ltd.

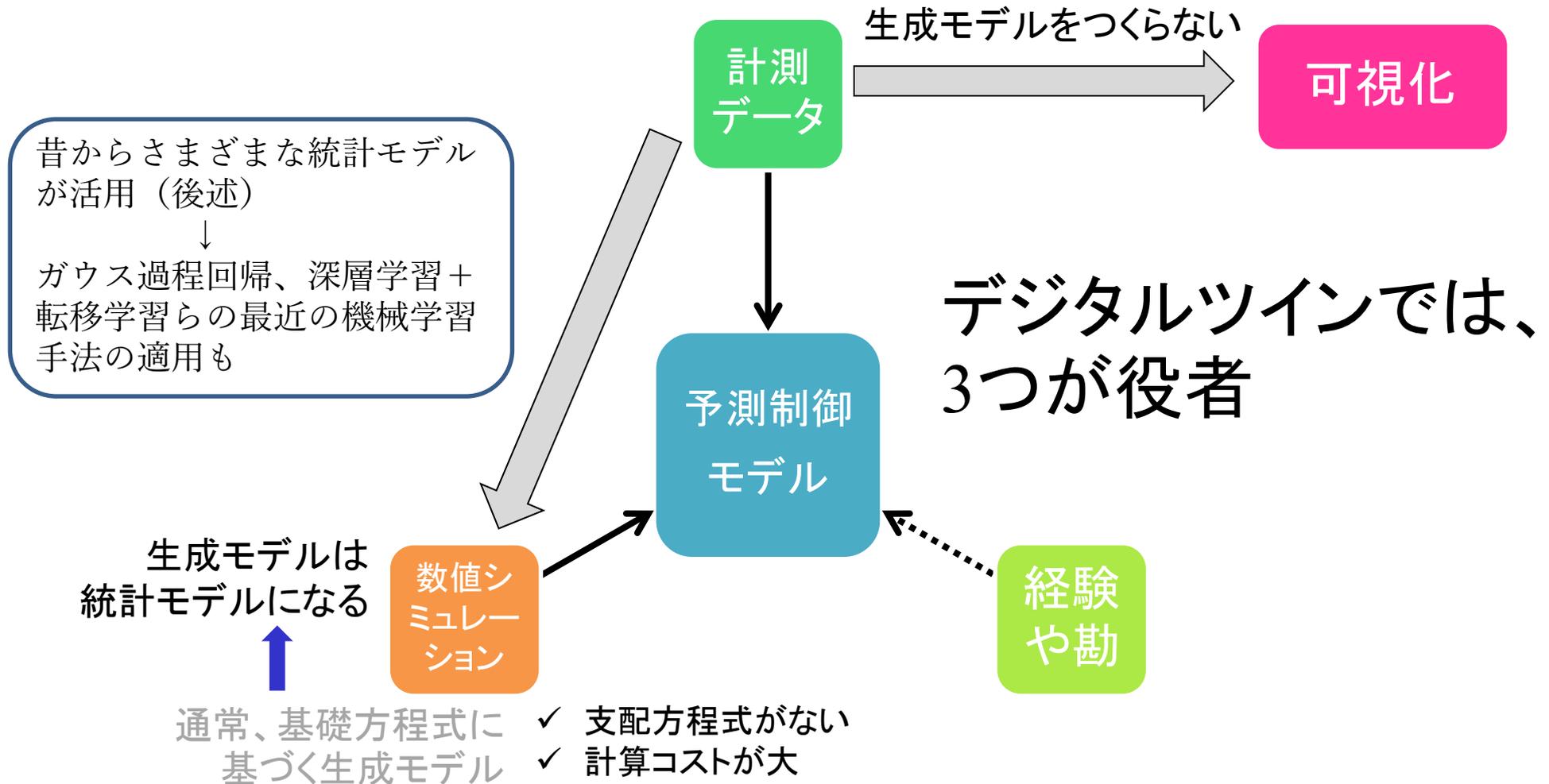
統計工学研究所 北川隆彦所長
東京海洋大学 大津清平名誉教授 ご指導
© Copyright 2009 Yokogawa Denki K.K. Co., Ltd.

出典 <http://rois.ac.jp/outline/pdf/saisyukougi.pdf>

<https://job.rikunabi.com/2016/company/top/r101900017/>

計測データ駆動型生成モデル: 予測モデル

従来のプロセス制御



生成モデルと識別モデル（関数）

Generative Model vs. Discriminative Model

生成モデル（広義）

$$G(z) : p(\mathbf{y} \mid \boldsymbol{\theta}, z)$$

データ パラメータ 説明変数(アスペクト)

生成モデルからベイズの定理を經由して条件付き確率を計算

$$p(C = C_i \mid \mathbf{y}, \boldsymbol{\theta}) = \frac{p(\mathbf{y}, C = C_i \mid \boldsymbol{\theta})}{\sum_i p(\mathbf{y}, C = C_i \mid \boldsymbol{\theta})}$$

クラス分類問題

例: $C=1$ or 0

生成モデル（狭義）

$$p(\mathbf{y}, C \mid \boldsymbol{\theta}) \quad \text{同時分布}$$

データ クラス

$$p(\mathbf{y} \mid C, \boldsymbol{\theta}_2) \cdot p(C \mid \boldsymbol{\theta}_1)$$

識別モデル

$$D(\mathbf{y}) : p(C \mid \mathbf{y}, \boldsymbol{\theta}) \quad \text{条件付き分布}$$

識別関数: \mathbf{y} を C に写像する関数

1980年代の統計学：頻度主義 vs. ベイズ統計



- ベイズ統計の有用性が統計コミュニティにもようやく受け入れられた時
第5部 何がベイズに勝利をもたらしたか
第16章 決定的なブレイクスルー
- 私は物理系出身であったため、概念理解に障害がほとんどなかった
物理学者や生物学者： プレートテクトニクス、パルサー、進化生物学、
汚染、環境、経済、健康、法律
- 当時の統数研の若手同僚に、物理系、数理工学系が複数人いた
- 直轄研→[国立]大学共同利用機関化(1985年)および
総研大が創立(1988年)された直後
統計科学専攻博士後期課程 第一期生二人 (1989年。一人は駒木東大教授) とともに勉強

ベイズの定理がなぜ今役立つのか？3つの理由

イギリスの牧師・数学者(1702 - 1761年)が発見
1763年に発表

x : 興味のある対象

y : データ

2. 対象の特徴をとらえるセンサー性能の向上
高精度センサーのコモディティ(日用品)化

ベイズの反転公式

$$p(\underbrace{x}_{\text{事後分布}} \mid \underbrace{y}_{\text{データ}}) = \frac{\underbrace{p(y \mid x)}_{\text{尤度関数}} \underbrace{p(x)}_{\text{事前分布}}}{\sum p(y \mid x) p(x)}$$

1. 膨大な数の積分(和)操作
には高速な計算機が必要
コンピュータの性能向上
周辺尤度

3. 対象の細かい情報を不確
実性を含めて数値化。個人の
情報を網羅的に収集
ストレージの廉価化

1980年代の統数研

MCMC: マルコフ連鎖モンテカルロ法
計算物理学からデータ駆動モデリングへ (伊庭による解説)

経験ベイズ = ベイズモデリング + 周辺尤度: Akaike (1980)

空間データ(点配置)モデル:
Ogata and Tanemura (1981, 1984)

画像回復(修復) Geman and Geman (1984) → 現在、スパースモデリング

状態空間モデル & 逐次ベイズフィルタ:
Kitagawa (1981, 1987)

現在：1980年代数理・情報技術の再興

- ニューラルネットワーク→深層学習
- 情報統計力学
 - ✓ イジングモデル→量子アニーリング
 - ✓ シミュレイトイドアニーリング
- 遺伝的アルゴリズム、遺伝的プログラミング
→超並列計算による大規模離散最適化
- 連続最適化(降下法)→確率的勾配降下法、
オンライン学習
- 制約付き最適化→スパースモデリング
- 次元圧縮→行列、テンソル分解

最適化

あこがれの統数研に助手として採用

•1989年(平成元年)に研究所に採用されました。当時赤池先生が所長であられ、先生から辞令をいただきました。

「君の論文※は僕が査読したんだよ」と辞令交付時に小さい声でおっしゃられました。いまだに真意がわかりません。(※経験ベイズを用いた応用光学分野の逆解析)

1988～1994 第八代所長
(1927～2009: 伊庭教授 撮影)

•統計数理研究所の助手は、もっとも恵まれた境遇にいることを十分認識して研究に励むように忠告を受ける。

統数研の人事採用方針

統計数理と数理統計は全く異なる概念。詳しくは本日の配布資料を参考にしてください。

- 数理統計、確率論に深い専門性をもつ人材
→ 理論の深化
- 物理、計算(機)科学(Computer Science)にも常に目配り
→ 計算の高度化、高速化
- データの質、量の観点で大きく変化のある分野から
→ 学際性を富ませ、未来分野を開拓

多様な分野を人でつなぐ

この3つをどうバランスさせるかが難しい

人工知能学会の栄枯盛衰

1986年創設

2017年
IBIS (統数研と長年にわたって共催)
電子情報通信学会 2種研
日本の機械学習研究の最大コミュニティ

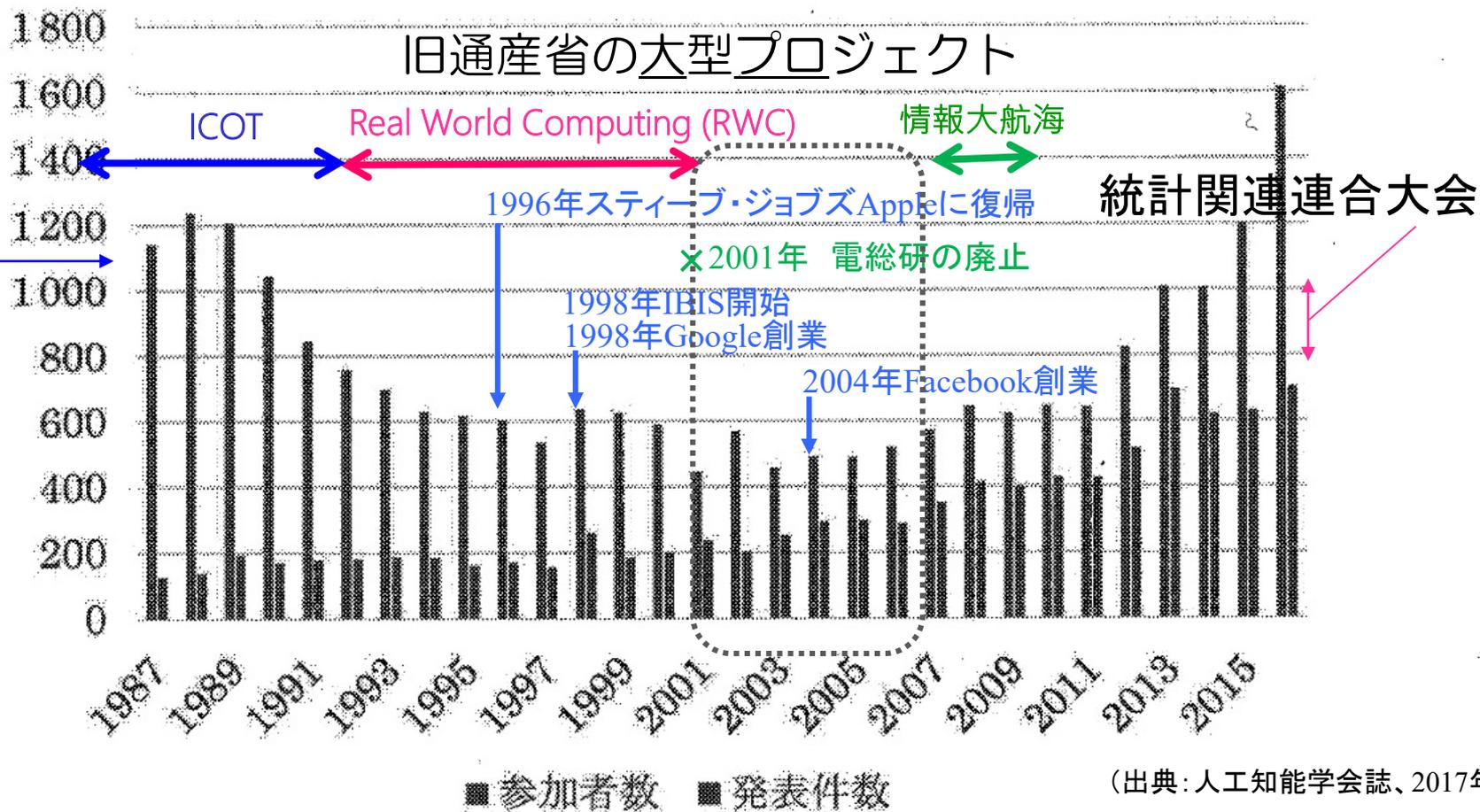


図4 全国大会参加者数と発表件数の推移

2016年10月に、それまで最高であった1992年10月の会員数を超える

忘れ去られた大プロ、RWC

RWC プロジェクトを回顧する

大津 展之* (電子技術総合研究所)

東アジア・ASEAN経済センター
西村事務総長

要旨

第五世代コンピュータプロジェクトに続く通産省の10年の大型国家プロジェクトとして、新情報処理研究計画、すなわち、リアルワールドコンピューティング (RWC) プロジェクトが、1992 (平成4) 年度に始まり2001 (平成13) 年度で終了した。

第五世代コンピュータプロジェクトが人間の情報処理の論理的側面 (記号処理) を追求したのに対して、RWC プロジェクトの狙いは、いわば直観的側面 (パターン処理) を新しい情報処理の枠組みとして基礎づけ、21世紀の高度情報化社会に向けての新たな情報処理パラダイムとしての「柔らかな情報処理」、すなわち、実世界における多様な情報をより人間に近い形で直接かつ柔軟に処理する知的情報処理の実現を目指し、そのための基礎と基盤技術を育成するものであった。

本稿では、RWC プロジェクトの計画、立案、推進に深く関わった研究者として、その背景、研究開発の内容と体制、成果について概要を紹介するとともに、そのキーコンセプト、そしてアプローチの枠組についても、私見や思いを交えて回顧する。

アウトライン

- i. 1980年代: ベイズ統計
- ii. 1990年代: **粒子フィルタ**
- iii. 2000年代: データ同化
- iv. 2010年代: 機械学習とAI
- v. 2020年代: そしてこれから
- vi. さいごに

私の1990年代：粒子フィルタ

Andrew Blake is Deputy Managing Director at the Microsoft Research laboratory in Cambridge, where he also leads the Machine Learning and Perception Group (MLP) with Prof. Christopher Bishop.

CONDENSATION (1998) Conditional Density Propagation

1960年代

統計物理…逐次モンテカルロ計算

Monte Carlo filter (1993)
Kitagawa

Bootstrap filter (1993)
Gordon *et al.*
(DERA, QinetiQ, DSTO)

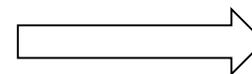
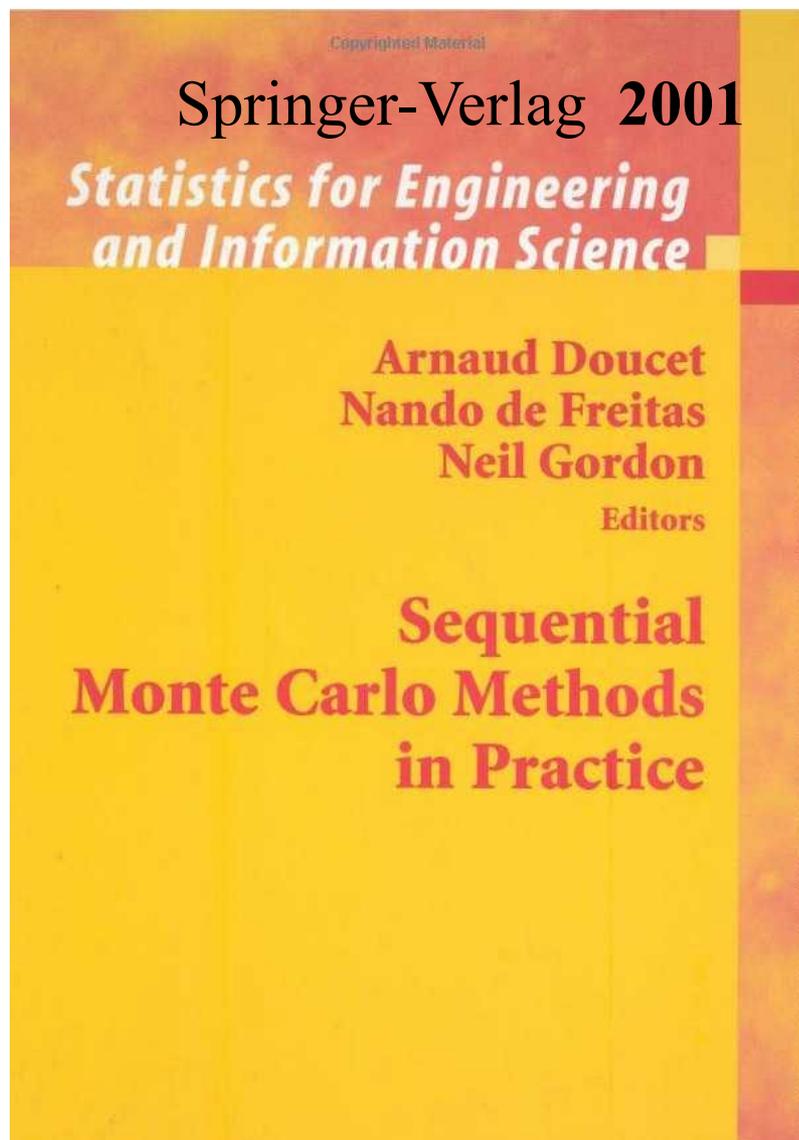
While the most immediate battle for Sony, Nintendo, and Microsoft is who can sell the most games and the most motion control hardware, it's clear that Kinect is a major technical achievement with some pretty *broad reaching implications for human / machine interaction going forward.*

Particle filter

PFにより国際的に若手の研究者と多く知り合える

共同執筆

2006年UBCにて



Oxford大統計学科教授

19

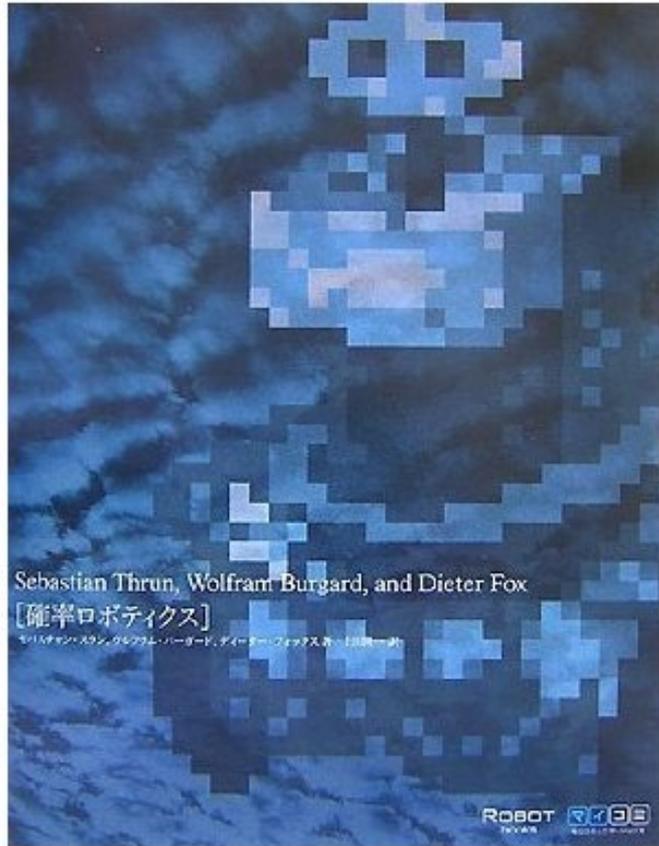
Particle Filters for Mobile Robot Localization

Dieter Fox
Sebastian Thrun
Wolfram Burgard
Frank Dellaert

20

Self-organizing Time Series Model
Higuchi, T.

確率ロボティクスとDARPA アーバンチャレンジ



● Sebastian Thrun (著), Wolfram Burgard (著),
◆ Dieter Fox (著), 上田 隆一 (翻訳)

JGfOs Japanese-German Frontiers of Science Symposium 2011 2010 2009 2008 2007 2006 2005

JGfOs 4 JGfOs2007

Date: November 2-4, 2007
Venue: Shonan Village Center (Kanagawa, Japan)
Participants: Total of 60, 30 from each side

Session Topics:

- Chemical Biology - New Approaches for Drug Discovery
- Cold Quantum Gases
- Early Life
- Economic Diversity
- Frontiers in Robotics
- Mechanisms of Chronic Pain

Frontiers in Robotics (Mathematics/Applied Mathematics/Computer Science)

PGM	Tomoyuki Higuchi	The Institute of Statistical Mathematics Research Organization of Information and Systems
PGM	Alois Knoll	Technical University Munich, Department of Informatics
Introductory Speaker	Tomohiro Shibata	Nara Institute of Science and Technology, Graduate School of Information Science Abstract
Speaker	Eiji Uchibe	Okinawa Institute of Science and Technology, Initial Research Project Abstract
Speaker	Wolfram Burgard	Albert-Ludwigs-University of Freiburg, Department of Computer Science Abstract



ちょっと脇道にそれますが、先端科学シンポジウム (JGFoS)



第3回 JGFOS

JGFoS³2006

日時 : 平成18年11月3日~5日
場所 : ホリデーインハイデルベルグ (ドイツ、ハイデルベルグ)

■セッショントピックス:

- [Acceptance of New Technologies](#)
- [Chemistry of Neuro-disorders: Their mechanism, diagnostics and drug development](#)
- [CO2 and Climate change: past - present - future](#)
- [Machine Learning and Prediction](#)
- [Neurogenomics](#)
- [Spintronics](#)

■プレスリリース

Machine Learning and Prediction (Mathematics / Applied Mathematics / Computer Science)		キーワード集	
PGM	樋口 知之	情報・システム研究機構統計数理研究所	教授・副所長 セッション アブストラクト
PGM	Alois Knoll	Technical University Munich, Department of Informatics, Robotics and Embedded Systems	Professor
スピーカー	石井 信	奈良先端科学技術大学院大学情報科学研究科情報生命科学専攻	教授 スピーチ アブストラクト
スピーカー	馬見塚 拓	京都大学化学研究所バイオインフォマティクスセンターパスウェイ工学	教授 スピーチ アブストラクト
スピーカー	Barabara Hammer	TU Clausthal	スピーチ アブストラクト
スピーカー	Bernhard Schokopf	MPI for Biological Cybernetics, Tübingen	スピーチ アブストラクト



先端科学シンポジウム

JSPSが日本側のホスト機関

FoS Alumni Messages No.6

「FoSの魅力：二つの理由」



樋口 知之

情報・システム研究機構 理事
統計数理研究所 所長
総合研究大学院大学統計科学専攻 教授

FoS参加歴:

[1st JGFoS](#) スピーカー
[3rd JGFoS](#) PGM
[4th JGFoS](#) PGM

2010年代に：数多くのPF研究者が深層学習に向かう

- 諸外国は、統計的機械学習のエキスパートが深層学習にどっと参入

- ^{ドイツ} S. Thrun (CMU) → Google + SAIL (Stanford AI lab.)
Google X, ユーダシティ, Kitty Hawk
- ^{ジンバブエ} N. de Freitas (Oxford) → Google DeepMind
- ^{イラン} Z. Ghahramani (Cambridge U.) → Uberの C.S.

アカデミア→産業界への人材流出傾向は欧米でも同じ

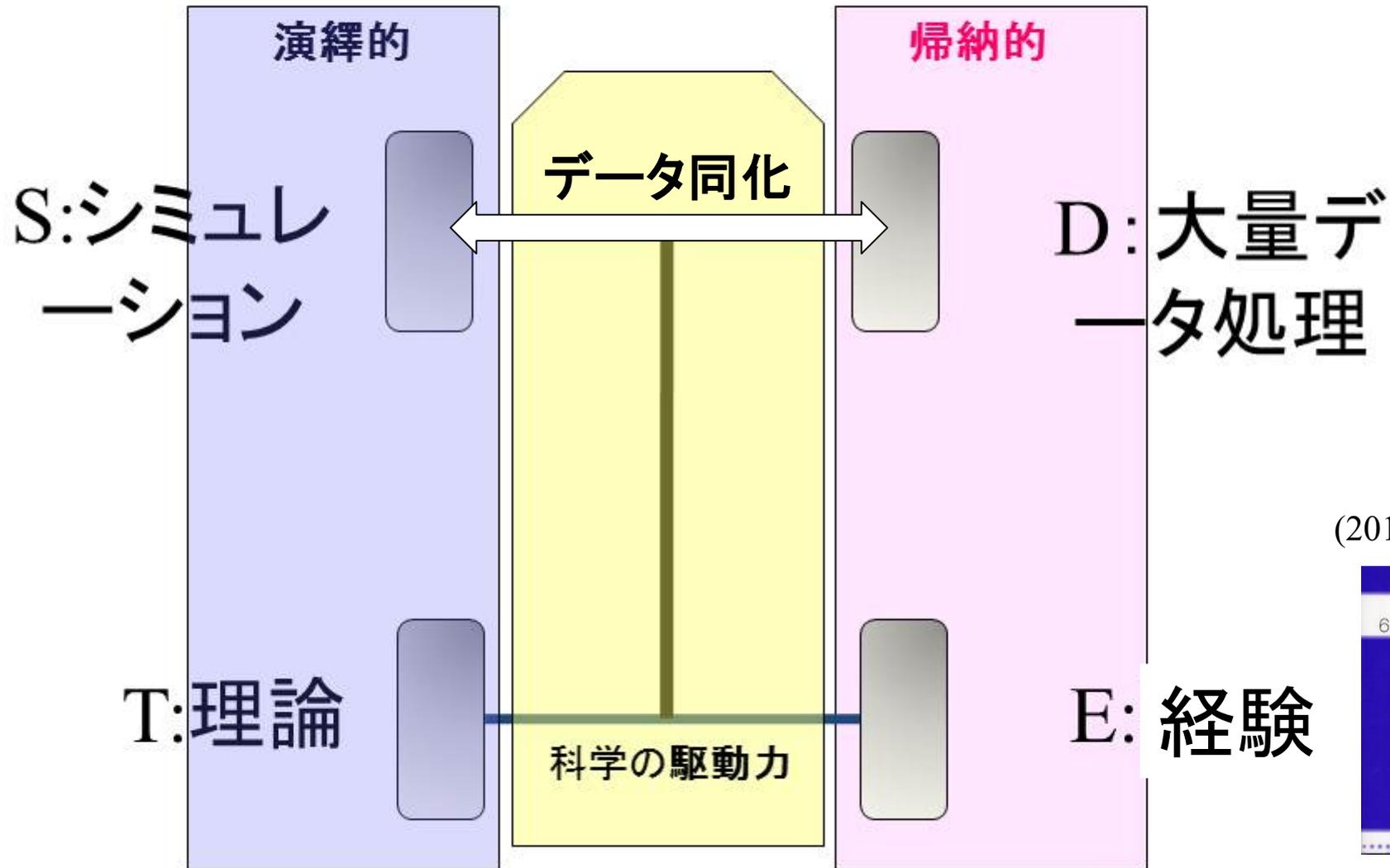
産業界(先端IT企業)とアカデミアのクロスアポイント

- 日本はこれまで、ベイズモデリング、統計的機械学習の専門家が少なかったのが問題
信号処理、時系列解析の基礎概念の習得不足も心配

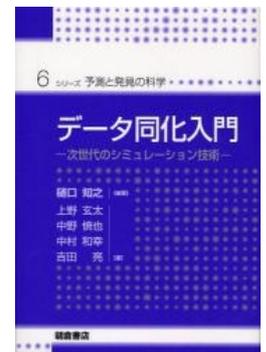
アウトライン

- i. 1980年代: **ベイズ統計**
- ii. 1990年代: **粒子フィルタ**
- iii. 2000年代: **データ同化**
- iv. 2010年代: **機械学習とAI**
- v. 2020年代: **そしてこれから**
- vi. **さいごに**

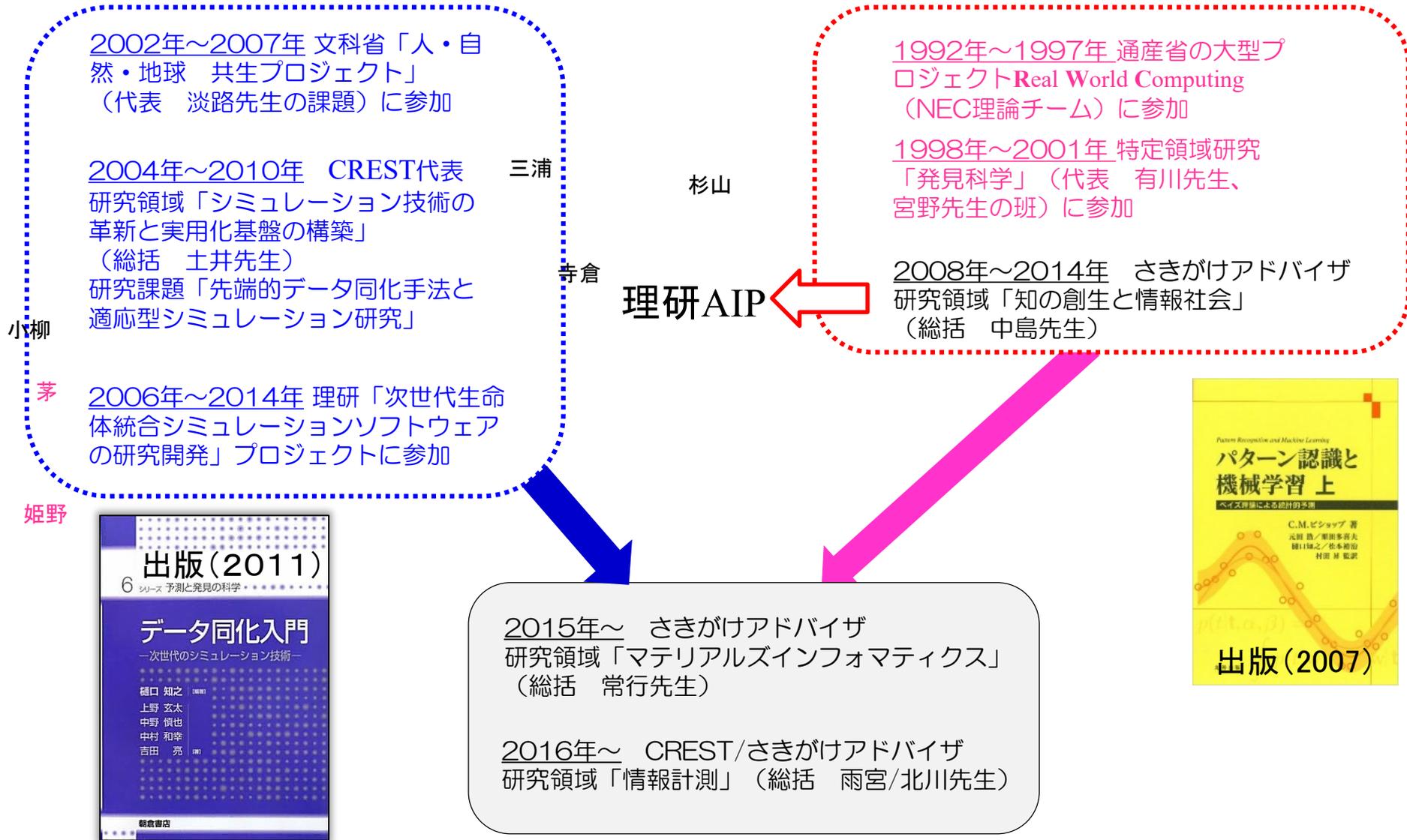
つなぐ：データ同化



(2011年9月刊行)



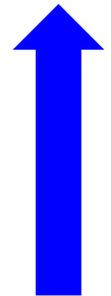
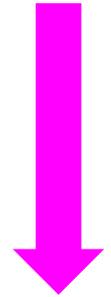
私の2000年代：データ同化と機械学習との邂逅



メゾスコピック・モデリング

従来の方法論
では限界

数理表現さ
れた法則



第一原理計算

材料

熱力学
弾性体

界面・粒界



量子力学

生命

疫学

個人差



ゲノム解析

経済

マクロ経済
モデル

ネットワーク



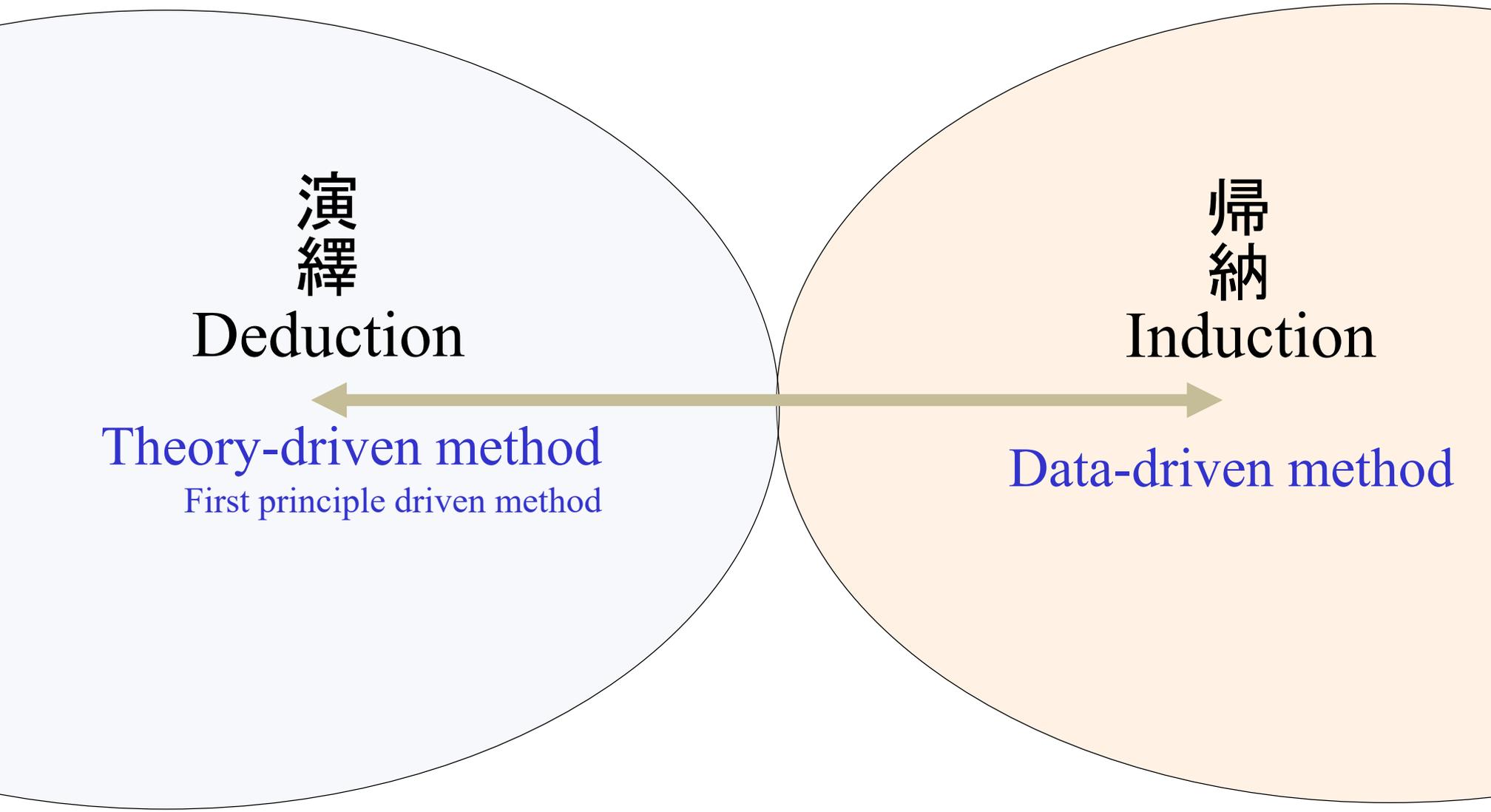
エージェント
モデル

マクロ

メゾスコピック

ミクロ

異質・相反する思考法



新幹線強風予報システムの開発に協力

(第十代所長 北川先生 & 第11代所長 樋口の業績)

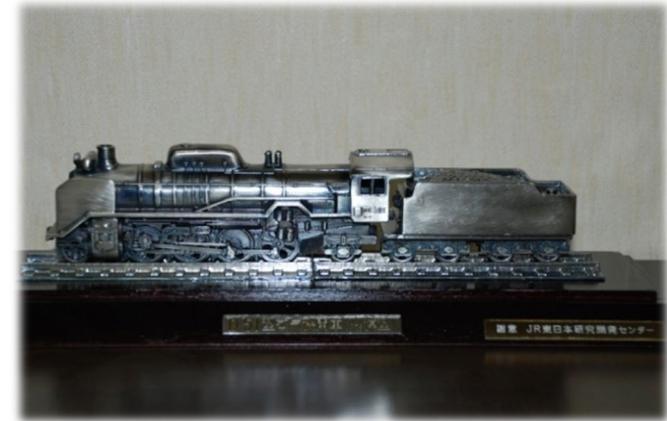
強風予報 列車の運行スムーズに

JR東日本が開発
実用化なら運休時間2, 3割減

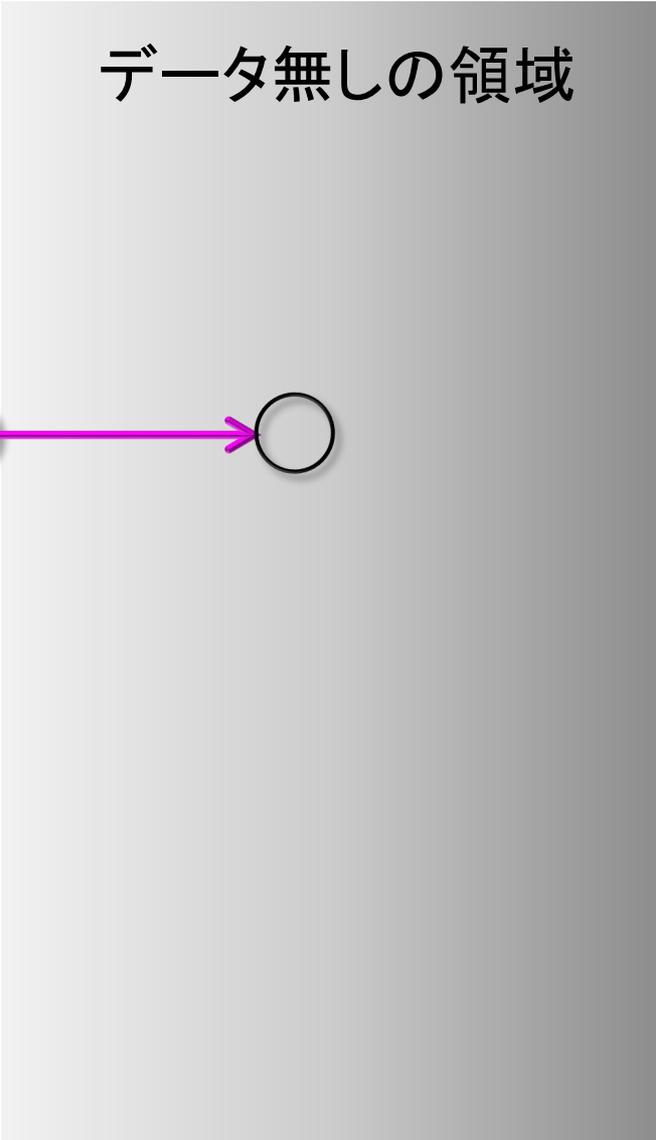
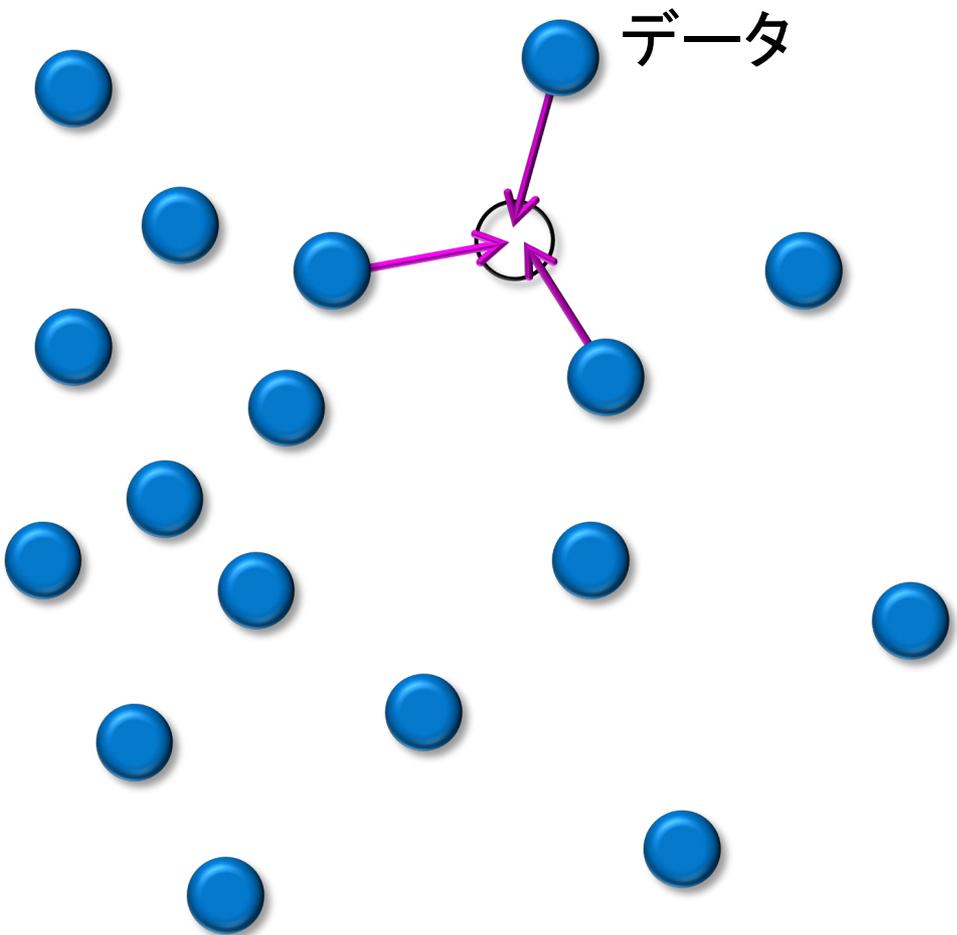
.....システムは、3分間隔で記録した風速データを統計学の手法で作った数式にあてはめ、その後の風速をコンピュータに計算させる。数式は「風が安定したとき」「荒れ模様の時」など6パターンあり、.....

朝日新聞平成14年2月17日(日) 14版第2社会面

状態空間モデル の利用



内挿と外挿



時系列予測は外挿的特性が強い

- 時系列データの長期予測は、基本的に外挿問題
- 一度も起きていない事象の前もっての予知(定量的“前兆”把握)は、内挿法では不可能
- 内挿(帰納法)だけでは、局所予測(線形+ランダムウォーク)の域を超えられない
- システムの動的変動のモデル化に関して、その精緻化が必須
- データの背後にある、現象に関する事前情報の具体的な数理的表現化

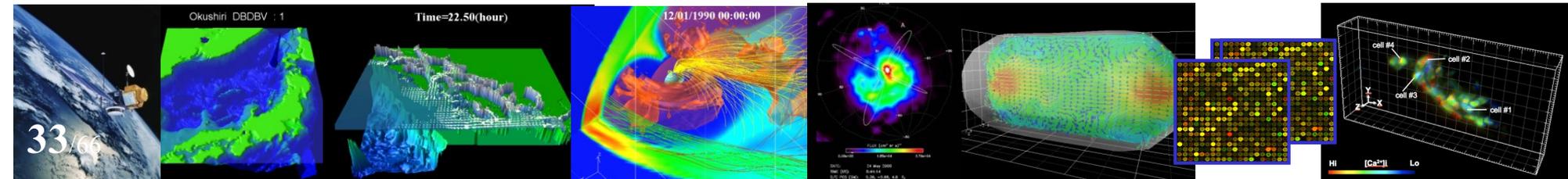
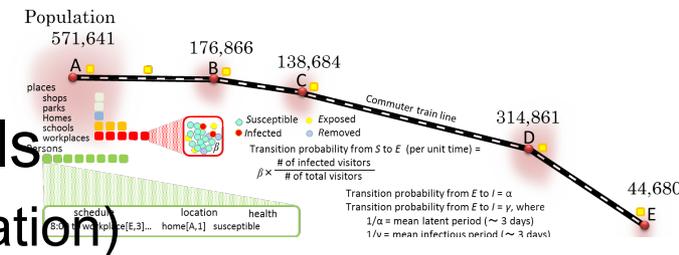
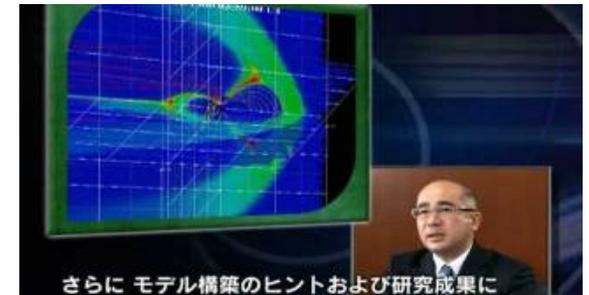
Application studies carried out by our group for **model** improvements

Research Projects:

- Typhoon trajectory
- Tsunami, Ocean tide
- Auroral phenomena
- 3D structure of ring current
- Acoustic waves
- Intercellular fluid dynamics
- Genome informatics
- Drug response prediction
- Neuronal circuits of whole nerve cells
- Influenza Pandemic (Multi agent simulation)



Watch YouTube “Data Assimilation R&D Center at ISM”.



データ同化で推定したい変数

- 時間に依存しないパラメータ

θ : モデルパラメータ (次元は < 20 程度)

\mathbf{x}_0 : 初期状態ベクトル(初期条件)

$$\dim(\mathbf{x}_0) = 10^6 \sim 10^8 \quad \text{Adjoint法}$$

- 状態ベクトルの列

$$\mathbf{x}_{1:T} = [\mathbf{x}_1, \dots, \mathbf{x}_T]$$

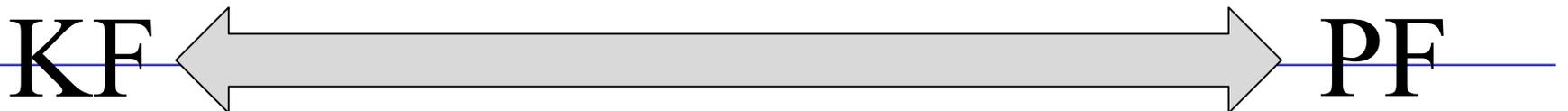
事後分布 $p(\bullet | \mathbf{y}_{1:T})$

成熟期にある逐次データ同化技術

Kalman Filter (KF) は1960年代から

- PF (Particle Filter) 1996年
- EnKF (Ensemble KF) 1994
- EnTKF (En transform KF) 2001
- LETKF (Local EnTKF) 2007
- MPF (Merging PF) 2007
- pMCMC (particle MCMC) 2010

次元 (状態ベクトルの次元、アンサンブルメンバー数) の違いと空間情報の局所性



アウトライン

- i. 1980年代: ベイズ統計
- ii. 1990年代: 粒子フィルタ
- iii. 2000年代: データ同化
- iv. 2010年代: **機械学習とAI**
- v. 2020年代: そしてこれから
- vi. さいごに

統計的推論法のパラダイムシフトを促す外的要因

ビッグデータ、ビッグデータ、...ビッグデータ

- (ちょっと前) クラウド、スマホ
- (今) IoT、AIアシスト (Amazon Echo, Google Home)
- (ちょっと先) Lifi ※ (次世代Wifi。今の100倍)、5G (次世代通信。10Gbps)

データ駆動サービス社会

- “役に立つ”問題、解きたい問題が爆発的に増加
- モノからサービス(体験)へ、全体から個(個別、個性、固有)へ
- 発見から予測へ、因果から相関へ
- 統計的モデリングから最適化関数設計へ、統計的推測から最適化へ

目的特化型計算機

- NVIDIA GPGPU
- (量子)アニーラー、AIチップ → 量子コンピュータ

データの観点で、何が変わったのか？

[教師あり学習]

$$z_i := (y_i, \mathbf{x}_i) \quad i = 1, \dots, N$$

ラベル

説明変数ベクトル

特徴ベクトル、記述子

サンプル数

$$p = \dim(\mathbf{x}_i)$$

データ (サンプル) 次元

2000年代始め

新NP問題

$$N \ll p$$

今でも、理論統計学の分野では世界的に研究が盛ん

2010年代

N がさらに激増

- 計算量が $O(N)$ では役に立たない
- オンライン学習 (バッチ学習・処理ではダメ)
- モデル等比較にはクロスバリデーションが標準

変数選択から、… この20年間の内挿手法の劇的な性能向上

線形・非ガウス→スパースモデリング

変数(要素)間の関係 最適化関数のクラス

Tibshirani, R. (1996)

Candès, E.J. and Tao, T. (2005)

非線形・ガウス→深層学習

Hinton, G.E. (2006)

線形・非ガウスモデルと非線形・ガウスモデルが
マシンで得られる時代になった！

モデリングの技



機械学習の有用性がさらに浸透する ダウンスケーリング：超解像

[教師あり学習]

$$\{ \mathbf{y}_i, \mathbf{x}_{0,i} \}_{i=1}^N$$

初期値、境界条件



$$\phi(\mathbf{y} | \mathbf{x}_0^*)$$

計測
データ

サロゲートモデル

統計的ダウンスケーリング

高解像度
シミュレー
タ

コンポーネントとし
てプラグイン

経験
や勘

数値シ
ミュレー
ション

低解像度初期条件・
境界条件
→ 高解像度の解

$$\mathbf{y} \equiv g(\mathbf{x}_0^*)$$

[シミュレーション]

力学的ダウンスケーリング

融合が進む：エミュレーション、仮想計測

マテリアルズインフォマティクスの
典型例の一つ

[教師あり学習]

$$\{y_i, z_i \equiv f(\theta_i)\}_{i=1}^N$$

記述子 ↓

$$\phi(y|\theta^*)$$

教師データの作成
に利用

計測
データ

y_i

機械学習

エミュ
レータ

計測モデルを
統計モデルで構築

記述子

パラメータ
 $z_i \equiv f(\theta_i)$

数値シ
ミュレ
ーション

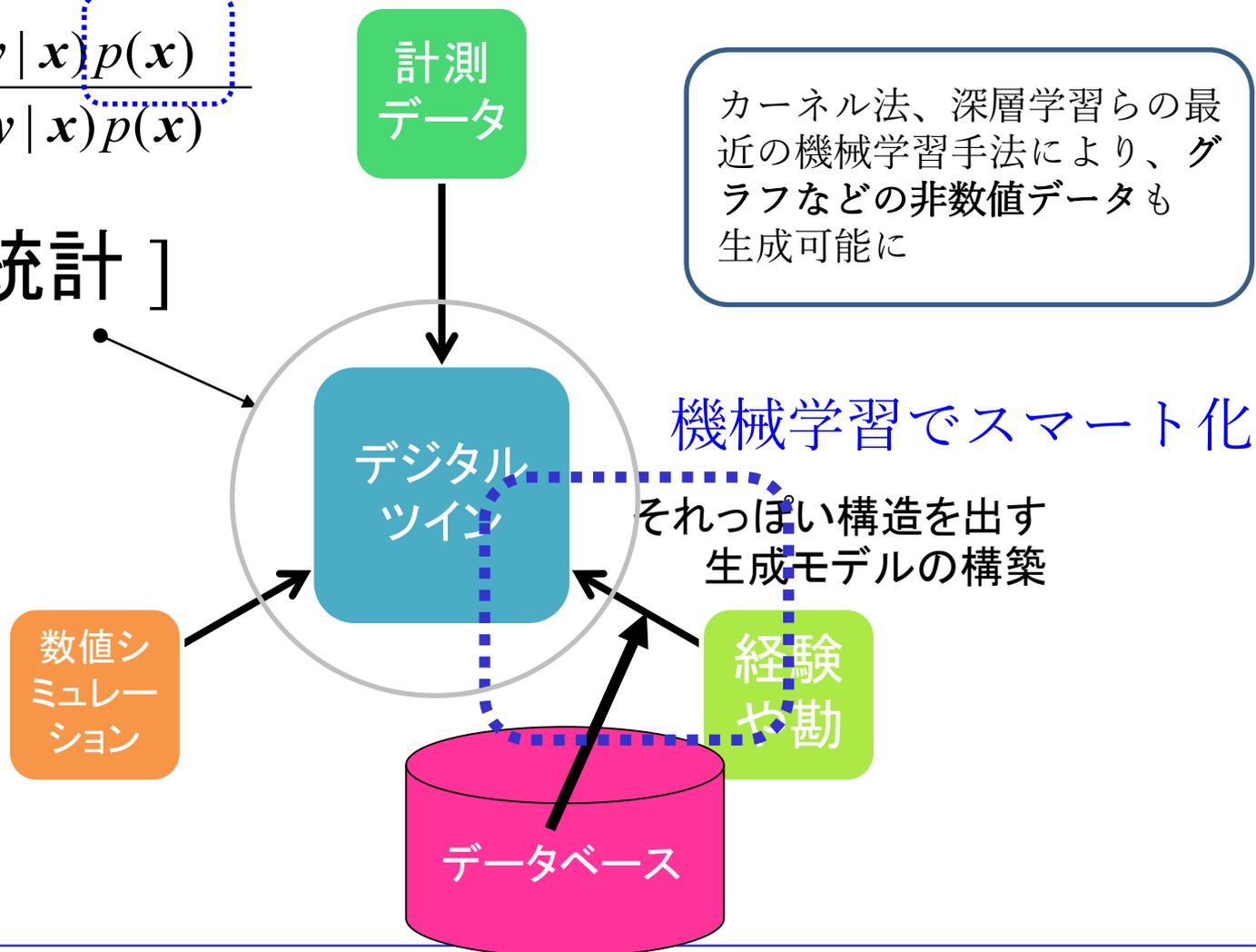
経験
や勘

基礎方程式に基づく生成モデル
でビッグデータを産出

暗黙知（経験と勘）を具現化する生成モデル

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x}) p(\mathbf{x})}{\sum p(\mathbf{y} | \mathbf{x}) p(\mathbf{x})}$$

[ベイズ統計]



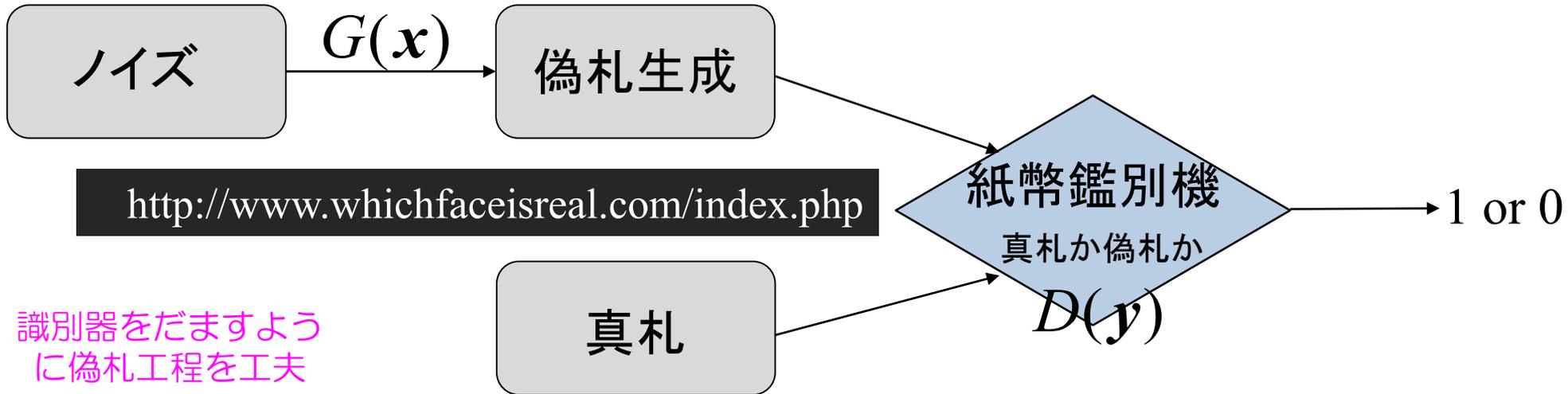
アウトライン

- i. 1980年代: ベイズ統計
- ii. 1990年代: 粒子フィルタ
- iii. 2000年代: データ同化
- iv. 2010年代: 機械学習とAI
- v. 2020年代: そしてこれから
- vi. さいごに

いたちごっこ：GANの基本アルゴリズムの比喩的解説

識別モデル $D(y)$ ：真札である確率

生成モデル $G(x)$ ：偽札を生成するモデル(偽札工程)



真札・偽札を
正しく識別

学習アルゴリズム

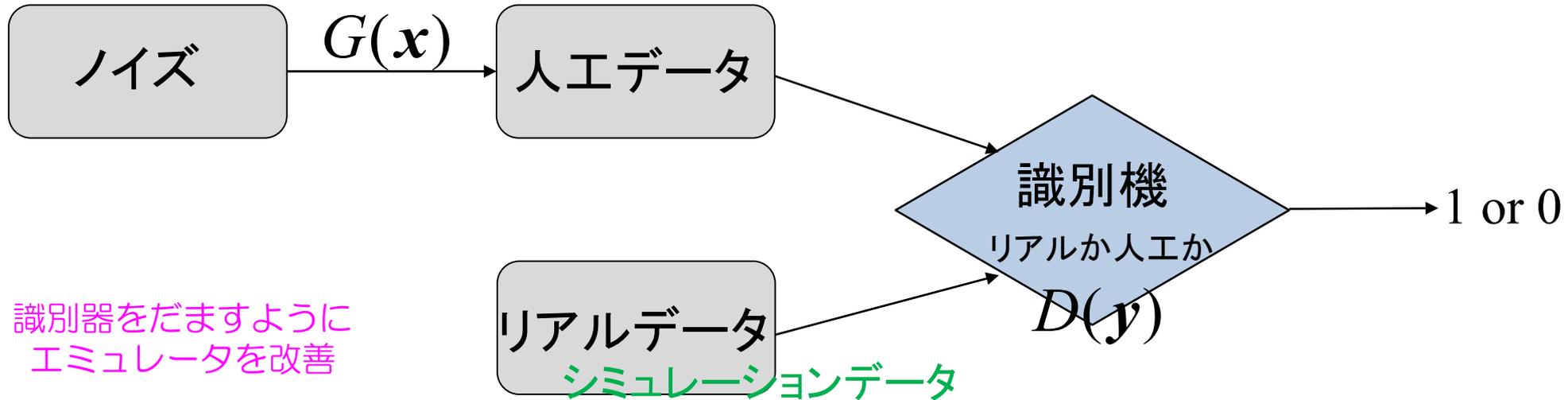
$$\min_G \max_D E_{y \sim p_{\text{data}}(y)} [\ln D(y)] + E_{x \sim p_x(x)} [\ln(1 - D(G(x)))]$$

① GANをつかってエミュレータをつくる

$D(y)$: リアルデータである確率

機械学習を使った研究は、 $D(y)$ の獲得に注力しているが、データが少ない場合 $G(x)$ の構成も同時に狙う研究のほうが元来筋がよい

$G(x)$: エミュレータ



識別器をだますように
エミュレータを改善

リアルデータと人工
データを正しく識別

学習アルゴリズム

$$\min_G \max_D E_{y \sim p_{\text{data}}(y)} [\ln D(y)] + E_{x \sim p_x(x)} [\ln(1 - D(G(x)))]$$

② 深層確率コンピューティングプロジェクト

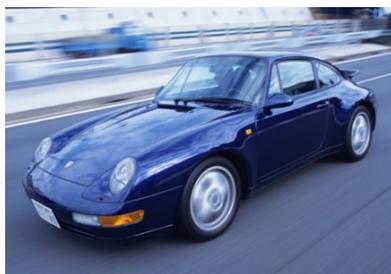
- 深層学習と状態空間モデルを融合し、時系列確率モデルの学習を容易かつ柔軟に実現
- エッジ側で時系列確率モデルを構築し、サーバには結果のモデルのみを送信する情報システムを確立
- 2018～2022年度の5年間  NEDO IoT部門

エッジ側 IoT (例)

屋内外環境



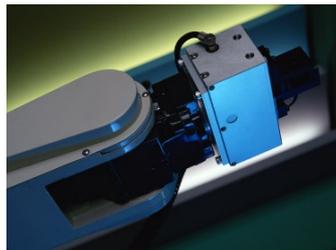
自動車



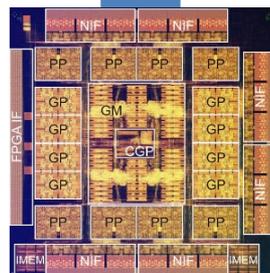
圃場



工作機械



各種センサからの
時系列データ



専用化により
学習に必要な
エッジ側消費
電力を1/100に

新アルゴリズム・
専用言語の開発

サーバ側 クラウド

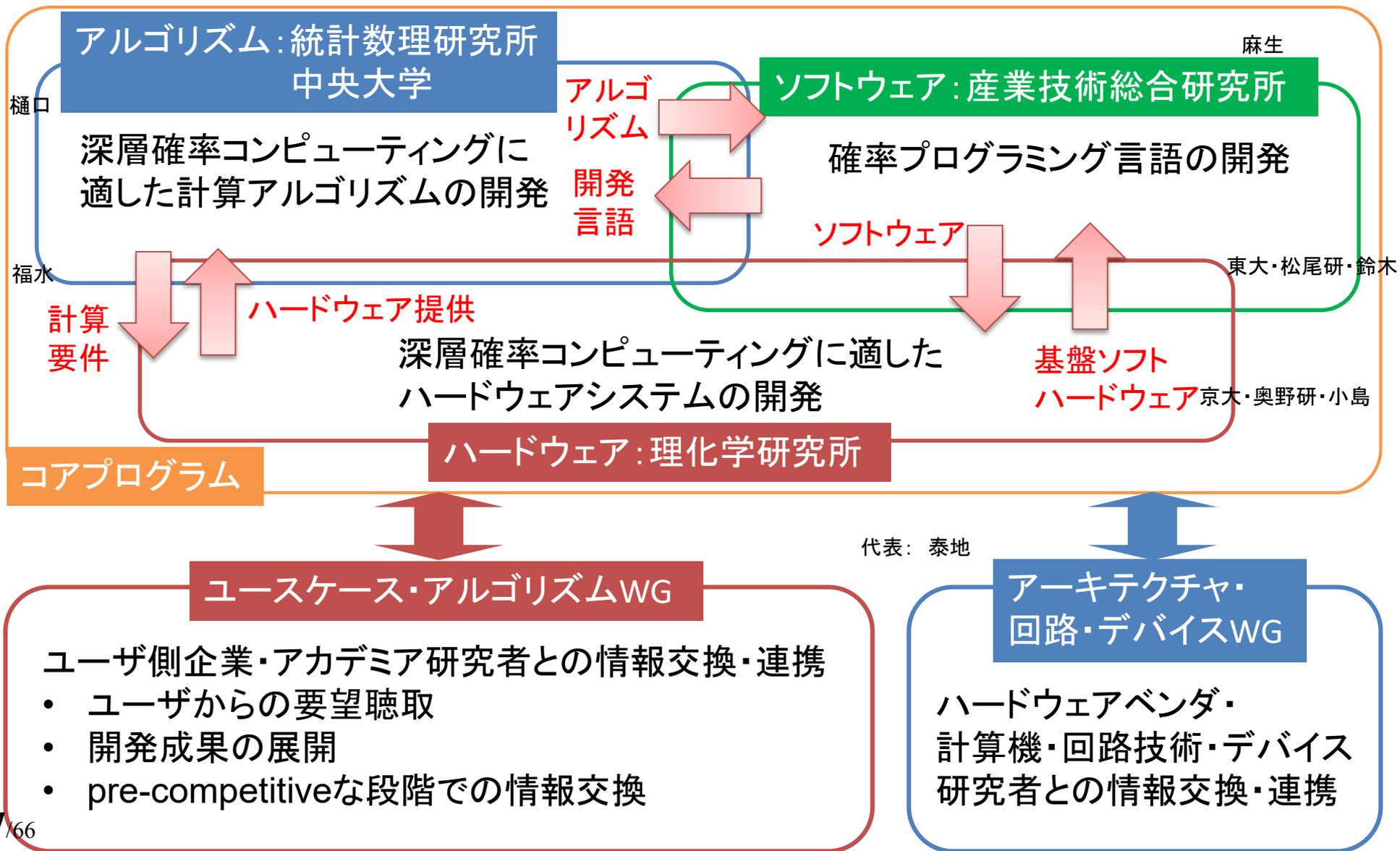


学習結果のモデルのみを送信する
ことで以下を実現

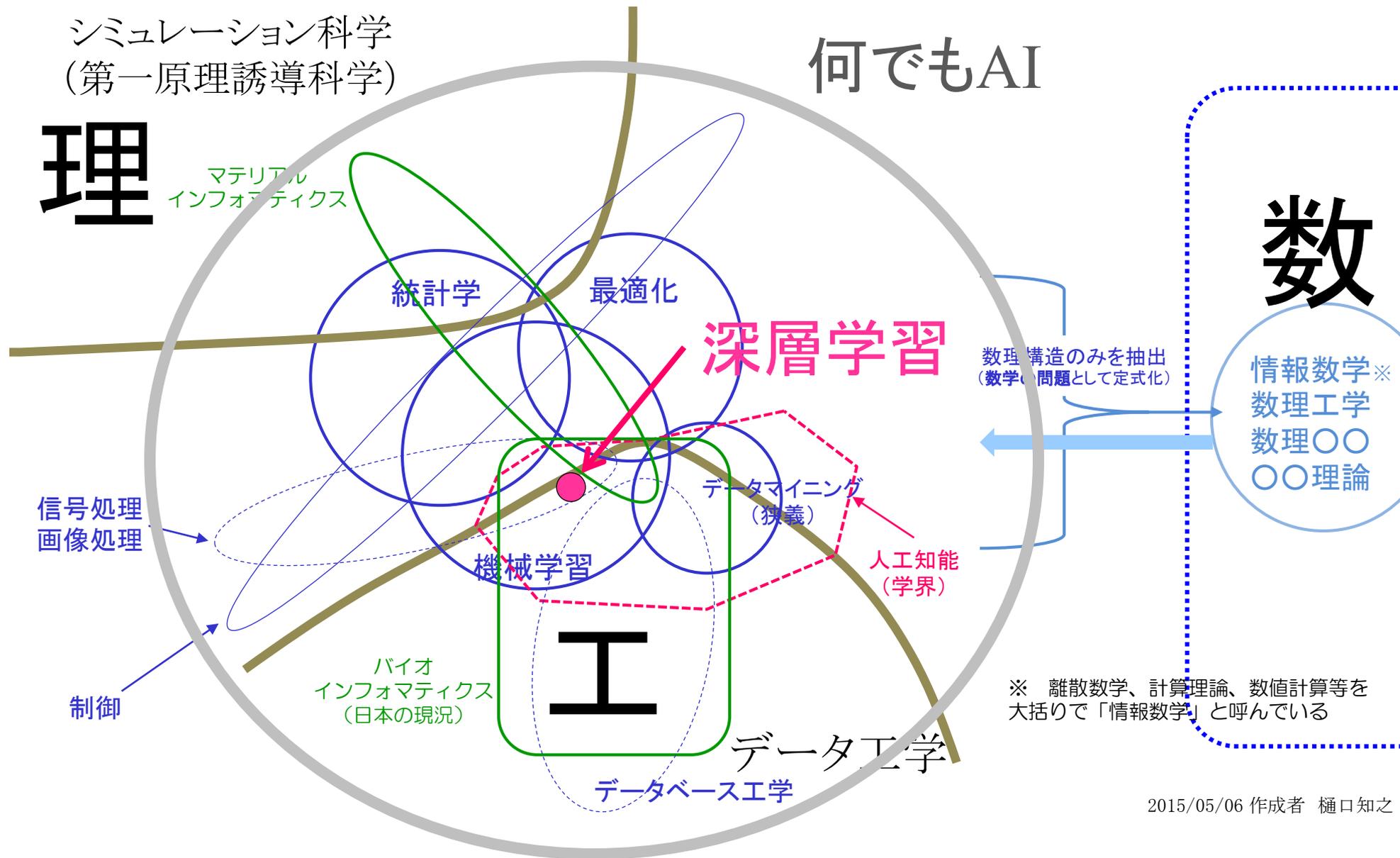
- ・必要な通信帯域を1/100以下に
- ・匿名化
- ・データ欠落に対する頑健性

研究開発の内容と体制

アルゴリズム・ソフトウェア・ハードウェアの3領域での密接な連携

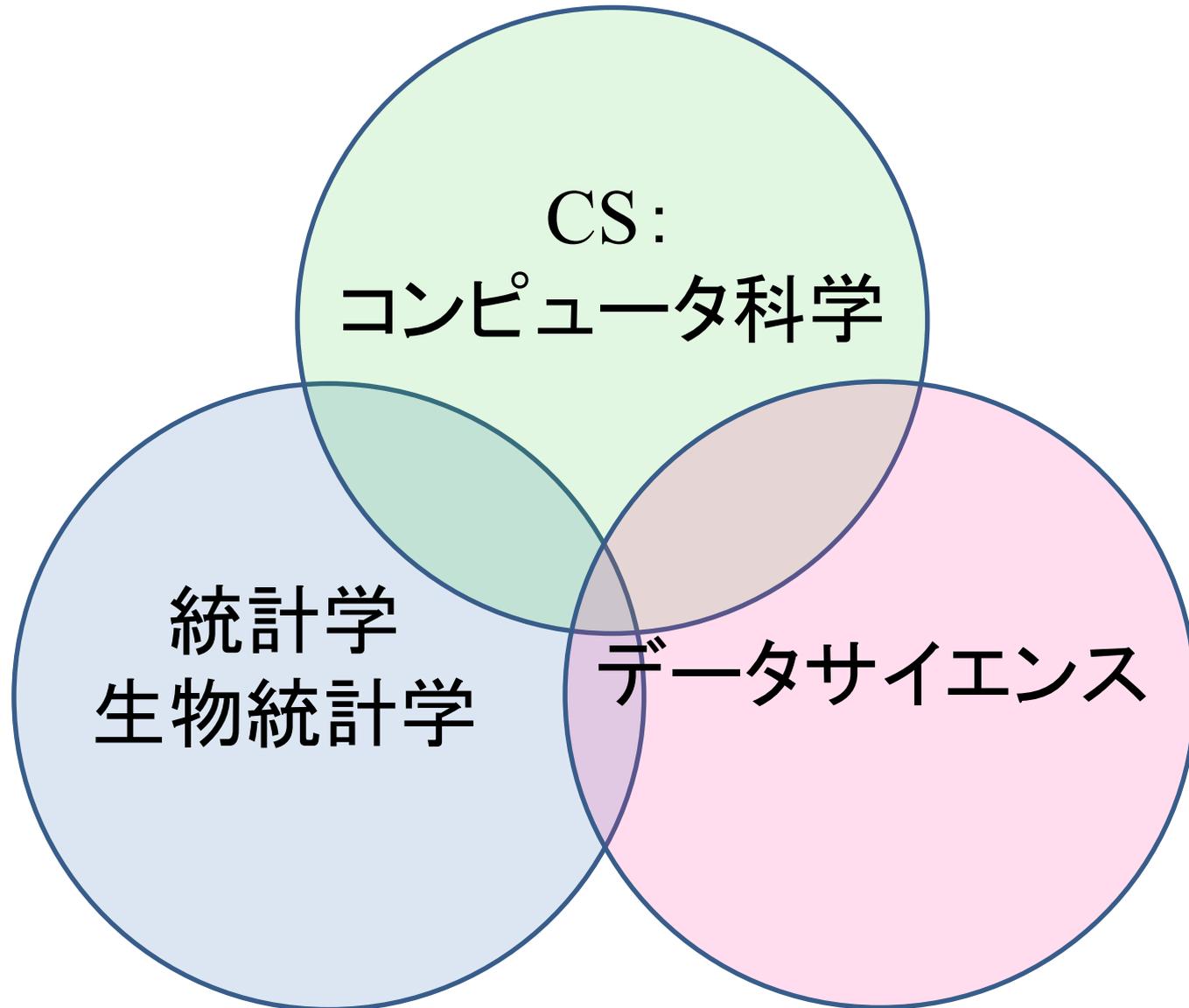


③高等教育：データに関連した数理分野の俯瞰図



2015/05/06 作成者 樋口知之

学部、研究科：世界と日本の歴史の違い



1839年

ナイチンゲール、
グラハム・ベル、
カーネギー

アウトライン

- i. 1980年代: ベイズ統計
- ii. 1990年代: 粒子フィルタ
- iii. 2000年代: データ同化
- iv. 2010年代: 機械学習とAI
- v. 2020年代: そしてこれから
- vi. さいごに

モデルとモデリング(再掲)

- 常にモデルを改良する姿勢
- ものの見方、捉え方を柔軟に変化
- 思考のプロセス自体を科学する

モデルを比較し選択するための羅針盤は必須
→ 情報量規準

究極的真理とモデル

構造が確認されている確率的な機構という特殊な場合を除き、期待の構成の仕方は我々の持つ知識や経験の使い方に大きく依存する。したがって、唯一無二の真の構造のようなものは存在しない。.....したがって、我々はより良いモデルの探求を通じて、常に未知の状態にある究極的な真理あるいは真の構造に迫るのである。

これは、極めて基礎的な科学的認識で、しばしば究極の真理と捉えたものでも、人類のもつ経験の累積にともないその内容が常に深化してきたことを考えれば、箱を開けてみればその内容が確定されるというような単純な事柄に関する真偽の議論の場合を除き、我々が追求する真理は、現在の知識に依存するという意味で相対的な、対象のひとつの近似を与えるモデルによって表現されるようなものに過ぎない、....

赤池弘次 「時系列解析の心構え」、朝倉書店(1995)

機械学習vs.統計的(ベイズ)モデリング

JSM2019 David van Dyk (Imperial College)

<https://www.imperial.ac.uk/~dvandyk/>

Astrostatisticsの専門家
LIGO (重力波)、ブラックホール撮像

- Data-driven vs. Science-driven
- Predictive model vs. Descriptive model
- Correlation vs. Causality

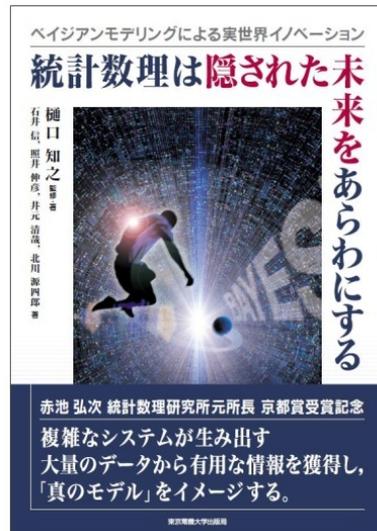
「認識科学」から「設計科学」へのシフト

「対象理解」から「機能の最適化」

へ興味が生じる

「真のモデル」

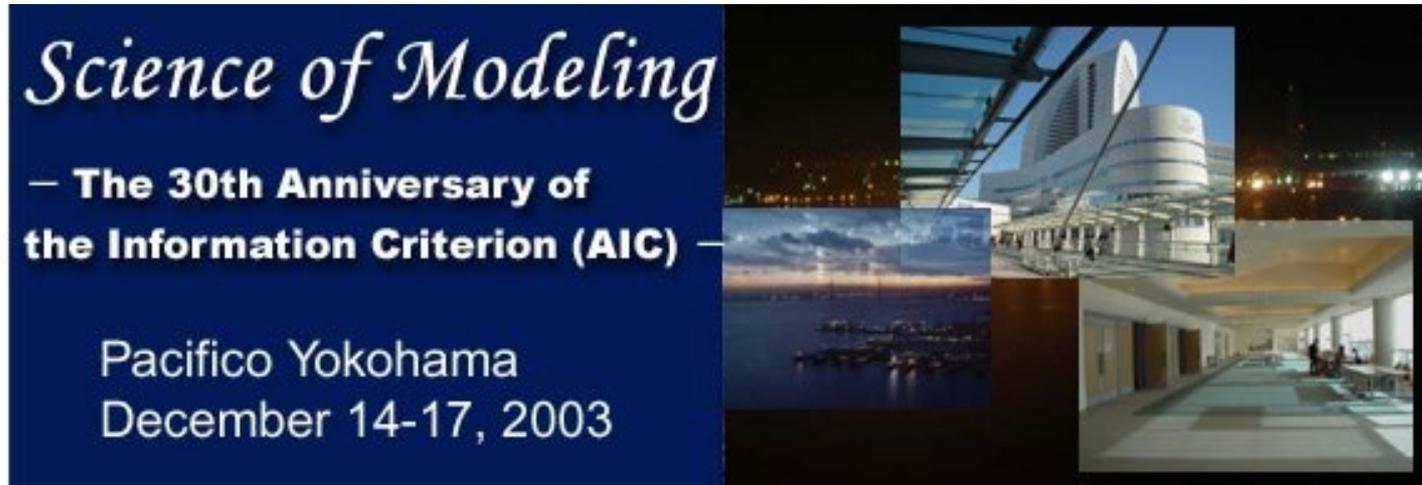
イメージ無くして理論の有効適用はありえない。「真のモデル」は、
当面の問題に対する最適なイメージを表現するモデルである。
このように考えれば、「真のモデル」など存在しないという見方は、
当面の問題のイメージの欠如の表現とも受け取られる。



井元 赤池先生
東大教授
樋口
石井 京大教授
照井 東北大教授

赤池弘次 「刊行によせて—AICとベイズモデル」、2007年 6月

Science of Modeling の創設



甘利先生

甘利先生
堀田先生

樋口
田邊先生
北川先生
赤池先生

赤池先生
経済評論家
田中直毅氏

赤池先生
H. Tong教授

平成の30年間はモデリング技術革新の30年

そして私にとって、平成の30年間は素晴らしい先生方との出会いの30年であった