"Materials research by Information Integration" Initiative
情報統合型物質・材料開発イニシアティブ
http://www.nims.go.jp/MII-I/index.html

since 2015

マテリアルズ・インフォマティクスの狙い

物質・材料研究機構　名誉フェロー・エグゼクティブアドバイザー
寺倉清之

マテリアルズ・インフォマティクスの現状と将来展望
2018/1/29

JSTイノベーションハブ構築支援事業 (H27-H31)

蓄電池材料 Gr
GL: 中山
副GL: 館山

1. 全固体電池
2. 多価イオン電池

磁石材料 Gr
GL: 三宅
副GL: 木野

1. 希土類磁石
2. 非希土類磁石

伝熱制御・熱電材料 Gr
GL: 徐
副GL: 塩見、篠原

1. 熱伝導制御
2. 熱電特性

マテリアルズ探索 Gr
GL: 小口
副GL: 中村

構造探索、相安定性、機能探索

トポロジカル解析 Gr
GL: 平岡
副GL: 赤木

トポロジー的解析による、物質・材料の構造に関する特徴抽出

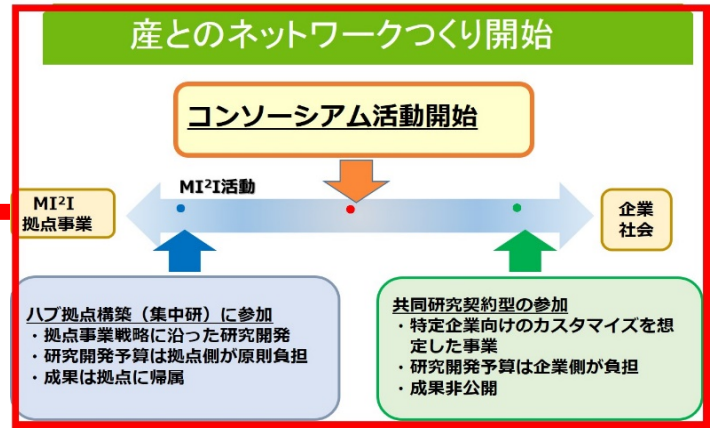
データ科学 Gr
GL: 津田
副GL: 田村

情報統合型物質・材料開発のための、方法論の開発、ツールの開発・整備

物質・材料記述基盤 Gr
TL: 吉田
副GL: 小山

物質・材料記述子: 収集、開発、整備、活用手法

H29.9 / 51社
コンソーシアム活動



統数研

北陸先端大

再委託先

データプラットフォーム

1. マテリアルズ・インフォマティクスとは何か？

2. データ科学のアプローチにおける
   記述子と目的変数

3. ベイズ推定の意義と適用例

   物理モデルをデータ解析に組み入れやすい。

   データが少なくても、存在するデータを活かした推定ができる。

4. 「少ないデータの問題」への対応
   転移学習（+ multi-fidelity model）

5. 予測と理解

物質・材料開発 Ⅰ

1957 BCS理論

その後、物性物理の終焉説が起こった。

その後の新物質発見の例

1976 カーボンナノチューブ（遠藤守信）
1985 フラレン
1986 銅酸化物超伝導体
1991 カーボンナノチューブ（飯島澄男）
2005 トポロジカル絶縁体の理論提案
2006 鉄系超伝導体
2007 トポロジカル絶縁体の実験的存在確認

これらの新物質発見が
基礎的な物性発現の機構
の提案と実証という
活発な物質科学の活動
の展開に繋がっている。

これらは、経験、洞察、ひらめき、偶然、などなどによる新奇物質発見

今のところ、データ科学的な研究とは明らかな繋がりをつけにくい。

・1/25 までの見解
・それ以後は、後述のように
　少し変わった。

物質・材料開発　II

科学技術の進歩に後押しされた物質開発

この場合は、目標とする物性や機能が明確なものが多い。

物質設計

物質設計は、本来逆問題なので実現が困難

＊Design principle によるスクリーニング

＊データ科学の活用
機械学習

(bio-informatics, cheminformatics, materials informatics)

課題はSTAGEで異なる

真鍋明氏作成

	STAGE I 新物質創成 例） C_{60}	STAGE II 物性極値化 Materials Genome	STAGE III 材料最適化 Integrated Computational Materials Engineering	STAGE IV 適用研究開発
内容	従来の特性限界超 物質探索	結晶構造あり 元素置換　ドープ 極値を探す	材料化 プロセス・組織構造 の最適化	システム設計 試作実証 信頼性確保
ポイント	コンセプトひらめき 実験発見 Abduction	傾向予測と実験 Deduction Induction	実験検証 特性ﾄﾚｰﾄﾞｵﾌ克服 Induction主体	Virtual　Prototype シミュレーション Deduction
データ共有	Unknown OPEN　知識	OPEN　一部CLOSE 物質データベース	特性CLOSE/OPEN プロセスCLOSE	固有材料CLOSE 一般材料OPEN
MIの期待	逆問題 特性→構造予測	結晶構造・特性相関 QSPR	特性・組織相関 プロセス・組織相関	短期間化 （時間、費用）
課題	**方法論研究**	**事例研究 手法選択**	**組織構造データ化 データ形式統一** （メタデータ）	**各種シミュレーション**

今後 ← 現状での主目的 → 今後

レシピ＋支配方程式

出口

計測インフォマティクス

計測の自動化
大規模実験施設　SPring8, J-PARC

STEM-EELS, Atom Probe Tomography,
SAXS, SANS

大型スパコン　京、ポスト京
不成功データの蓄積6

既にbig data の時代

・big data を活用した物質探索
・**実験の効率化**
big data の自動解析

JST CREST, さきがけ

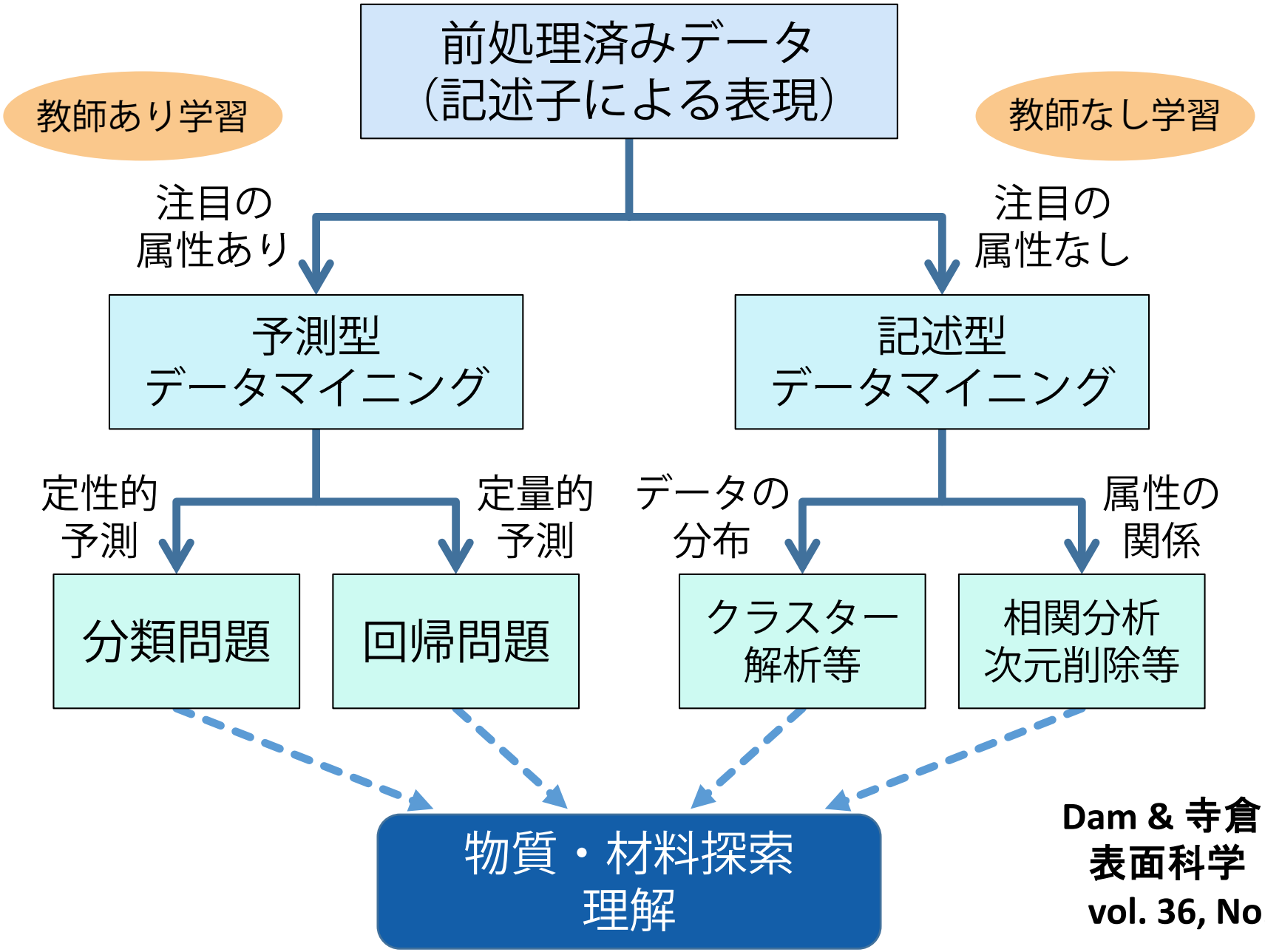
計測技術と高度情報処理の融合による
インテリジェント計測・解析手法の開発と応用

情報統合型研究の3つの主要目標

1. 解析から予測へ、予測から設計へ
   解析型研究から開拓型研究へ
2. 複雑な現象や機能を制御している原因の解明
3. 実験・計算における解析型研究の高速化

蓄積したデータからの学習に基づく逆問題対応

個々の解析への対応
特に、計測インフォマティクス
ただし、これもデータ科学で
上記の1, 2 と同じ。



Dam & 寺倉
表面科学
vol. 36, No.10, p.507 (2015)

回帰解析

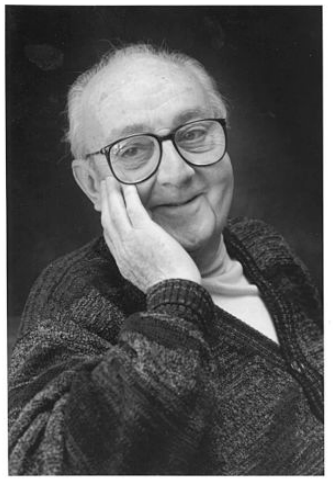
機能: $F = f(x_1, x_2, \cdots)$ (1)

記述子: $x_1, x_2, \cdots$ (2)

目的の機能Fと記述子$(x_1, x_2, \cdots)$の間の関係式(1)を導くだけでなく、記述子$(x_1, x_2, \cdots)$を決める。

記述子:①機能 F を制御する物理量であり、
②その機能に関して、対象の間の類似性を測るのに適している。

All models are wrong;
some models are useful.

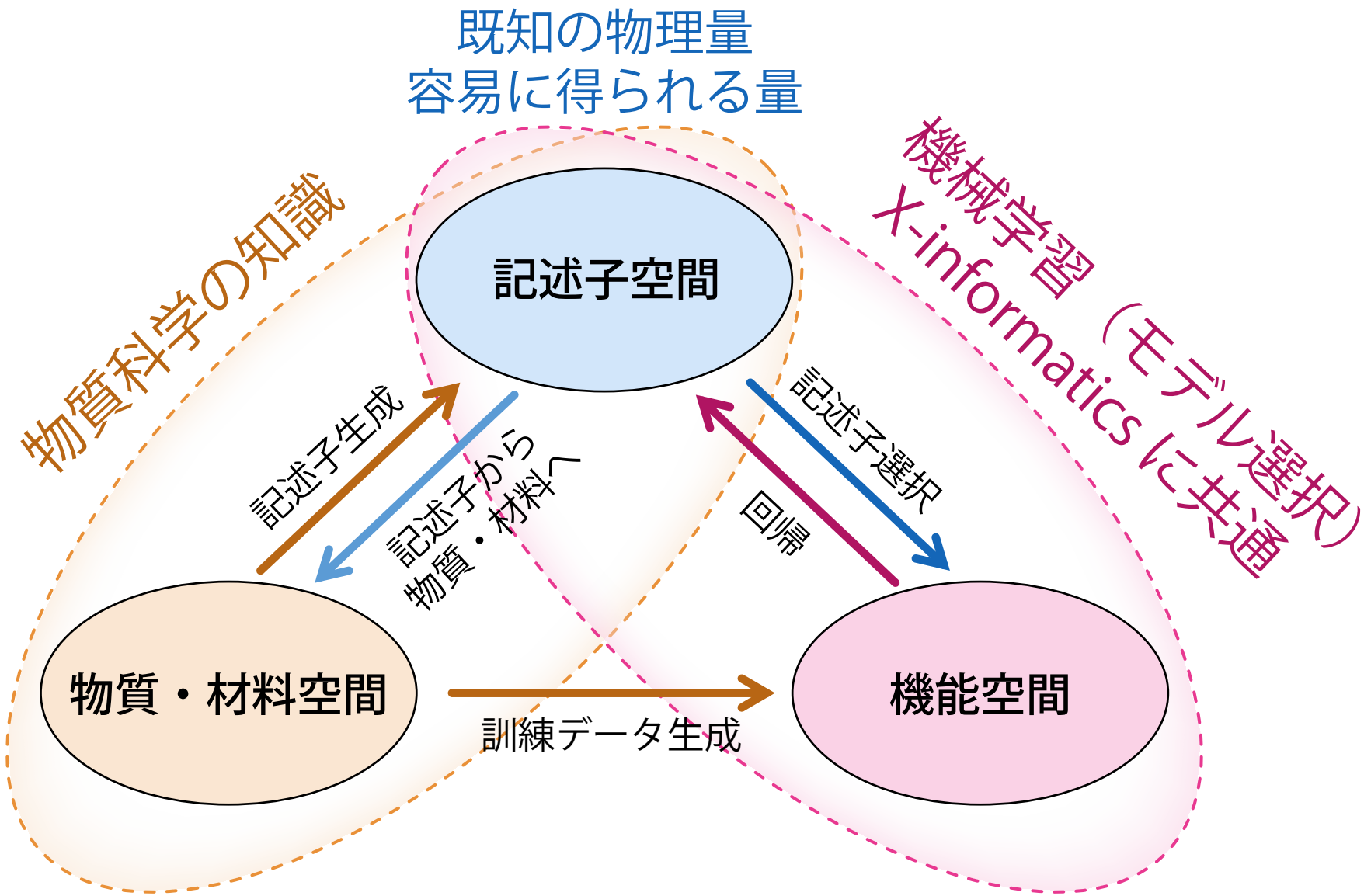


George E. P. Box

データからの回帰で得られる式(1)は、相関関係と呼ばれる。
相関関係は必ずしも因果関係ではない。 ➡ 予測の検証が必要

相関関係が、因果関係になっているかどうかの判定もデータ科学の課題。

物質・材料に機械学習を適用



現状では、
AI にはできない。

知性

知能

山本一成氏
（ポナンザ開発者）

課題の設定：研究としては常にここが最重要

教師あり学習の場合

目的変数（狙いの機能を測る量）の設定
測定や計算が容易な量が望ましい。

：後述の転移学習に関係

うまく選ぶことが成功の鍵。専門知識が必要。

記述子（目的変数を制御する量）の設定
既知、あるいは測定や計算が容易な量

記述子の整備は MI[2]I でも重要課題
第5回チュートリアルセミナーで、吉田亮氏が説明

機械学習の手法の選択

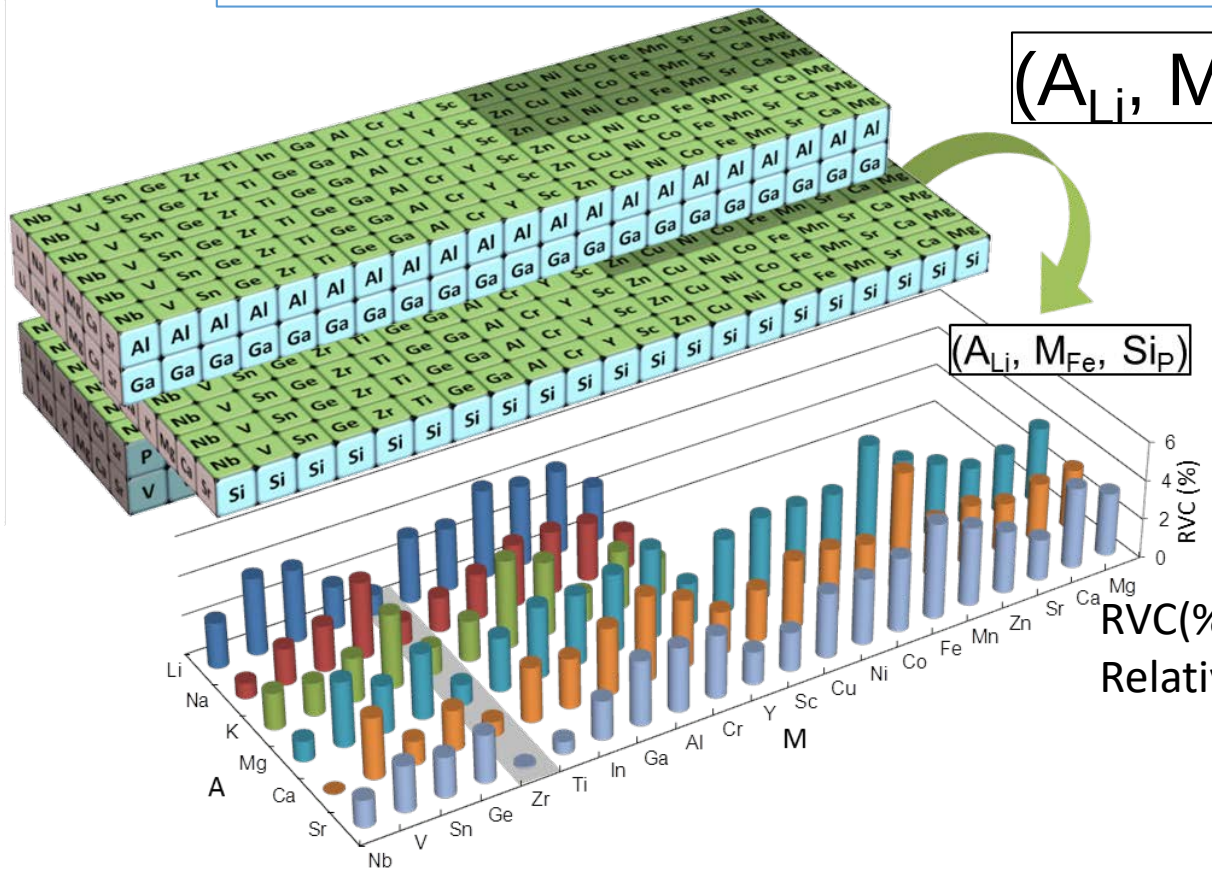
目的変数の選択の例

cosubstitution in $LiFePO_4$

京都大学
田中功　グループ

M. Nishijima et al. Nat. Commu. 5, 4553 (2014)

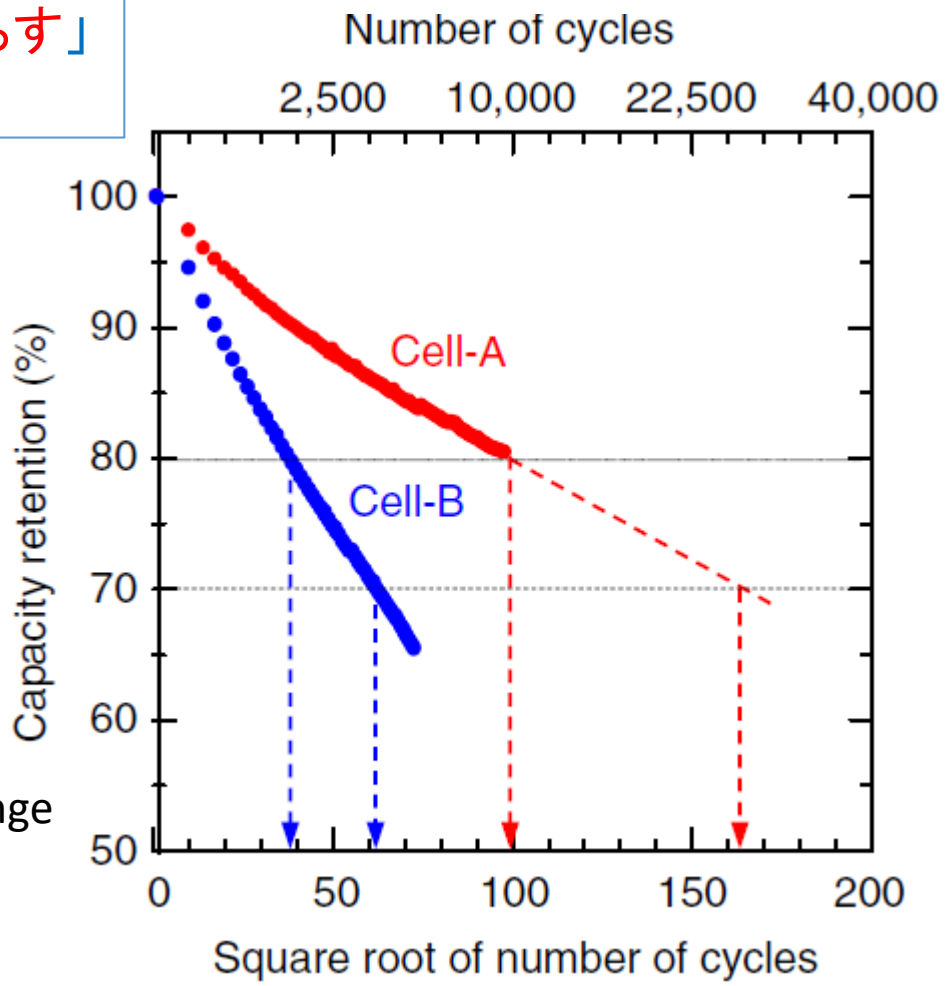
長寿命化を扱うのに、「充電時と放電時の体積変化を減らす」というデザインプリンシプルを立てた。



RVC(%):
Relative volume change

Total number of DFT calculations: 2,000

実証



寿命が6倍

形状記憶合金の劣化：マルテンサイト変態での温度ヒステリシス

Nature Commun. 7, 11241 (2016)

ARTICLE

Received 2 Sep 2015 | Accepted 4 Mar 2016 | Published 15 Apr 2016

DOI: 10.1038/ncomms11241 OPEN

Accelerated search for materials with targeted properties by adaptive design

Dezhen Xue[1,2], Prasanna V. Balachandran[1], John Hogden[3], James Theiler[4], Deqing Xue[2] & Turab Lookman[1]

the complex search space. Our strategy uses inference and global optimization to balance the trade-off between exploitation and exploration of the search space. We demonstrate this by finding very low thermal hysteresis (ΔT) NiTi-based shape memory alloys, with $Ti_{50.0}Ni_{46.7}Cu_{0.8}Fe_{2.3}Pd_{0.2}$ possessing the smallest ΔT (1.84 K). We synthesize and characterize 36 predicted compositions (9 feedback loops) from a potential space of ~800,000 compositions. Of these, 14 had smaller ΔT than any of the 22 in the original data set.

多元系形状記憶合金 Ti-Ni-Cu-Fe-Pd での温度ヒステリシス最小化
最適化組成を予測して、実証

実験とデータ科学の協働　　9回の iteration で最適化実現

最適組成合金 $Ti_{50.0}Ni_{46.7}Cu_{0.8}Fe_{2.3}Pd_{0.2}$

機械学習の手法

１．Bayes 推定
２．転移学習
に絞る。

Bayes 推定

同時確率: $P(X,Y) = P(Y \mid X)P(X) = P(X \mid Y)P(Y)$

事後確率　　尤度　　事前確率

Bayes の定理: $P(X \mid Y) = \dfrac{P(Y \mid X)P(X)}{P(Y)} \propto P(Y \mid X)P(X)$ (1)

原因　結果　　結果　原因

物質　物性　　物性　物質

逆構造物性相関　　構造物性相関

原因と結果の関係の向きを逆転できる。 ➡ 逆問題に好都合。

与えられた結果を与える原因を求めるには、事後確率を最大にする X を探すことになる。

計算機の発達により、これは MCMC (Markov-chain Monte Carlo) 法などによって解かれるようになった。

Bayes 推定の特徴

文献

E. T. Jaynes: Confidence intervals vs. Bayesian intervals (1976), http://bayes.wustl.edu/etj/articles/confidence.pdf
T. J. Loredo: The promise of Bayesian inference for astrophysics (1992)
http://www.astro.cornell.edu/staff/loredo/bayes/promise.pdf
V. Dose: Bayes in five days, https://www.researchgate.net/publication/245581841_Bayes_in_five_days
Udo von Toussaint: Bayesian inference in physics, Rev. Mod. Phys. 83, 943 (2011).

データからの情報

$$\text{Bayes の定理: } \underbrace{P(X \mid Y)}_{\text{事後確率}} = \frac{P(Y \mid X)P(X)}{P(Y)} \propto \underbrace{P(Y \mid X)}_{\text{尤度}} \underbrace{P(X)}_{\text{事前確率}} \qquad (1)$$

1. 事前確率：解析対象のデータを見る前に知られている事前知識を組み込める。
   - データを解析して得られた知識を、そのデータに関する事前確率に組込んではいけない。(Dose の解説)
     これを繰り返すと、どんどん分散が小さくなる。← 過学習
   - 一つのデータの解析で得られた知識を、別の関連データの解析の事前確率に使うことによって、予測精度が著しく向上する。(Bayes 更新)
2. 尤度：データからの学習。解析対象の問題で、適切なモデルの導入が可能。

データから学習してデータを見る前の知識を改善する。

観測データのセット$\{(y_i, \vec{x}_i); i=1\sim N\}$に対する解析

1. 尤度におけるモデルの活用　パラメータ推定

パラメータ群Θを含むモデル（法則）$G(\vec{x}_i;\Theta)$により、

$$y_i = G(\vec{x}_i;\Theta) + \varepsilon_i \tag{5}$$

で表す。ε_iはランダムノイズで、分散がσ^2のガウス分布に従うとして、尤度は

ノイズのバラつきと
データそのもののバラつきは別物

$$p(y_i \mid \vec{x}_i, \Theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)\exp\left\{-\frac{1}{2\sigma^2}(y_i - G(\vec{x}_i;\Theta))\right\} \tag{6}$$

Θについての事前分布を$p(\Theta)$とすると、その事後分布は

$$p\left(\Theta \mid \{y_i, \vec{x}_i : i=1,N\}\right) \propto p(\Theta)\prod_{i=1}^{N} p(y_i \mid \vec{x}_i, \Theta) \tag{7}$$

事後分布を最大にするもの(MAP推定)。

Cu_2O 薄膜結晶の吸収スペクトルの解析

第６回チュートリアルセミナー (2017/11/1) における赤井一郎先生の講演

K. Iwamitsu, S. Aihara, M. Okada and I. Akai, J. Phys. Soc. Jpn. 85, 094716 (2016)

Cu_2O での励起子の Bose-Einstein 凝縮が期待されている。

１．励起子の寿命が長いことが望まれるので、スピン３重項にある、パラ励起子を対象にする。

２．MgOで挟まれた Cu_2O では、僅かな格子不整合により、励起子に対して引力的ポテンシャルが誘起されている可能性が高い。これを定量的に検証する。

吸収スペクトルのベイズ分光

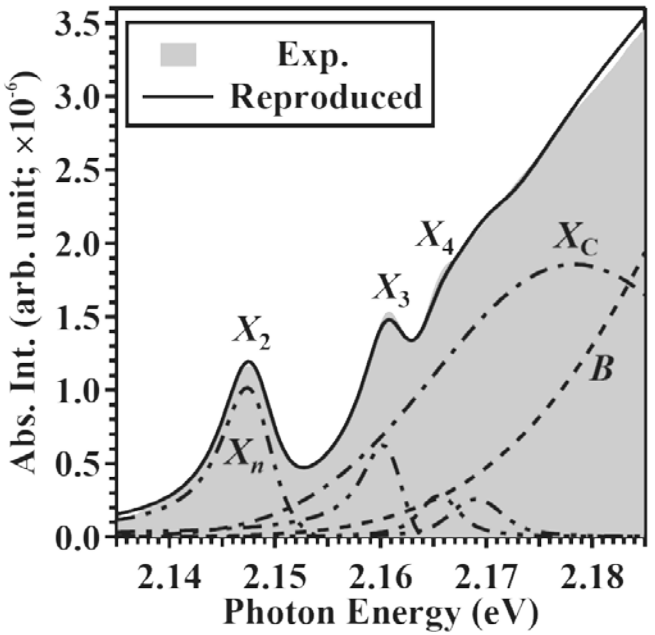
物理モデル: $G(x_i;\Theta)$

- X_n 励起子の遷移エネルギー E_n

$$E_n = E_g - \frac{R_y}{n^2}, \qquad (n = 2,3,\cdots,n_\infty)$$

- X_n 励起子の吸収強度 f_n

$$f_n = f_0\left(\frac{n^2-1}{n^5}\right), \qquad (n = 2,5,\cdots,n_\infty)$$



- 励起子ピーク: X_n $(n = 2\sim20)$
  - 非対称ローレンツ形状
    - ✓ 均一幅 Γ_2, Γ_X

$$X_n(E;E_n) = f_n \times \frac{1}{\pi}\frac{\Gamma_n + 2A_n(E-E_n)}{(E-E_n)^2+\Gamma_n^2}$$

  - 不均一幅 γ_X

Voigt関数:

$$\int X_n(E;E_n+d\epsilon)\exp\left(-\ln 2\frac{\epsilon^2}{\gamma_X^2}\right)d\epsilon$$

（Xc, B についても同様）

- バックグラウンド
  - B: 双極子禁制バンド間吸収

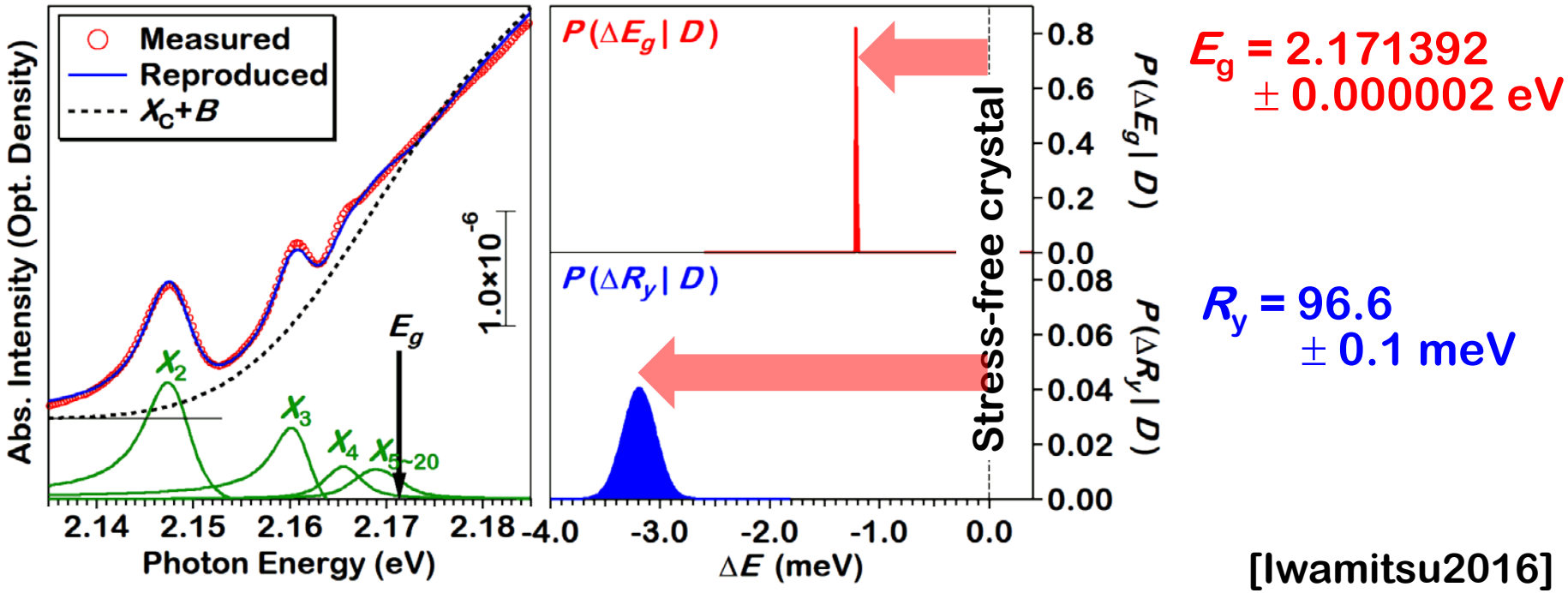
$$X_C(E) = f_{X_C}\pi\beta(1+\beta^2)\frac{\exp(\pi\beta)}{\sinh(\pi\beta)}, \quad with\ \beta = \sqrt{Ry/E-E_g}$$

  - X_C: 励起子連続帯バンド吸収

$$B(E) = f_B(E-E_g)^{3/2}$$

Burn-inの後、メトロポリス法による250万回のサンプリング

➢$\sigma_{\mathrm{noise}} = 4 \times 10^{-8}$ (吸収強度) ~ 物理モデル $G(x_i; \Theta)$ のRMSD



E_g = 2.171392 ± 0.000002 eV

R_y = 96.6 ± 0.1 meV

[Iwamitsu2016]

- 事後確率分布 $P(\Delta E_g|D)$, $P(\Delta R_y|D)$ は極めてシャープ
- E_g , R_y は明確・有意に変化！→ 励起子トラップポテンシャル形成

吉田亮

chemoinformatics:

吉田亮、池端久貴(統数研)、本郷研太(北陸先端大)

H. Ikebata, K. Hongo, T. Isomura, R. Maezono and R. Yoshida,
J. Computer-Aided Mol. Design 31, 379-391 (2017)

定量的構造物性相関 (Quantitative Structure-Property Relation: QSPR)
から
逆定量的構造物性相関 (IQSPR)
へ

分子部品の組み合わせから、自然言語処理の利用への発展

Bayesian inference による

解析・予測から設計へ

の印象的な成果

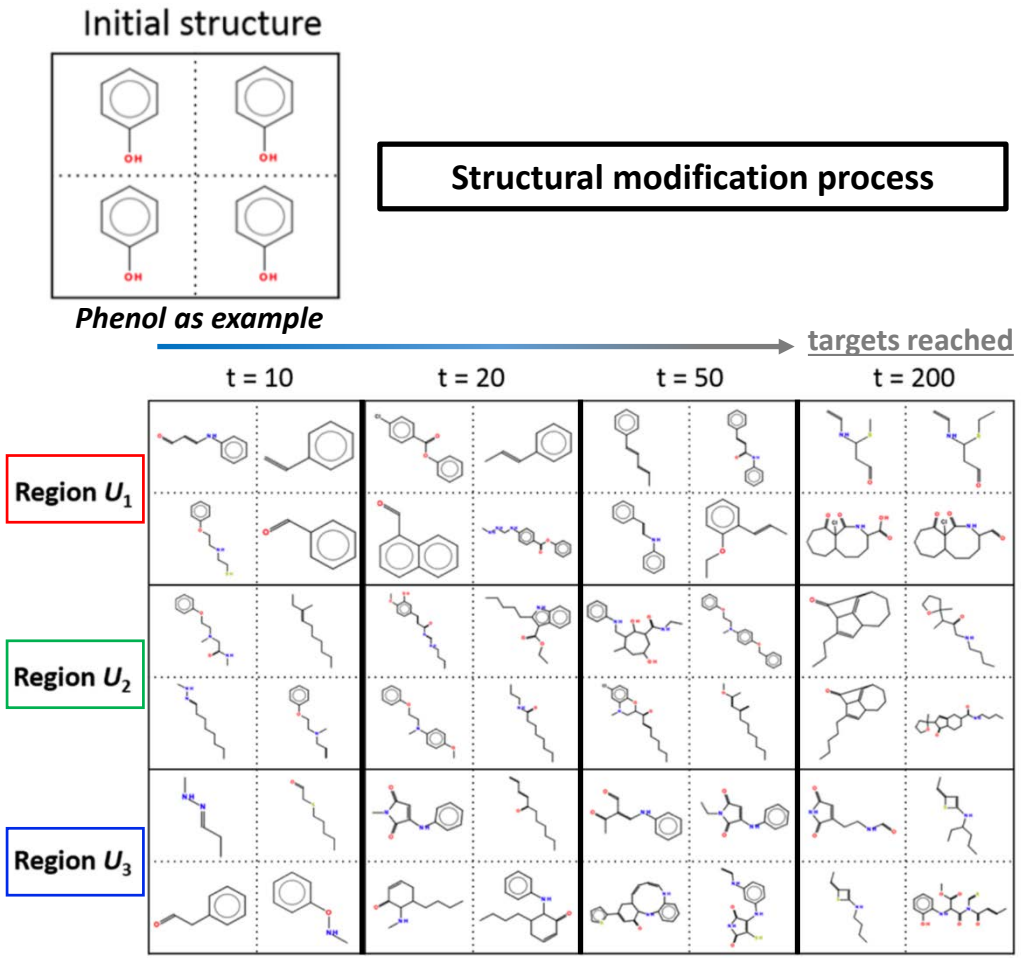
SPACIER
外挿へ

分子設計については、
津田先生等の MCTS, RNN を用いた ChemTS

DEMO: BAYESIAN MATERIALS DESIGN

- Data: 16,674 organic compounds in PubChem and DFT properties
- Properties: HOMO-LUMO gap, internal energy
- Targeted regions: U_1, U_2, U_3

SOFTWARE '*iqspr*'
https://cran.r-project.org/web/packages/iqspr/index.html



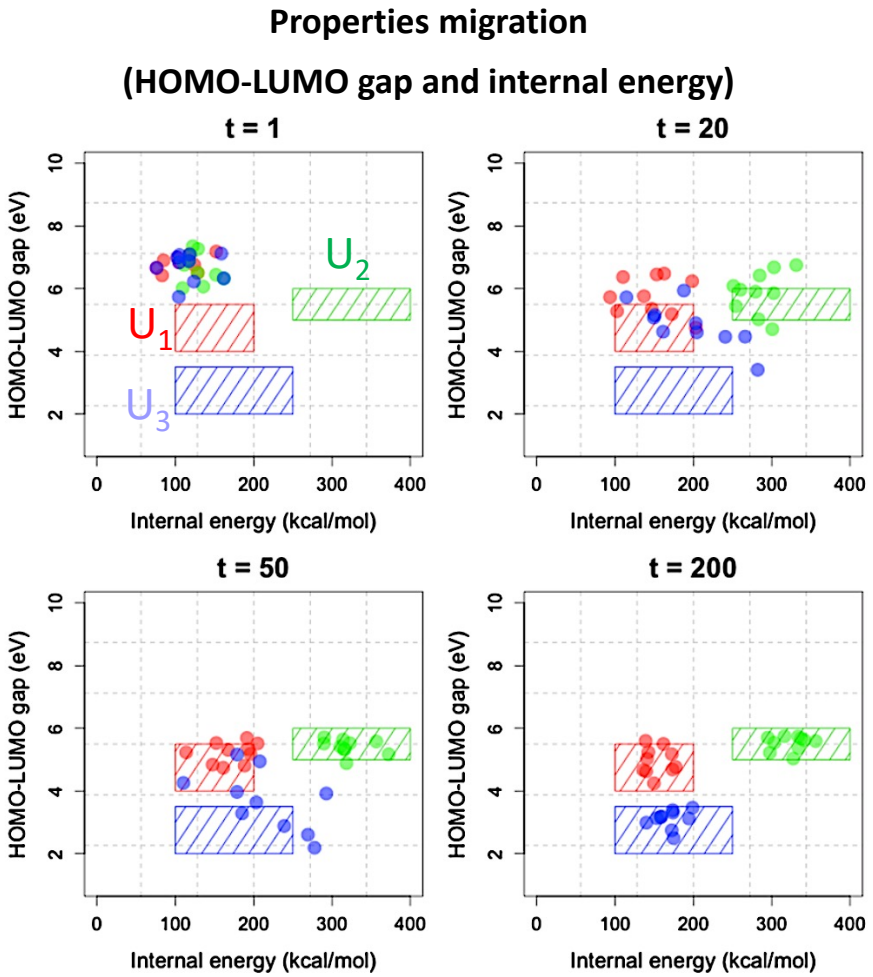


Fig. 4-a (left), 4-b (right) from Ikebata et al., J. Compt. Aided Mol. Des., 2017

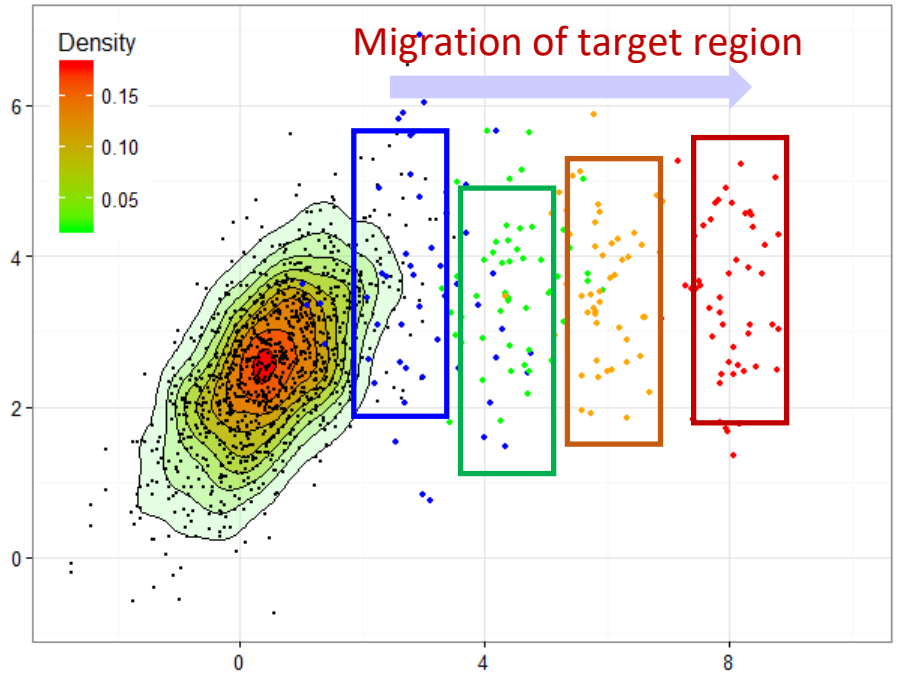
“iqspr”の拡張機能としての外挿への試み

Lambard, G
ISM

SPACIER Go Beyond Interpolative Prediction

Difficulty of jumping up to an extreme property region because of exceedingly low accuracy of QSPR models due to the lack of existing materials

Adaptive selection of the target property region by successively controlling the trajectory of created materials



Choice for the temporary target region

Acquisition function of SPACIER

$$A_{U,\lambda}(S, Y^*) = -\left\| Y^* - \hat{f}(S) \right\|^2 - \sigma(S) - \lambda \inf_{Y \in U} \| Y^* - Y \|^2$$

Design point for the DFT calculation

転移学習 (Transfer Learning)

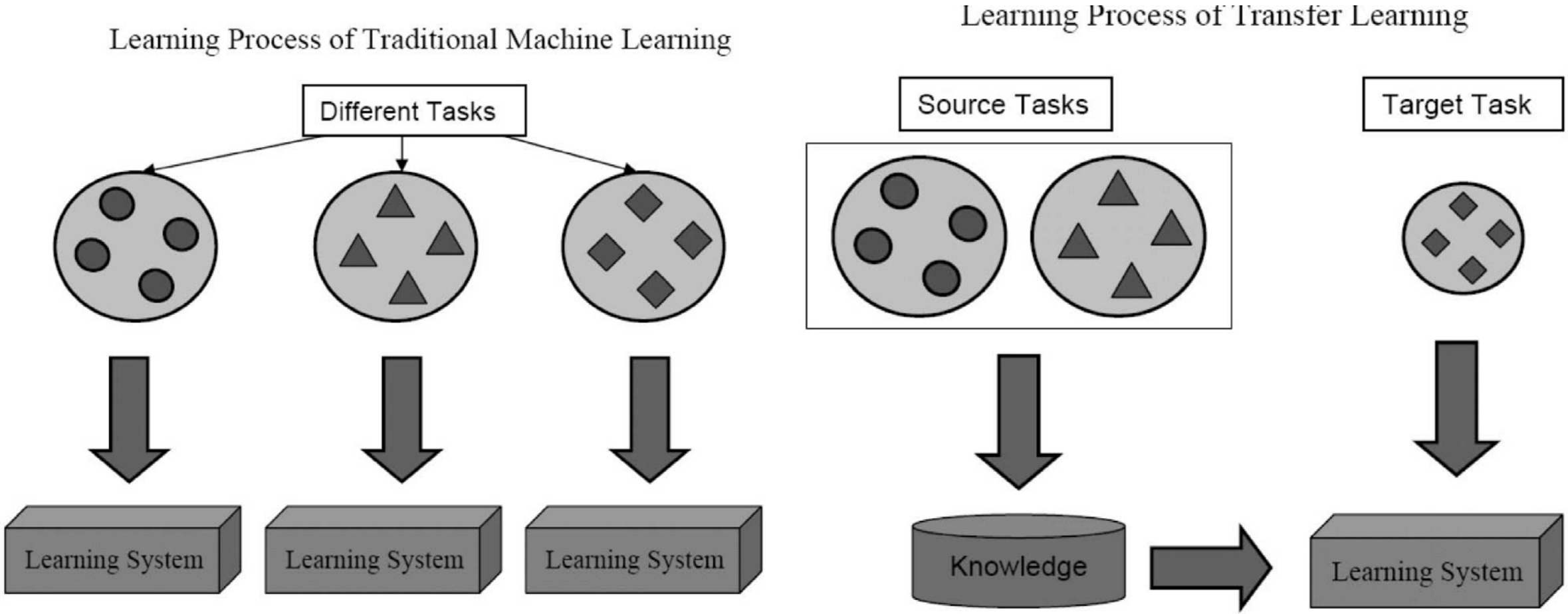
関連しているが異なる部分もあるデータから、
目的の問題にも利用できる情報・知識だけを
取り込んで、より予測精度の高い規則を得る。

神嶌敏弘　　人工知能学会誌 25巻4号 (2010).
http://www.kamishima.net/archive/2010-s-jsai_tl.pdf

転移学習が役立つ2つのケース

1. 似たような回帰のタスクを行う場合に、回帰の効率と精度を上げる。(GB の例)
   むしろ、multi-task と呼ぶほうが適切（配布物に修正）
2. ある目的とする量 A のデータが少ない場合、その量と相関があり、データ取得が容易なために沢山のデータがある量 B というような状況では、量 B の回帰結果を量 A の回帰に使う。（熱伝導の例）

データが得やすい量を通して、データが得にくい量の学習をする。



S. J. Pan and Q. Yang, A Survey on Transfer Learning,
IEEE Transactions on Knowledge and Data Engineering, 22, 1345 (2010)
https://www.cse.ust.hk/~qyang/Docs/2009/tkde_transfer_learning.pdf

最近、転移学習が物質科学の課題に適用されるようになった。

対応粒界の
安定性の議論

multi-task ?

1. H. Oda, S. Kiyohara, K. Tsuda and T. Mizoguchi,
   Transfer Learning to Accelerate Interface Structure Searches
   J. Phys. Soc. Jpn. 86, 123601 (2017)

2. T. Yonezu, T. Tamura, I. Takeuchi and M. Karasuyama,
   Knowledge-Transfer based on Cost-effective Search for Interface Structures
   arXiv: 1708.03130v1

3. M. L. Hutchinson, E. Antono, B. M. Gibbons, S. Paradiso, J. Ling and B. Meredig,
   Overcoming data scarcity with transfer learning
   https://arxiv.org/pdf/1711.05099.pdf

4. 吉田先生（統数件）らによる、
   熱伝導度の予測に、scattering phase space (SPS) のデータを活用

5. 小野先生（KEK）ら
   実験の分光スペクトルの解析に、シミュレーションの解析結果を転用

予測と理解

津田先生

（第１回JST workshop （2013/2/11) では）

機械学習の目的は次の2つ

１．原因究明

２．予測

それぞれの目的を徹底的に追及する場合は、2つは別の課題。

山本一成氏 （ポナンザの開発者）

『人工知能はどのようにして「名人」を超えたのか？』

第2章：黒魔術とディープラーニング　ー科学からの卒業ー

第1節：機械学習によってもたらされた「解釈性」と「性能」のトレードオフ

・・・

最9節：還元主義的な科学からの卒業

真鍋氏作成

ARTICLE 世界1番の棋士が人工知能に負けた

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

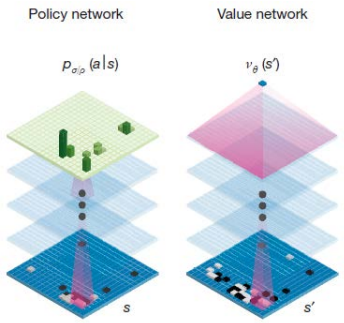
10^{360}

人工知能技術

・**Deep Learning**

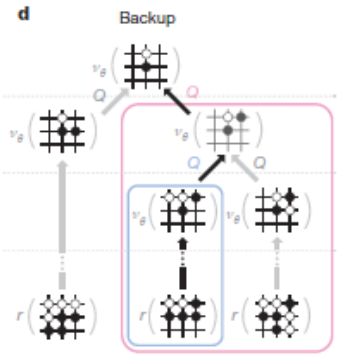
報酬関数　F(p,v)

p; policy network
v; value network



機械学習

・**強化学習**
・**モンテカルロ木探索**



棋譜データベース

16万棋譜

3000万局面

計算パワー

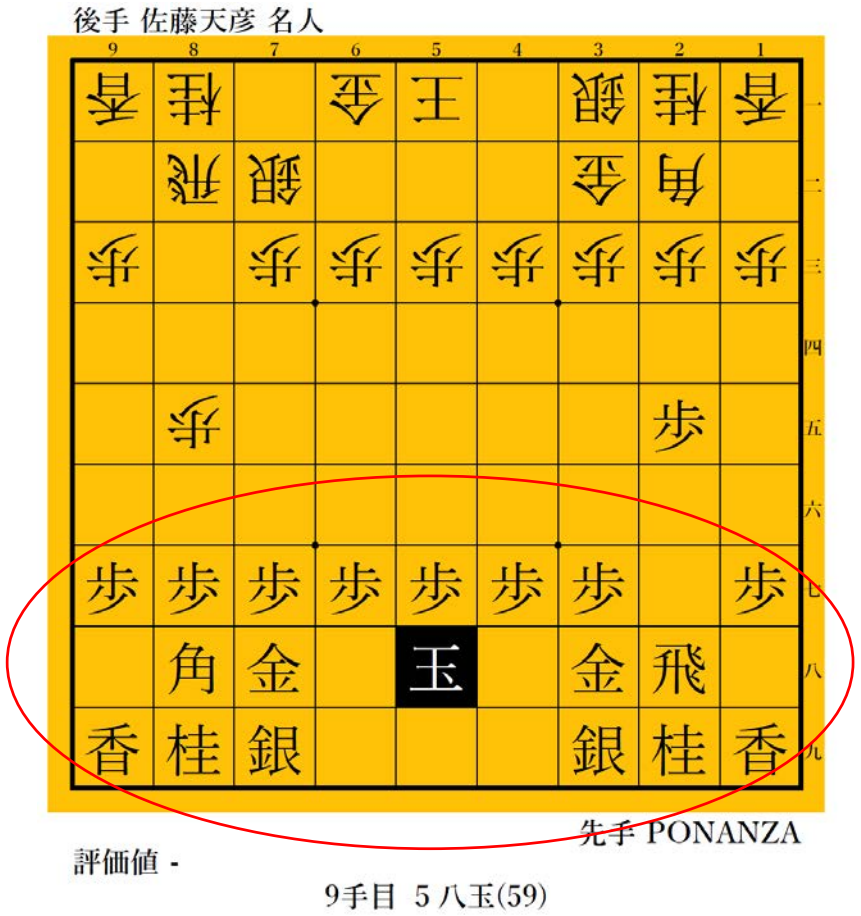
Google Cloud
1202CPU
176GPU

人間は絶対に指さない1手目

1手目　3八金

羽生義治
「人間は絶対指さない手ですね。」

中住居に近い形



羽生義治
「この局面になると、・・・『3八金もある手ではないか』と思えるところが、一つの驚きでもあります。」

佐藤天彦名人　対　ポナンザ
第1局　2017年4月

1手目は結局は正しい予測だったが、通常の理解を超えていた。

中住居に近い

人工知能の「最適解」と人間の選択：NHK 出版新書

Bayesian inference

少数データへの適応性と解釈性
不確実性への対応

「ひらめき」、「勘」につながる？？
STAGE I の課題

Bayesian Deep Learning：
deep learning への不確実性の取り込み
認知（deep learning) ＋推論 (Bayesian inference)
・・・・

Deep Learning

big data と 解釈不可能性
決定論的

過去の対局の碁譜のデータを入力　　碁譜のデータ入力不要

囲碁：アルファ碁、あるいは Master

将棋：Ponanza, あるいは elmo

これらは、人工知能（機械学習）の到達点の例になっている。
強化学習により、データ数が小さいことが問題にならなくなった。

マテリアルズ・インフォマティクスへの水平展開を。
　囲碁での探索法（MCTS）のプログラムは津田グループで開発済み