

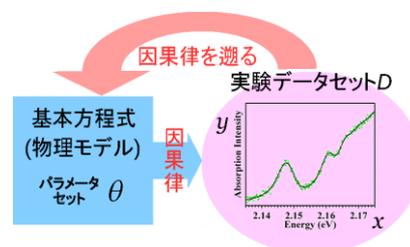
# データ駆動科学の物性科学への適用と、放射光計測への展開

赤井 一郎

熊本大学 パルスパワー科学研究所

2016年3月 Google の囲碁 AI 「AlphaGo」 が、人類最強と言われる棋士を破ったニュースを覚えておられるでしょうか？ この開発者である D. Hassabis は関連した対談で、CERN<sup>1</sup>に関して次の様な発言をしています。「I think it'd be cool if one day an AI was involved in finding a new particle.」<sup>2</sup>これは CERN の膨大なデータ解析に AI を適用すれば、未発見の素粒子を発見出来る可能性について述べたもので、AI for science の重要性を述べていると理解できます。

これまで物性研究では、因果律に基づいて実験データを解析する手法が取られてきました。最小二乗法がその典型で、因果律の元になる物理モデルとそれに含まれる物理量が「確定している」、因果律(図の右矢印)に従って「確率的に」実験データが得られると考えます。データ駆動科学はその逆の立場に立ち、ベイズ推定により因果



律を遡ります。つまり測定されたデータは測定終了時に「確定している」と考え、重畳するノイズを仮定した上でベイズの定理を使って、物理モデルの物理量が「確率的に決まる」と考えます。このベイズ推定では、たった1つの測定データセットから物理量の分布が推定可能で、高度なベイズ推定法を用いると、例えば多峰性スペクトルで、一切先入観なしにピークがいくつ含まれているかまでも推定(モデル選択)することが出来ます。更に、例えば時間的振動するデータ波形に対して、そのデータに含まれる振動周波数成分を分解するためフーリエ変換がよく用いられますが、観測時間の逆数程度以下の精度で周波数を特定することは出来ません。しかし振動波形をモデル化したベイズ推定により、極めて高精度に周波数を推定することが可能になります。

データ駆動科学はこの様なベイズ推定と、AI 技術へと繋がるスパースモデリング (SpM)<sup>3</sup>の枠組みがあります。熟練した実験家は、一見複雑に見えるデータを眺めていると、その中に含まれるデータ要素(スペクトルならピーク構造、振動波形なら振動成分)を「直感的に」見抜くことができる場合があります。SpM はそれを自動化するもので、データに内在する特徴構造で重要な主要成分を数学的な手法を用いて抽出することを可能にします。SpM は、物性研究はもとより、ブラックホール等の天体観測<sup>3</sup>、生体の CT 画像解析<sup>3</sup>等に適用されつつ有り、これからの発展が期待され、活発な研究が進められています。

講演では、光物性分野の研究でこれまで取り組んできたベイズ推定の例や、フーリエ変換に取って代わる新規解析法としての SpM の例を紹介します。また、ベイズ推定や SpM を放射光計測に組み合わせることによって期待できることについても触れたいと思います。

<sup>1</sup>欧州原子核研究機構：世界最大規模の素粒子物理学の実験施設

<sup>2</sup><http://www.theverge.com/2016/3/10/11192774/demis-hassabis-interview-alphago-google-deepmind-ai>

<sup>3</sup><http://sparse-modeling.jp/>

**IPPS** データ駆動科学

## データ駆動科学の物性科学への適用と、放射光計測への展開

熊本大学 パルスパワー科学研究所  
赤井 一郎

第10回九州シンクロtron光研究センター  
研究成果報告会 @ サンメッセ鳥栖  
2016年8月3日(水)

SAGA-LS 2016/08/03 1

**IPPS** データ駆動科学

最新の情報科学の手法をデータ解析に導入して  
既存のボトルネックを解消し、新しい研究展開を開く。

**実験屋こそ取り組むべき** (データ解析の価値観が変わる)

- データを骨までしゃぶる
  - 間接測定される物理量の統計的精度を評価
  - 物理モデルの妥当性の評価
  - 既存の解析法を「嚙呑み」にしていなか、それが Best solutionか?
- 実験の最適化
  - 極限測定で、必要な精度を得るための実験条件は?
  - 実験を行う前の事前シミュレーションによる実験の効率化
- 理論研究へのフィードバック
  - 解析に用いる物理モデルの良し悪しを判定

SAGA-LS 2016/08/03 2

**IPPS** データ駆動科学

- データを骨までしゃぶる
  - 間接測定される物理量の統計的精度を評価
  - 物理モデルの妥当性の評価

J. Phys. Soc. Jpn., in press (2016).

MgO基板に挟まれたCu<sub>2</sub>O薄膜結晶(励起子量子凝縮相)  
格子歪由来の励起子トラップポテンシャル形成?  
E<sub>g</sub>の精度良い決定が必要

2P Yellow Series  
Absorption Intensity (arb. unit)  
Photon Energy (eV)

Band Gap Energy (eV)

Rydberg Constant (meV)

$$E_n = E_g - \frac{R_y}{n^2}$$

E<sub>g</sub> = 2.171392 eV ± 2 μeV  
R<sub>y</sub> = (93.6 ± 0.1) meV

[1] Eur. Phys. J. B. 86, 194 (2013). SAGA-LS 2016/08/03 3

**IPPS** データ駆動科学

- データを骨までしゃぶる
  - 間接測定される物理量の統計的精度を評価
  - 物理モデルの妥当性の評価

既存の解析法を「嚙呑み」にしていなか、それが Best solutionか?

どのモデルで解析すべきか?  $y = ax + b$ ,  $y = ax$ ,  $y = b$

情報学的自由エネルギー

F(y=b)  
F(y=ax)  
F(y=ax+b)

SAGA-LS 2016/08/03 4

**IPPS** 測定行為: ベイズの定理

- ベイズの定理

欲しいものはこれ!

$f(x_i; \theta)$	$\theta$	$y_i$
誤差論	確定的	確率的
ベイズ推定	確率的	確定的

$$P(\theta, D) = \underbrace{P(D|\theta)}_{\text{測定行為}} P(\theta) = \underbrace{P(\theta|D)}_{\text{測定が済めば確定値}} P(D)$$

パラメータセット  $\theta$   
 $\vec{r}, N(r), \dots$

実験データセット  $D = \{x, y\}$

因果律

SAGA-LS 2016/08/03  
図出典: Elements of Modern X-ray Physics 2<sup>nd</sup> Ed. 5

**IPPS** データ駆動科学の最初の一步

- これまでの誤差論  
「 $\theta$ は真値で確定されていて、データがノイズで確率的に与えられる」と考える。
- ベイズ推定  
「データが確定されていて、 $\theta$ が確率的に決まる」と考える。

$f(x_i; \theta)$	$\theta$	$y_i$
誤差論	確定的	確率的
ベイズ推定	確率的	確定的

$$P(\theta|D) \propto \exp\left(-\frac{N}{\sigma_{\text{data}}^2} E(\theta)\right) P(\theta)$$

事後確率

事前確率

因果律

パラメータセット  $\theta$   
 $\vec{r}, N(r), \dots$

実験データセット  $D = \{x, y\}$

因果律

図出典: Elements of Modern X-ray Physics 2<sup>nd</sup> Ed.

★ データ駆動科学 = データを出発点と考える

SAGA-LS 2016/08/03 6

### IPPS ベイズ推定を実装するアルゴリズム

- グリッドサーチ (全パラメータの総スキャン)
  - 例えば5つのパラメータで有効桁2~3桁、 $P(\theta|D)$ の計算が1 ms かけるとすると、2,500年 by (i5-CPU)
- マルコフ連鎖モンテカルロ法 (MCMC: Markov chain Monte Carlo methods)
  - P. J. Green, *Biometrika* **82**, 711 (1995).
- メトロポリス法 (Metropolis-Hasting method)
  - N. Metropolis, et al., *J. Chem. Phys.* **21**, 1087 (1953)
  - C. Andrieu, et al., *Machine Learning* **50**, 5 (2003).
- レプリカ交換モンテカルロ法
  - R. H. Swendsen, et al., *Phys. Rev. Lett.* **57**, 2607 (1986).
  - K. Hukushima, et al., *J. Phys. Soc. Japan* **65**, 1604 (1996).
  - K. Nagata, et al., *Neural Networks* **28**, 82 (2012).

SAGA-LS 2016/08/03 7

### IPPS Virtual Measurement Analysis PESの仮想計測解析

- ベイズ推定による計算機実験
  - 必要精度を得るために必要な積算量は?  $\sigma_{data}$

SAGA-LS 2016/08/03 8

### IPPS LASSO法 (Least Absolute Shrinkage and Selection Operator)

1次ノルムを、非スパース性を示すペナルティ項として付加して最小化

R. Tibshirani, *J. Royal Stat. Soc. B* **58**, 267 (1996).

$$E(\theta) = \frac{1}{2N} \sum_{i=1}^N [y_i - f(x_i; \theta)]^2$$

非スパース性を示すペナルティ項

$$\hat{x} = \arg \min_{\vec{x}} \left( \|\vec{y} - \vec{A}\vec{x}\|_2^2 + \lambda \|\vec{x}\|_1 \right)$$

λ: スパース性の制御パラメータ

- λ = 0: 最小二乗法と同じ解
- 0 < λ < ∞: 有限個のスパース解
- λ = ∞: 主要素が0個 = データ全てがノイズ

SAGA-LS 2016/08/03 9

### IPPS 1次ノルムのペナルティ項でスパース解を得る (K=1)

- 簡単な例
 
$$\begin{pmatrix} \vec{y} \\ \lambda \end{pmatrix} = \begin{pmatrix} 2 & -1 & 3 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$x_1 \neq 0, x_2 = 0, x_3 = 0$  の場合

LASSO法  $f(\vec{x}, \lambda)$  の極小点

$$\hat{x} = \arg \min_{\vec{x}} \left( \|\vec{y} - \vec{A}\vec{x}\|_2^2 + \lambda \|\vec{x}\|_1 \right)$$

SAGA-LS 2016/08/03 10

### IPPS EXAFSのSpMフーリエ解析

- EXAFS
 
$$\chi(k) k^3 \propto \sum_j N_j t_j(k) \frac{k^2}{R_j^2} \exp \left[ -2 \left( k^2 \sigma_j^2 + \frac{R_j}{\Lambda} \right) \right] \sin(2kR_j + \delta_j(k))$$

$$\hat{x} = \arg \min_{\vec{x}} \left( \|\vec{y} - \vec{A}\vec{x}\|_2^2 + \lambda \|\vec{x}\|_1 \right)$$

SAGA-LS 2016/08/03 11

### IPPS まとめ

- ベイズ推定
  - 特徴要素の高精度推定と分離
  - モデル選択
- スパースモデリング(SpM)
  - スパースに主要素・特徴構造の抽出
- 仮想計測解析(VMA)
  - 計測のシミュレーション
  - 高難易度計測の最適・効率化

汎用性は極めて高く、放射光計測との有機的連携が期待できる。

SAGA-LS 2016/08/03 12