

# 研究報告書

## 「物性測定における『熟練』と機械学習の統合的アプローチ」

研究タイプ: 通常型

研究期間: 平成 27 年 12 月～平成 31 年 3 月

研究者: 中島 千尋

### 1. 研究のねらい

情報統計力学の知見を最先端の実験データ処理に適用し、トップレベルの材料科学実験技術に機械学習手法を導入することで、飛躍的な精度と画期的な情報抽出を実現する。特に、実験研究者の斬新なアプローチや、複雑な手順へ熟練を十分に活かしつつ、それを最大限に補助するように機械学習を精選して導入・融合する形で、新規な基盤的手法を確立する。それをもって、日本独自のマテリアルインフォマティクス形態を提案する。

新規な基盤技術のための萌芽的な実例を、以下の 3 つの研究課題を通して積み上げる。

課題[1] 実験機器の画期的利用法と圧縮センシングの併用による物質構造測定

課題[2] 複雑な手順を要する画像解析に対する機械学習の適用による自動化

課題[3] データ解析を用いた橋渡しによる複数の測定技術の統合

課題[1]は実験研究者の斬新な発想を活かし、機械学習の導入によって実用レベルの手法へと引き上げることを目指すもの、課題[2]は複雑・精密な手順への熟練を機械学習で再現することで自動化・効率化を目指すもの、課題[3]は従来個別の実験手法であったものを機械学習によって円滑に補完しあうことで統合的な新たな手法となすことを目指すものである。いずれも実験科学者の熟練を活かしつつ新規な手法を確立するための、それぞれ別の角度からのアプローチである。

### 2. 研究成果

#### (1) 概要

HAADF-STEM 電子顕微鏡による金原子ナノクラスターの観測に圧縮センシングを導入し、3 次元の原子配置の再構成に取り組んだ。電子顕微鏡による複数の撮像を利用した 3 次元構造の再構成は近年研究され始めた課題であるが、現在の電子顕微鏡は CT 的な利用法を想定して作られておらず、多方向からの観測には大きな手間がかかる。また電子線照射の繰り返しによる観測フィールドのコンタミネーションならびに観測試料そのものの劣化を伴うため、多数の（逆行列の構成に十分な）観測データを得ることは実験原理的にも難しい。そこに少数の撮像からの高精度再構成が可能なアルゴリズムの導入により上記の汚染や劣化の軽減をはかり、2 万原子程度のクラスターに対して 3 枚の撮像からの再構成結果を得た

再構成結果を吟味する過程で、結果の信頼性を評価する手法を開発する必要があった。圧縮センシングを素朴に実行して得られる結果は点推定によるものとなり、そのままではエラーバーを伴わない。そこで、観測データからのリサンプリングにより多数の再構成結果を生成し、その統計的性質（主に分散）から再構成の正否を判別する手法の開発に着手し、その手がかりとなる振る舞いを見出した。

また、リサンプリングから得られる分散は再構成結果に対するに留まらず、仮想的な観測点を設定し、その仮想観測の結果に対しても計算することができる。これを利用し、分散の大き

い観測点に対して実際の観測を実行して潰してゆくことで、測定精度を効率的に向上させてゆくことができると考えた。この手続きを用いた観測の最適化の手法を構想している。

## (2) 詳細

### (A) HAADF-STEM 撮影像への圧縮センシングの適用

評価関数

$$L(x) = \frac{1}{2} \|y - Ax\|_2^2 + \frac{1}{2} \lambda_1 \|Bx\|_2^2 + \lambda_2 \|x\|_1$$

( $x$  は原子位置のベクトル(信号) $y$  は観測データ、 $A$  は観測行列、 $B$  は全変動行列)に基づく圧縮センシングと結合成分解析を用いて図 1(A)に示した 3 枚の撮像から図 1(B)に示した 3 次元原子配置を得た。変数  $x$  の総数は 125073、観測データ  $y$  の総数は 14672、得られた配置の原子の総数は 24571 個。全原子配置の像を得るとともに、隣接する格子点における原子配置の解析から、触媒活性が高いと見られる原子を抜き出した(図 1(C)参照)。

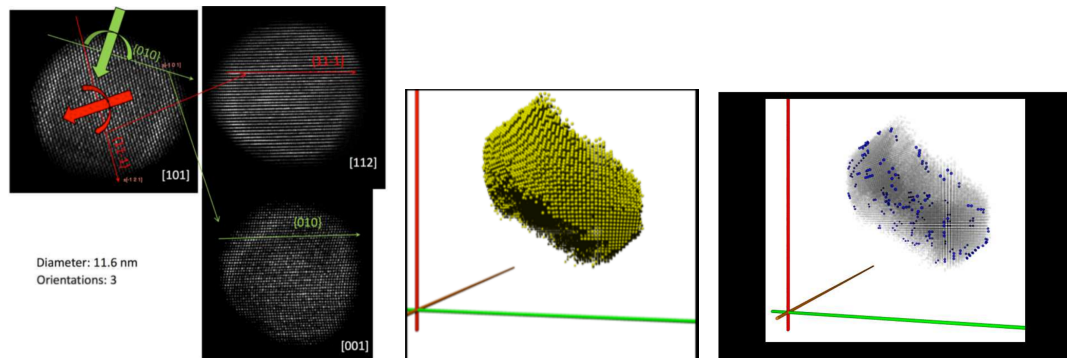


図 1：左から、(A) HAADF-STEM による撮影像、それぞれ[001]、[101]、[112]方向から撮影。(B) 再構成した原子配置。(C) 触媒活性が高いと見られるサイトを抜き出した図。

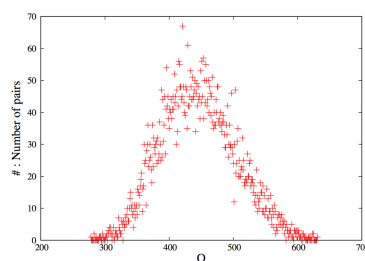


図 2：リサンプリングにより得られた Hamming 距離の分布(横軸は差異の見られた格子点の数、縦軸はヒストグラムの値)。

また、図 1(A)の観測データを確率 1/120 でランダムに欠損させて 120 個のリサンプリングデータを生成し、それぞれについて再構成を行うことで 120 個の再構成像を用意し、それらの間で差異を統計的に調べることで再構成結果のぶれを見積もった(図 1(C)参照)。

### (B)圧縮センシング適用結果に対する信頼性評価

上記(A)で得られた結果に対し、前述の通りリサンプリングを利用した統計的性質の評価を行った。この点を掘り下げ、圧縮センシングにより得られた結果に対して誤差の評価、ならびに(ひいては)再構成の正誤判定を行う手法の開発に着手した。

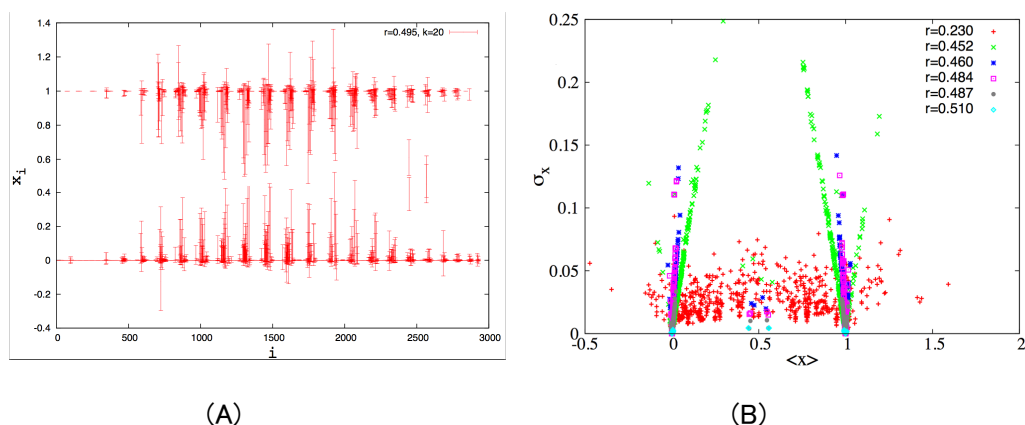


図 2 (A)原子配置の人工データに対してリサンプリングにより得られた変数別の期待値と分散(横軸は変数  $x$  のインデックス、縦軸は再構成結果の期待値、エラーバーは分散)。(B)同人工データのリサンプリングに関して、観測データ  $y$  の数を種々にとった場合の期待値-分散の散布図。データ量( $y$  の要素数)の多い観測(図中赤紫、青、緑の点)と少ない観測(図中赤点)で振る舞いが定性的に異なる。

圧縮センシングはより少ないデータから信号の再構成を可能にするとはいえ、Donoho-Tanner ならびに樺島-和田山-田中により解析されたように、正確な再構成に必要なデータの下限がある。電子線 CT の場合にも同様の下限があると考えるのが自然であるが、実データの再構成においては、Donoho ならびに樺島らのランダムな原信号の解析とは異なり、原信号にはほぼ必ず何らかの構造がある。多様な実データに見られる構造に対して再構成限界(相境界)を解析的に計算することは困難である。本研究では、要素数 2929、うち原子の占有数 599 の人工データの再構成に対して、リサンプリングにより得られる分散等を計算した。結果、全観測データ量に依存して、再構成された信号の変数別の分散の振る舞いに図 2(B)のような定性的な違いが見出された。

### (C)リサンプリング分散を用いた観測点配置の最適化に向けて

観測データからのリサンプリングにより、仮想的な射影に対する結果  $y$  の分散の計算を行った。射影軸の選択により予測分散の大きさは変わるが、軸が既存のもののいずれかに近い場合は分散の値は小さく、また既存のいずれの軸ともより直交に近い場合ほど分散が大きいことが見出された。(図 3 参照)

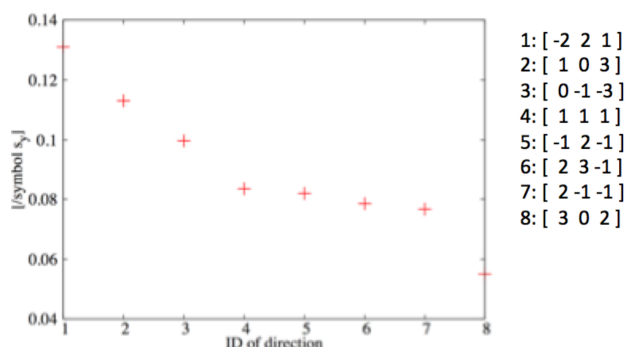


図 3:  $[001]$ ,  $[101]$ ,  $[112]$  の 3 方向からの観測データからの再構成に対し、新たな方向 (図右部) からの観測を加える場合の、 $y$  の分散 (変数について平均) の値。

既存のものと近い方向についての射影は観測済のそれと類似するため新規な情報に乏しく、大きく異なる方向からの射影図ほど取得済みのいずれとも独立な情報を含むという直観が成り立つが、上記の結果はその意味で直観と整合する。この結果は直観との整合性の確認の域を出ていないが、これが計算結果として実証されたことはリサンプリングに基づく分散の大きさが独立な情報の度合いと関連することを意味しており、この分散を情報取得の指標とすることによる観測の最適化への期待を根拠づけるものである。

### 3. 今後の展開

圧縮センシングの応用研究は、その黎明期においては『いかに少ないデータで従来以上の推定結果が得られるか』を全面に出した研究が多かった。しかし今後は導入された各分野において、同手法の定着に伴い、もとの定量科学としての文脈に即した発展をするだろう。それに伴い、実データへの圧縮センシングの適用結果に対するエラーバー評価、ならびに正誤判の手法の重要性は増していくと期待できる。

リサンプリングにより予測分散を計算することによる観測点配置の最適化は、物性測定 of 効率化・省力化に直接つながることが期待できる。またこの手法はベイズ最適化と類似するが、分散を与える揺らぎのソースが異なる。この点についての情報科学的な研究も興味深いと考えられる。

### 4. 自己評価

原子クラスターに対する電子線 CT については、おおむね予定通りの成果を挙げることができた。また、圧縮センシングの結果に対する信頼性評価の手法開発は、当初の研究計画には入っていなかったが、電子線 CT の現状を勉強するうちに研究対象として組み込まれていった。これは当初の計画にあった内容を補完する上で必要な内容だったと考えている。また、リサンプリングによる予測分散を用いた観測の最適化も、同内容から発展したものである。



当初の計画とは異なる方向への発展が多く、全体的に題目と異なる内容となっていた点は反省すべきだと思う。しかしながら、電子線 CT を中心に、実験条件からくるデータ取得への制限について掘り下げること、新たな問題意識ならびに研究の方針を発見できた点は今後の展開につながる成果だと考える。他方、当初計画に組み込んでいた課題には未着手のものがあ、これは今後の課題である。

## 5. 主な研究成果リスト

### (1) 論文(原著論文)発表

なし

### (2) 特許出願

研究期間累積件数:0 件(公開前の出願件名については件数のみ記載)

### (2) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

#### 学会発表

1. 中島千尋、“電子線 CT による原子像再構成における再構成可能性判定”、日本物理学会第 73 回年次大会 講演番号 22aK703-2 (東京理科大学野田キャンパス, 2018/3/22-25).
2. Chihiro H. Nakajima, “Electron beam CT using HAADF-STEM- Toward reliability evaluation of reconstruction with real data”, 2<sup>nd</sup> PHYSICS INFORMED MACHINE LEARNING (Inn and Spa at Loretto, Santa Fe, USA, 2018/1/21- 25).
3. Chihiro H. Nakajima and Masayuki Ohzeki, “3D image reconstruction of a gold nano-cluster by compressed sensing”, PHYSICS INFORMED MACHINE LEARNING (Inn and Spa at Loretto, Santa Fe, USA, 2016/1/19- 22).

#### 講演(招待講演・依頼講演等)

1. “TEM イメージからの 3 次元ナノポーラス構造の再構成”、応用物理学会結晶工学分科会主催：第 22 回結晶工学セミナー(工学院大学,2017/12/12)
2. “HAADF-STEM を用いた電子線 CT そして実データによる再構成の信頼性評価に向けて”、第 7 回計算統計物理学研究会(東北大学, 2017/9/25-26)
3. “実験データ処理のマテリアルズインフォマティクス：CT 画像を例に”、土木学会応用力学委員会主催：応用力学フォーラム(岡山大学, 2017/1/20)

#### 著作物

1. “TEM イメージからの 3 次元ナノポーラス構造の再構成”、応用物理学会結晶工学分科会主催：第 22 回結晶工学セミナー(上記講演参照)講演概要