

(技術資料)

計測インフォマティクスの紹介

世木 隆^{*1} (博士(理学))

Introduction to Measurement Informatics

Dr. Takashi SEGI

要旨

材料開発に必要な多くの物理・化学分析技術と機械学習を併用する計測インフォマティクスは、従来にない解析速度や付加価値創出が期待されている。本稿では、放射光実験にベイズ最適化を組み合わせることで測定時間を短縮化した事例のほか、深層学習により画像解析の省力化を実現した事例、データの品質を向上させた超解像についてそれぞれ例示した。さらに、誰でもAI技術を利用できるインフラストラクチャーの社内整備について紹介した。計測インフォマティクスは、ドメイン知識（ここでは分析分野の専門知識）、データサイエンス、および情報工学の学際技術であり、これら三つの技術分野を組み合わせた取り組みの重要性を示した。

Abstract

Measurement informatics, which combines many physical and chemical analysis techniques required for material development with machine learning, is expected to realize unprecedented analysis speed and creation of added values. This paper exemplifies a case where the measurement time has been shortened by combining a synchrotron radiation experiment with Bayesian optimization, a case where labor saving in image analysis has been realized by deep learning, and a case where data quality has been improved by super-resolution imaging. Also introduced is an in-house preparation of infrastructure allowing everyone to use AI technology. Measurement informatics is an interdisciplinary technology involving domain knowledge (here, expertise in the field of analysis), data science, and information engineering. The importance of efforts that combine these three technical fields has been demonstrated.

検索用キーワード

計測インフォマティクス, 機械学習, ベイズ最適化, セグメンテーション, 深層学習, 超解像, AI, SEM

ま え が き = 現在、試作、理論、および計算によって支えられていた材料開発にデータサイエンスを組み合わせる材料インフォマティクスが注目されている。いっぽうで、素材が持つ特性や構造などの情報を高効率で取得・解析する計測インフォマティクスの熟成によって材料データのビッグデータ化が求められている。ここでは、主に計測実験、計測データ解析、品質向上の局面へ機械学習を応用した結果についてそれぞれを例示し、最後に機械学習利用を支えるプラットフォームについて紹介する。

1. 計測実験に対する機械学習の応用

本章では、実験時間の短縮のために、放射光実験のビームラインへ機械学習を適用した事例について紹介する。

国立研究開発法人理化学研究所が運営する大型放射光施設 (SPring-8) は世界最高クラスの輝度・エネルギー・指向性を持つ X 線放射光を生み出す照射施設として知られている。国立研究開発法人量子科学技術研究開発機構 (QST) は上記施設内に専用のビームライン BL11XU を所有している。BL11XU では、核共鳴散乱、X 線吸収端微細構造、X 線非弾性散乱などの静的な実験を行うことができる。また、分子線エビタキシー蒸着装置と表面 X

線回折計を併用することにより、結晶成長のその場実験などの動的な実験が可能である。この X 線回折計は、X 線を μm オーダへ集光して局所的な回折パターンが得られる特徴を持っている。このため、例えば数 $10 \mu\text{m}$ 四方の薄膜試料における空間的な配向変化などの情報が得られる。

本実験に供する試料は Si(001) 基板上に成膜された Si-Ge 薄膜であり、組成と配向が分布している特徴を持つ。このような材料の計測実験を行う場合、特定のピーク強度が最も高い場所をあらかじめ探索する予備的な測定を行うことが多い。例えば、特定のピーク強度 z を探索する場合、グリッドスキャンにより $1 \mu\text{m}$ 径の X 線を $5 \mu\text{m}$ おきに $150 \mu\text{m}$ 四方の試料の xy 空間へ走査すると、 $31 \times 31 = 961$ 回の計測を行う必要があり、多くの時間を要する。

そこでこの計測時間を短縮化するために、ベイズ最適化を用いて X 線回折計を制御する取り組みを行った。ベイズ最適化はブラックボックス関数に対する大域的最適化計算として知られており、多くの最適化問題への応用が広がっている¹⁾。

この実験ではサンプルの座標 (x, y) に対するビーム強度 z を与える関数 $f(x, y)$ が上述したブラックボックス関数に対応すると考える。この関数をガウス過程

^{*1} (株)コベルコ科研 技術本部 計算科学センター

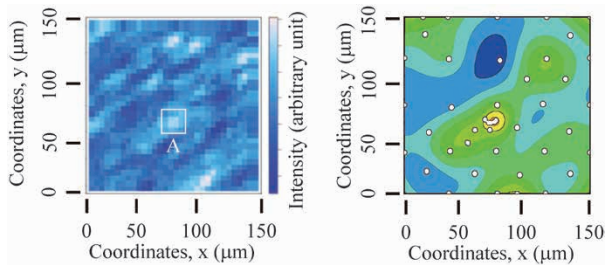


図1 Si-Ge 薄膜のXRDパターンから得られた特徴マップ(左) 実験、(右) 予測
Fig.1 Experimental (left) and predicted (right) feature map of X-ray diffraction pattern obtained from Si-Ge film

$G(x, y)$ で近似すると未測定点の期待値と分散が評価できる。この時、未測定点の分散を小さくするように後述の戦略に従ってデータを逐次的に追加すると $f(x, y)$ を近似する $G(x, y)$ が得られる。

実際の適用事例を図1に示した²⁾。図1(左)のコンターは実測で得られた結晶配向を反映する特徴マップを示した。この強度は、ダイヤモンド構造(022)回折ピークの逆格子空間の情報を非負値密度行列因子³⁾により4つの基底ベクトルへ分解した結果として得られる係数を反映したものである。グリッドスキャンでこのデータを得るためには約3時間の計測時間(計測ポイント961点)を要した。

いっぽう、バイズ最適化を用いて左図のポイントAを探索する場合、まずランダムな位置で10点程度を測定する。つづいて、その結果に基づいて初期の獲得関数を見積り、期待改善(Expected Improvement)戦略に基づいて測定点を逐次的に追加して獲得関数を更新する。これによって近似された特徴マップを図1(右)に示した。図中の白点は実際に測定した箇所を示しており、結果として48点のみの測定によってポイントAの場所を探索することができた。すなわち、グリッドスキャンを使った場合より測定時間を1/20に短縮化できる結果が得られた。

2. 計測データ解析に対する機械学習の応用

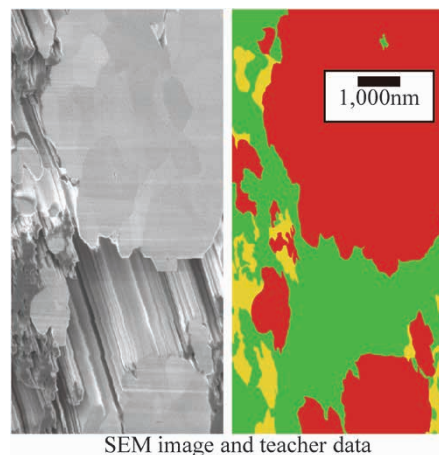
計測データ解析に機械学習を適用するメリットとして、従来専門家が行っていた作業の多くを自動化できる点が挙げられる。例えば、走査型電子顕微鏡(Scanning Electron Microscope: SEM)を用いると材料が持つ化合物や結晶粒の大きさ、形、数などの情報が取得できる。画像データ解析においては、従来、組織や素材に対する二次電子像や反射電子像に対する知識と経験を持った専門家が対応していた。自動化した画像データ解析においては、市販のレタッチソフトなどを使って化合物相ごとに分類した後、画像解析ソフトを用いて数値化する手続きを行う。

こうした一連の手続きにおいては、結晶が複雑で観察範囲が数視野にまたがる場合などは解析者への負担が大きくなるという問題があった。しかしながら、今のところ、これらの問題解決にはピクセルごとにオブジェクトを分類するPyramid Scene Parsing Network⁴⁾やU-net⁵⁾などの良いセグメンテーション手法が貢献することが知

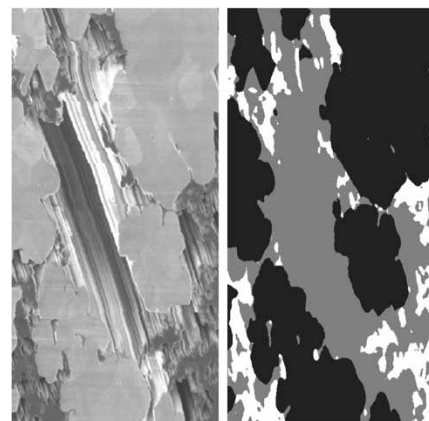
られている。ここではSEGNET⁶⁾を基にカスタマイズした自動分類例を図2に示した。この図は機能性材料についてGaイオンを使った集束イオンビーム(Focused Ion Beam:FIB)加工機を用いて得た断面SEM像である。この視野には酸化物と炭素材料が含まれ、その間には空隙が存在する。これらを画像解析の専門家によってそれぞれ赤、黄、緑で着色した結果を図2右に示した。

この画像解析を対象とした機械学習においては、元データである観察結果(図2左)と着色によりラベル付けされた教師データ(図2右)のペアから構成される数多くの教師データを使って双方の相関を取得(学習)させる。その学習成果を未学習の断面SEM像に適用することによって自動的に画像が分類できる。その結果の一例を図3に示す。酸化物(黒)、炭素材料(白)、空隙(灰色)が高精度で分けられた(図3右)。この画像分類に要する計算時間は、市販のグラフィック処理ユニット(Graphics Processing Unit:GPU)を搭載した計算機であれば数秒で完了する。このため、画像データ解析に必要な時間は大幅に削減できる。

ここでは機能性材料への事例を示したが、金属などの構造材料でもこの種の画像データ解析ニーズは大きく、また、SEM以外にも光学顕微鏡などでも利用可能である。ただし、必要になる教師データの数は試料の多様性などに依存するため、サンプルに対応して必要十分なデ



SEM image and teacher data
図2 セグメンテーション用のSEM画像(左)と教師データ(右)
Fig.2 SEM image (left) and teaching data (right) for segmentation task



SEM画像(左)と深層学習を用いた分類結果(右)
Fig.3 SEM image (left) and predicted data (right) by using deep-learning

ータ数を用意する必要がある。ロバスト性を担保するためには観察倍率・加速電圧などの測定条件だけでなく、エッチングや電子染色などの前処理条件なども含め、実験計画法などを用いながら決定していくことになる。

3. 機械学習を用いた品質の向上

本章では機械学習技術を用いたデータ品質向上について、Slice & Viewによる3次元立体構造構築の事例を紹介する。Slice & Viewでは、FIB-SEM複合装置を用いることによってサンプル内部の所望する断面を連続的に観察した材料組織が撮影できる。

FIB-SEM複合装置を用いた観察では、FIB加工機を用いてサンプル断面を構築した後、その新生面をSEMで観察する。これを100回繰り返すことによって図4に示したような立体構造を構築することができる。素材によって条件が異なるが、 xy 面内の $10 \times 10 \mu\text{m}$ の観察断面に対し、 z 軸方向に100 nm程度の加工ピッチで行うことが多い。

ここで、SEM像の空間分解能は一般に10 nm以下とされているなかで、加工ピッチを100 nmとしている。このため、結果として得られる立体構造の解像度も100 nmに制約を受けることが懸念される。こうした懸念を解消するためには、10 nmピッチで加工すれば良い

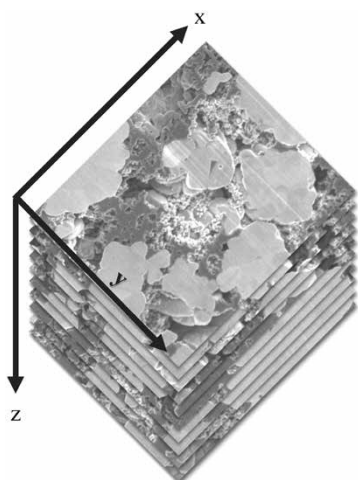


図4 FIB-SEM観察の概念図
Fig.4 Schematic drawing for FIB-SEM observation

ということになる。しかしながらその場合、数十時間もの観察時間が必要になることから現実的ではない。そこでここでは、実験を行うことなく10 nm相当の解像度を持った立体構造を推定する方法を採った。

上述したとおり、 xy 面内には10 nm相当の空間分解能を持った像が存在する。したがって、アプローチとしては、この情報を利用して深さ方向(z 軸方向)の組織構造を高解像化することとした。このようにして得られた高解像は超解像(Super Resolution: SR)と呼ぶ。超解像技術としてはSuper Resolution-Generative Adversarial Network⁷⁾などの手法がよく知られているが、ここではSuper Resolution Residual Attention Module (SRRAM)を採用した⁸⁾。この手法では、入力情報から2次元コンボリューション層を経て深層特徴量を抽出し、画像の解像度を2倍に高める。この工程を4回繰り返すことによって最大16倍の超解像データを作成する。

超解像計算の結果を図5に示した。超解像計算の対象は第2章で評価した試料と同一試料である。一番右の像(図5(e))は従来解像度で得られた断面画像であり、このデータを入力として深さ方向(z 軸方向)の画像の解像度を向上させた。SRRAM(図5(b))、Lanczos法(図5(c))、および画像補完技術である最近傍補間(図5(d))を使って同様の処理を行った結果を正解画像であるGround truth(図5(a))と共に示した。このGround truthは比較のために行われた10 nm相当の高解像度実験による実測結果である。SRRAMで得られた超解像計算結果を見ると従来の補完技術では得られなかった微細な構造が再現されていることがわかる。

さて、この超解像計算と2章で示したセグメンテーション技術とを組み合わせることにより、立体構造解析を行うことが可能になる。ここでは、この立体構造解析には市販ソフトを用いた。従来解像度で行った場合と比較した解析結果を表1および表2に示した。表1は酸化物、炭素材料、および空隙の体積を示し、表2は酸化物と炭素材料、酸化物と空隙、そして炭素材料と空隙が隣接する接触面積を示している。

体積値は従来解像度と超解像との差は小さい。他方、とくに炭素材料と空隙が構成する接触面積は大きく異なる

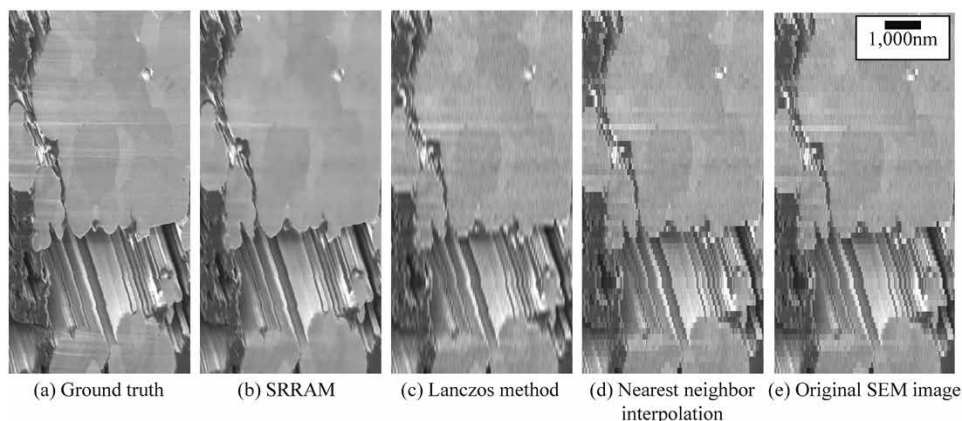


図5 従来解像度と超解像の比較 ((a) 高解像度実験 (解像度10 nm相当), (b) SRRAMによる超解像, (c) Lanczos補間, (d) 最近傍補間, (e) 従来解像度 (解像度100 nm相当))
Fig.5 Comparison of (a) experimental high-resolution SEM image (resolution 10 nm equivalent); Ground truth, (b) SRRAM, (c) Lanczos method, (d) Nearest neighbor interpolations, and (e) original observation (resolution 100 nm equivalent)

表1 3次元立体解析 (体積)
Table 1 3-dimensional structure analysis (volume)
(μm^3)

Object	Conventional	SR
Oxides	145.3	150.5
Carbon	24.7	26.4
Void	113.9	106.8

表2 3次元立体解析 (接触面積)
Table 2 3-dimensional structure analysis (contact surface area)
(μm^2)

Object1	Object2	Conventional	SR
Oxides	Carbon	55.1	68.9
Oxides	Void	252.2	329.5
Carbon	Void	332.4	556.2

る結果が得られた。これは、従来粗視化されて見えなかった炭素材料などの細かな構造が反映された結果と推定される。

今回の手法は、100 nm 条件に必要な実験時間しか使っていないにもかかわらずデータの品質が向上したことを示した。このような技術は、例えば100 nm よりも細かな構造情報を扱う触媒やリチウムイオン二次電池などへの応用が期待される。

4. 機械学習のプラットフォーム

実験や解析に機械学習技術を援用することによって省力化や品質向上が可能になる。(株)コベルコ科研では、社員は誰でもこのような機械学習サービスを利用できる環境を計算科学センターに構築した。この機械学習サービス計算機は、アクセスポイント部、実際に機械学習の演算を行う AI 部、および利用履歴やアクセスログを保持するデータベース部の3種のサーバ群を有機的に運用した自社製システムである。計算科学センターではこれを“AIサーバ”と呼称している。利用者は解析対象となる電子データ(画像またはcsv形式ファイル)を本サーバへアップロードし、必要な解析内容を選択することによって短時間で結果が得られる。

本サービスにおいては、実行型ファイルを社内に配布するのではなくサーバ形態を採用した。これは、AIの汎化(はんか)性能を維持するためには定期的に追加学習を行うことが有効であり、そのためには利用の度に自然とデータが集まる本形態が適するためである。

ユーザはWebブラウザを使って必要な時にAIサーバにアクセスでき、e-mailとファイル共有サービスを用いて計算結果を受け取ることができる。本サービスはWindows標準のブラウザとメーラがあれば利用できる。

いっぽう、機械学習技術を用いて開発した新しい解析プログラムは、pickleあるいはh5pyでシリアルライズ(直列化)された学習済みモデルとプリ・ポスト処理を記述したPythonスクリプトを準備するだけで良い。計算科学センターにおけるコードレビューを受けた後に、新たなAIサービスとしてリリースできる(2020年現在)。AI技術の活用にはインフラ整備が重要であり、このよ

うな取り組みを通じ計測インフォマティクス技術が定着化していくと考えている。

むすび=機械学習を応用することにより、実験時間や解析時間の短縮化だけでなく、データの品質を向上させることができることも示した。学際領域の色合いが強いインフォマティクスを実用化するためには、材料や分析分野の専門知識(ドメイン知識)やデータサイエンスだけでなく、さらに情報工学分野を加えた三つの技術領域を全て組み合わせることが大切になる。これらの活動を通じ、神戸製鋼グループの社会貢献へ寄与したい。

1章で述べた放射光実験は、高輝度光科学研究センター課題(No. 2016A3553, 2017A3553, 2018A3553, 2018A3585, 2018B3587, 2019A3587, 2019B3586, 2020A3586)によりSPring-8 BL11XUで実施した。また、文部科学省委託事業ナノテクノロジープラットフォーム課題(課題番号:A-18-QS-0005)のQST微細構造解析プラットフォームおよび理化学研究所革新知能統合研究センターの支援を受けて実施した。さらに、理化学研究所革新知能統合研究センターの沓掛健太郎先生をはじめ、名古屋大学大学院医学系研究科の松井孝太先生、量子科学技術研究開発機構量子ビーム科学部門の佐々木拓生先生、豊田工業大学神岡武文先生のご協力を得て実施した。ここに感謝申し上げます。

参考文献

- 1) B. Shahriari et al. "Taking the Human Out of the Loop: A Review of Bayesian Optimization". Proceedings of the IEEE. 2016, Vol.104, p.148-175.
- 2) 沓掛健太郎ほか."放射光X線回折パターンの特徴抽出と空間マッピング(II)". 第80回応用物理学会秋季学術講演会講演予稿集. 2019, p.18-036.
- 3) A. Cichocki et al. "Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations". IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences. 2009, Vol.E92-A, p.708-721.
- 4) H. Zhao et al. Pyramid Scene Parsing Network. arXiv.org e-Print archive, 2016, 1612.01105. <https://arxiv.org/abs/1612.01105v2>, (参照 2020-12-14).
- 5) O. Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv.org e-Print archive, 2015, 1505.04597. <https://arxiv.org/abs/1505.04597>, (参照 2020-12-14).
- 6) V. Badrinarayanan et al. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv.org e-Print archive, 2015, 1511.00561. <https://arxiv.org/abs/1511.00561>, (参照 2020-12-14).
- 7) C. Ledig et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv.org e-Print archive, 2016, 1609.04802. <https://arxiv.org/abs/1609.04802>, (参照 2020-12-14).
- 8) J-H. Kim et al. RAM: Residual Attention Module for Single Image Super-Resolution. arXiv.org e-Print archive, 2018, 1811.12043. <https://arxiv.org/abs/1811.12043v1>, (参照 2020-12-14).