

機械学習・統計解析を用いた量子ビーム計測の最適化と計測データ解析

伊藤 優成, 小野 寛太

1 はじめに

近年、機械学習や統計解析技術の爆発的な普及に伴い、さまざまな計測手法における計測データの取得および解析の手法が急速に進化している。この急速な進化は量子ビーム計測にも大きな影響を与え、効率的な計測と先進的なデータ解析に関する研究が活発に進められている。

量子ビーム計測とは、X線、中性子線、ミュオン、陽電子線などの量子ビームを使用して材料の内部構造や化学状態を解析する手法である。X線は、その透過性と元素選択性を活かして、結晶構造解析、実材料の欠陥・不純物の検出、医療分野での非侵襲的な体内断層画像の取得（X線CTスキャン）、さらに考古学における古代遺物の内部構造調査などに広く利用されている。中性子線は原子核との相互作用に基づき、X線との相互作用が小さいためX線では捉えにくい軽元素や、磁気モーメントを持つ物質の研究に非常に有効であり、金属水素化物や磁性材料の内部構造と磁気特性の研究などに用いられている。陽電子線は材料の表面構造や欠陥状態の解析に適しており、材料科学では半導体の欠陥状態の研究による電子デバイスの性能向上に寄与し、医学では陽電子放射断層撮影（PET）を通じた体内生化学的プロセスの観察やがん診断に活用され、化学分野では触媒材料の解析や開発に重要な役割を果たしている。ミュオンはX線よりも透過性が高く、また軽元素への感度も高いなどの利点から、試料の元素深度分布の分析などに用いられている。このように、量子ビーム計測技術は物理学、化学、生物学、医学といった自然科学の広範な分野で不可欠な技術となっている。

特に、X線領域における量子ビーム利用研究の分野では、シンクロトロン放射光施設の高輝度化と計測装置の高度化により、ハイスループットな計測が可能となり、従来は観測が困難であった微細な構造や動的な現象の解析が実現されている。このようなハードウェア技術の進化は、計測の幅を広げ、科学的な発見や技術革新を大きく促進している。しかし、その一方で、得られる計測データは非常に大規模化しており、計測データの処理や解析には高度な情報技術と大規模な計算資源が不可欠となっている。このような背景に加え、昨今のデジタルト

ランスフォーメーション（DX）の潮流から、機械学習を用いた解析手法の高精度化・高速化・自動化のための研究がますます重要な課題として認識されている。

本稿では、ハイスループット化が進んでいるX線粉末回折計測と、X線の高輝度化により計測データの大規模化が著しいX線吸収分光イメージング計測を例として取り上げ、機械学習や統計解析を用いることで解析の高度化と計測最適化がどのように進められているかについて、最新の事例を交えながら詳しく説明する。具体的には、機械学習によるデータ解析の効率化、統計解析を用いた実験設計の最適化、さらにこれらの技術がもたらす研究の新たな可能性について論ずる。

2 粉末X線回折計測（XRD）

2.1 粉末XRD解析の難しさ

粉末X線回折（XRD）計測は、粉末状にした試料にX線を照射し、反射されたX線の角度と強度のプロファイルを解析して物質の結晶構造を特定する手法であり、量子ビームを用いた計測の中でも最も広く用いられている手法の一つである。粉末XRD計測は放射光施設だけでなく、実験室で計測可能な装置も市販されており、実験室での計測では1試料あたり数十分、放射光施設では24時間で5000試料を超えるハイスループット計測が実現されている。これにより、新しい試料を合成するたびに粉末XRD計測を行う研究者も少なくない。

しかし、粉末XRD解析には多くの難しさが伴う。代表的な解析手法であるリートベルト解析には、熟練者の経験と技術が必要であるとされている。リートベルト解析は、初期構造モデルに基づいて非線形最小二乗法を用いてモデルフィッティングを行うことで、結晶構造を詳細に推定する方法である。結晶構造の精密な推定には、格子定数、原子位置、熱振動パラメータ、サイト占有率など、多くのパラメータがかかかっており、これらのパラメータは完全に独立しているわけではなく、相互に影響し合う。そのため、すべてのパラメータを同時に最適化すると構造精密化が収束しないことが知られている。さらに、粉末XRD計測では、観測されるバックグラウンドと試料由来のピークとの区別が難しい場合がある。バックグラウンドは、試料由来ではない成分であり、これを適切に取り除くことが解析の精度に大きく影響す

る。例えば、試料の非晶質成分、試料台や周囲の環境からの散乱、検出器の特性によるノイズなどがバックグラウンドに含まれる。これらのバックグラウンド成分を正確にモデル化し、実際のデータから除去するには、経験と専門知識が必要であり、これが粉末 XRD 解析を難しくする一因である。研究者はバックグラウンド関数や精密化するパラメータの順番を試行錯誤しながら解析を行う必要があり、熟練者でも 1 試料あたり数時間を費やすことが多い。

2.2 粉末 XRD 解析への機械学習の適用

このような解析の難しさを解決するために、機械学習が適用される事例が増えている。尾崎ら¹⁾の研究では、リートベルト解析を行う際の研究者の試行錯誤がブラックボックス最適化という枠組みに類似していると捉え、最適化問題の解法の一つであるベイズ最適化を用いて自動解析を実現した。このアプローチは、研究者が従来行っていた意思決定プロセスを機械学習に置き換えた非常に明快な方法である。

リートベルト解析を行う際、研究者はフィッティングの良さを示す指標である R_{wp} がより小さくなるように解析手順を試行錯誤する。これは、図 1 に示すように解析のパラメータ（精密化するパラメータの順番など）を入力し、その結果として得られる R_{wp} を出力と考えた場合、ブラックボックス最適化、すなわち入出力の関係性が未知の状態での最適化問題を解いていることに対応する。そこで、ブラックボックス最適化の代表的な解法であるベイズ最適化を用いて最適化問題を解くことで、研究者の試行錯誤を代替することができ、人手を介することなく自動で解析を行うことが可能となる。これにより、従来は熟練した研究者が数時間かけて行っていた解析が、標準的なワークステーションを用いて 30 分程度で可能となる。

このような自動データ解析の利点は昼夜を問わず次々と自動で解析を行える点にある。例えば、夜間に一連の

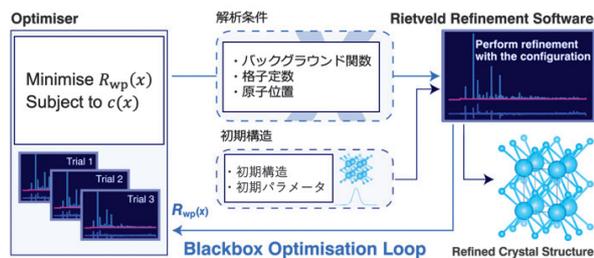


図 1 ブラックボックス最適化を用いたリートベルト解析の模式図

Schematic diagram of automated Rietveld analysis with blackbox optimization. Originally published in npj Computational Materials 6, 75 © The Author (s) 2020. Published by Springer Nature. Licensed under CC BY 4.0.

測定データの解析を行っておけば、翌朝には研究者はその解釈に専念することができる。また、自動解析は研究者の先入観によらない解析を行うことを意味し、従来の専門家の人手による精密化のプロセスで見落とされていた解析結果を得ることに成功している。

リートベルト解析のように、多数のパラメータを最適化する必要がある場合には、ベイズ最適化を用いることで、パラメータ探索の効率が飛躍的に向上する。ベイズ最適化は、試行錯誤の過程をモデル化し、効率的に最適解を探索する手法である。計測データ解析には非線形最小二乗法に基づくモデルフィッティングを用いることが多いため、最適化問題として定式化し、適切な最適化手法を用いることにより、従来の手法では困難であった複雑なパラメータ空間の探索が可能となり、解析精度の向上が期待できる。また、機械学習モデルを用いることで、データからバックグラウンド成分を自動的に抽出し、除去することも可能である。これにより、研究者の負担を大幅に軽減し、解析の信頼性を向上させることができる。

「機械学習を用いた XRD 解析」というと、計測結果からリートベルト解析のような精密化プロセスを経ずに、直接構造を出力する方法をイメージすることが多いが、これは従来の手法から完全に離れた新しいアプローチである。それに対し、従来の研究者が行っていたリートベルト解析プロセスをそのまま機械学習に置き換えるこのアプローチは、過度な機械学習への置き換えを避けることができ、非常に実用的である。このように従来の手法を踏襲しつつ機械学習を活用することは、透明性を保ちながら信頼性の高い解析結果を得るために非常に重要である。

ここまで述べたように、機械学習の適用により、粉末 XRD 計測の解析プロセスが大幅に効率化・自動化され、精度も向上している。特にブラックボックス最適化を用いたアプローチは、研究者の試行錯誤を機械学習に置き換えることで、自動化と精度向上を両立させている。これにより、研究者はより創造的な研究に集中できるようになり、さらに多くの発見が期待される。今後は、深層学習を用いた解析手法の進展により、ますます高度な解析が可能となるであろう。

3 X線吸収分光イメージング計測

3.1 X線吸収分光イメージング

X線吸収分光イメージング (XAS-Imaging) は、空間各点の電子状態や局所的な化学状態を反映した X線吸収分光スペクトル (XAS) を取得することができる計測手法であり、電池や触媒などのさまざまな材料の解析に広く用いられている。シンクロトロン放射光施設を利用したこの手法は、光学顕微鏡程度の分解能で大視野を計測するものから、10 nm 以下の分解能で計測するもの、

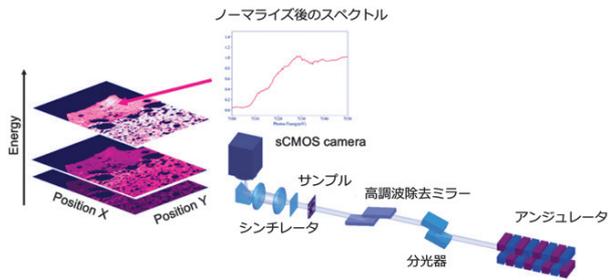


図2 XAS-Imaging 計測のセットアップ

さらに X 線の高い透過性を活かした 3D イメージングまで、さまざまなバリエーションが存在する。図 2 は、大視野計測に適した XAS-Imaging のセットアップを示している。単色化した X 線を試料に照射し、その透過光をシンチレータで可視光に変換し、sCMOS カメラで撮像する。X 線のエネルギーを変化させながら各エネルギーでイメージングを行うことで、空間各点のスペクトルを得ることができる。実際の計測では、試料の位置を変化させてタイリング計測を行うことで、1 cm×2 cm 程度の広範囲を計測することが可能である。

近年、X 線吸収分光イメージングを含むイメージング技術の高分解能化と大視野化が急速に進展している。計測対象がマクロスケール全域にわたりミクロスケールで均一な試料であれば局所領域を高分解能で計測すれば十分な情報が得られる。しかし、一般に実際の材料はマクロスケールでは均一であるように見えてもミクロスケールでは不均一な化学状態や構造を持ち、そのことが材料特性を決定付けていることが多い。そのため、試料全体を大視野かつ高分解能で計測する必要がある。以上のような背景から、大視野かつ高分解能計測で得られる膨大なデータを効率的に解析するためには、機械学習の技術が欠かせない。

以下に、X 線イメージング計測への機械学習の適用事例を二つ紹介する。

3.2 統計解析を用いた XAS-Imaging の計測最適化

自然科学において、実空間での観察は極めて多くの情報を人間にもたらす。そこで、高分解能かつ大視野化が必要となってくるが、これらはトレードオフの関係にあるため、計測時間の増大は避けられない。シンクロトロン放射光施設の性能は年々向上しているが、その高輝度化には限界があり、大視野かつ高分解能の計測が進むにつれて、1 ピクセルあたりのフォトン数が少なくなり、計測時間が長くなる。

XAS-Imaging は、X 線のエネルギーを変化させながらイメージング計測を行うため、エネルギー点の数が計測時間に直接影響する。1 ピクセルあたりに必要とされるフォトン数は統計的に求まるが、どれだけのエネルギー点で計測すれば必要十分なのかは、どのようにして

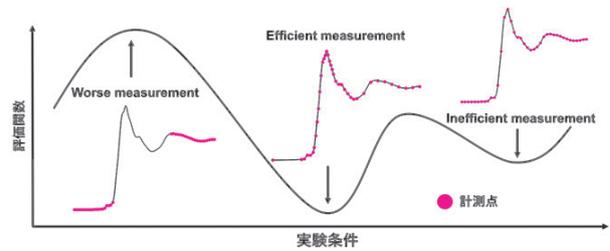


図3 Bayesian Experimental Design により最適な計測点を得る模式図

決めれば良いのだろうか。

従来は研究者の経験に基づいてエネルギー点が決定されていたが、伊藤ら²⁾は統計理論に基づいた実験計画を用いることで、従来の計測時間の 70 % で同程度の精度が得られることを示した。この研究では、研究者がどのように最適な計測点を定めるかに着目している。熟練度が高い研究者は、多くのスペクトルを見た経験があり、どのエネルギー領域を重点的に計測すべきかを経験的に認識している。そこで、この研究では、事前情報および想定される計測ノイズをもとに、最適な実験条件を決定するためのベイズ実験計画法 (Bayesian experimental design) を活用し、計測点の評価関数を定め、図 3 に示すようにそれを最小化することで最適な計測点を得ている。まず、この方法では事前情報を確率分布として与え、その確率分布をもとに繰り返しスペクトルを生成し、その生成したスペクトルを計測ノイズのもとで擬似的に計測した際の「計測の良さ」を計算する。「計測の良さ」とは、離散的な計測点からどれだけ正確に真の (正解の) スペクトルを得られるかを示す指標である。この操作を繰り返すことで、「計測の良さ」の期待値を計算し、それが選んだ計測点の評価値となる。スペクトルを生成し、擬似的に計測して評価する作業は一見途方もなく思えるが、事前情報の確率分布の形に制限を与えることで、解析的に行うことが可能となり、標準的なノート PC で数十分の計算を行えば最適な計測点を出力できる。毎回同じ測定対象・計測ノイズであれば、熟練者の経験的な計測点の決定は高い精度を持つが、計測対象やノイズレベルが変化する実際の測定においては、統計理論に基づく本手法は合理的であり有用である。

さまざまなスペクトルに対して計測点の評価を行い、「計測の良さ」の期待値を考えるプロセスは、熟練者が計測点の良し悪しを経験的に獲得するプロセスを解析的に再現していると言える。すなわち、粉末 XRD 解析と同様に、研究者が行っているプロセスを定式化することにより統計解析に置き換えたものである。

この手法を用いることで、計測時間の大幅な短縮が可能となる。従来の経験に基づく手法では、計測者の主観や経験に依存する部分が大きかったが、ベイズ実験計画法により客観的かつ再現性の高い計測条件の設定が可能

となる。これにより、同一の試料について異なる研究者が計測を行っても、一貫した結果が得られるようになる。また、この手法は、計測時間の短縮に加えて、計測データの質を向上させる効果も期待できる。最適な計測点を選定することで、ノイズの影響を最小限に抑えつつ、必要な情報を効率よく取得することが可能となる。

さらに、この手法の応用範囲は広く、さまざまな材料や測定条件に対応できる柔軟性がある。例えば、新しい材料の特性評価や複雑な化学反応の追跡など、従来の手法では困難だった応用にも対応できる可能性がある。これは、材料科学や化学、物理学など、広範な自然科学分野での研究に有用なものであり、他の計測・分析分野でも活用されることが期待される。

3.3 機械学習を用いた XAS-Imaging のデータ解析

XAS-Imaging のデータ解析も計測同様に機械学習技術の発展により、劇的に進化している。特に、大規模なデータセットの中から重要な微小領域を検出する「干し草の中から針を探す (Needle in a haystack) 問題」を解決するための手法が注目されている。本節では、具体例を交えて、この問題を解決するための手法とその応用について詳述する。

XAS-Imaging は、上述したように材料の化学状態や電子状態を高空間分解能で可視化する強力な手法である。しかし、計測データが膨大になるため、従来の解析手法では重要な微小領域を見落とすリスクがあった。伊藤ら³⁾は、図4に示すように各ピクセルでのスペクトルを教師なし分類することで、微小な未知の相を発見する手法を開発した。この手法により、実材料の4億ピクセルもの計測データから、従来の解析では見落とされていた0.5%の少数相を発見することに成功した。

物理計測データを解析する際、少数相はノイズとして扱われることが多い。これは、多数のデータ中に少数しか存在しないものは、一般的に異常値やノイズと考えら

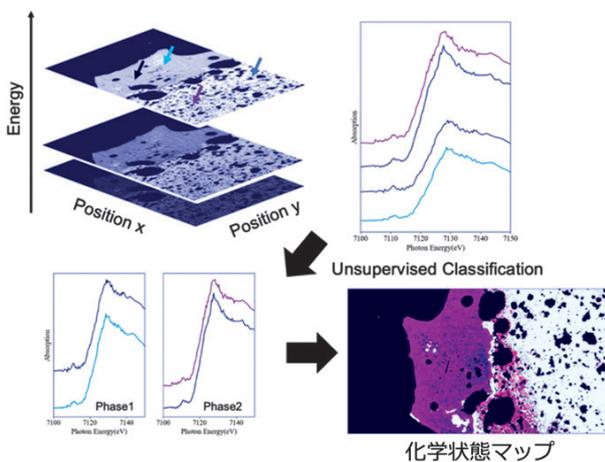


図4 クラスタリングによる解析の模式図

れ、無視される傾向があるためである。従来の機械学習手法では、この問題を解決するための具体的な基準が存在しなかった。伊藤ら³⁾は計測ノイズの定量値を基に分類の基準を定めた。研究者にとって最も合理的な分類の基準とは、二つのデータに計測ノイズを加味しても異なったものとして認識される場合には二つのデータは別の分類であるとし、ノイズに埋もれる程度の違いであれば同じ分類であるとするものである。この基準に基づく分類は、自然で合理的な方法である。

従来の解析手法では、大規模な計測データの解析には限界があった。これは、計測データが膨大になると、すべてのデータ点の組み合わせについて計算することが困難であり、重要な情報を見落とす可能性が高まるためである。例えば、距離尺度に基づく一般的な分類手法である階層的クラスタリングでは、データ点数が N のときの計算量は $O(N^2)$ となるため、非現実的な計算量となる。

伊藤ら³⁾は、この問題を解決するために、図5に示すような計算量を $O(N)$ とする手法を提案した。この手法では、まずスペクトルをエネルギー点数次元のベクトルと考え、すべてのスペクトルはある空間上に存在すると考える。計測ノイズが存在することから、この空間上でスペクトルはぼやけたように存在するため、このぼやけの大きさ (計測ノイズの大きさ) に基づいて空間を分割する。分類を行う際には、それぞれの計測スペクトルが空間のどの領域に対応するかを求めるだけで済み、 $O(N)$ の計算量で分類が実現する。さらに、空間的な分類の分布は偏りやすいこと、すなわち空間的に近い点は近い化学状態を持つ傾向があるため、画像処理的な手法でノイズ削減を行うことができる。この方法により、4億ピクセルもの実材料データに対して合理的な分類が可能となっている。

この手法では、研究者が二つのスペクトルの一致・不一致を判断するプロセスを、膨大なスペクトルのすべてに適用したことがポイントであり、研究者の考え方に裏

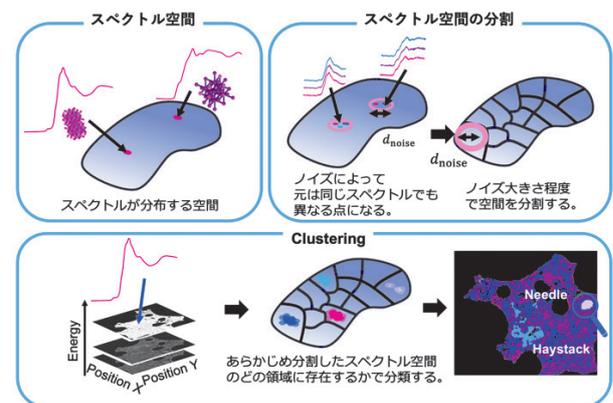


図5 提案するクラスタリング手法の模式図

付けられた手法を定式化することの強さが示されている。

4 おわりに

本稿では、ハイスループット化が進展している XRD と、大規模化計測データが取得されるようになった XAS-Imaging を例として、機械学習や統計解析を用いた計測最適化と計測データ解析について最新の事例を紹介した。近年は、機械学習手法の発展により、機械学習や統計解析を適用することが容易になり、従来の手作業による計測や解析では不可能だった高度な実験が実現できるようになっている。しかしながら、適切な問題設定と問題の定式化が最も重要であるとわれわれは考えている。問題解決の前に、適切な問題設定と問題の定式化を十分に考えないと、問題解決に失敗するリスクが高い。失敗の原因として、統計的あるいは数学的に適切でないツールの使用が挙げられる。われわれは合理的な実験計画やデータ解析を目指しており、そのためには数学的に正しさが保証された統計・機械学習手法を用いることを心がけている。

本稿で紹介した事例で強調したい点は、粉末 XRD 解析におけるリートベルト解析の自動化のように、従来の研究者が行ってきた解析プロセスをうまく機械学習に組み込むことが重要であるということである。これにより、透明性と信頼性の高い解析結果が得られ、研究の精度と効率がどちらも飛躍的に向上する。また、このような研究は、従来の計測・解析手法に精通した研究者の知見があってこそ成り立つものであり、分析技術の専門家

と機械学習の専門家とのコラボレーションにより、さらなる発展が期待される。

本稿で紹介した手法が、研究者にとって新たな道を切り拓く一助となることを願ってやまない。

文 献

- 1) Y. Ozaki, Y. Suzuki, T. Hawaii, K. Saito, M. Onishi, K. Ono : *NPJ Comput. Mater.*, **6**, 75 (2020).
- 2) Y. Ito, Y. Takeichi, H. Hino, K. Ono : *NeurIPS-AI4Mat* (2024).
- 3) Y. Ito, Y. Takeichi, H. Hino, K. Ono : *Sci. Rep.*, **14**, 22549 (2024).



伊藤 優成 (Iro Yusei)

大阪大学大学院工学研究科博士前期課程在学 (〒565-0871 大阪府吹田市山田丘 2-1)。《現在の研究テーマ》計測インフォマティクス、量子ビーム計測。《趣味》ピアノ。



小野 寛太 (Ono Kanta)

大阪大学大学院工学研究科 (〒565-0871 大阪府吹田市山田丘 2-1)。東京大学大学院理学系研究科化学専攻博士課程修了。博士 (理学)。《現在の研究テーマ》自律駆動実験、マテリアルズインフォマティクス、量子ビーム科学。《主な著書》“量子ビーム物質科学”, (共著), (共立出版)。《趣味》街歩き。

E-mail : ono@ap.eng.osaka-u.ac.jp