公開

（事後評価・研究者用）

研　究　報　告　書

「物質の結晶構造を高速に予測するデータ解析技術の開発」

研究期間： 2018 年 10 月～2020 年 3 月
研究者番号： 50174
研 究 者： 鈴木　雄太

1.　研究のねらい

材料開発では、(1)材料の設計、(2)合成、(3)特性評価(実験・計測)のサイクルを回して研究が進められる。これらの各ステップは、機械学習やロボットの活用により効率化や自動化が試みられており、特に特性評価のステップでは、自動計測装置の普及により、手法によっては1日あたり数千サンプルを超える大量のデータが取得できるようになりつつある。しかし、計測データから材料パラメータを抽出するためのデータ解析は、熟練者の経験と勘に頼る部分が多いことから多大な労力と時間を要し、現代の材料開発におけるボトルネックの一つとなっている。

そこで本研究では、材料の最も基本的かつ重要な特徴である結晶構造(原子の並び方)の特徴に着目する。結晶構造を知るための代表的な測定法であるX線構造解析の実験データ(X線回折パターン)を解析するためには、一般的に物理モデルのフィッティングを用いた解析が行われるが、この解析は熟練者でも1データあたり数時間以上を要する複雑な作業である。そこでこの物理モデルを用いた解析に代わり、実験データから結晶構造を高速に予測する機械学習モデルを構築することで、高速かつ自動的なデータ解析を実現する。さらに本研究では、構築したモデルの識別規則を解析することを通じて、これまで熟練者が経験的知識として暗黙に用いてきた法則を抽出し、具体的に定式化することを目指す。

2.　研究成果

(1)概要

材料計測データ解析において時間と手間を要するステップである、物理モデルのフィッティングを機械学習で代替するというコンセプトのもと、X 線回折(XRD)パターンからの結晶構造推定を高速かつ自動で行うことを目指して研究を行った(図 1)。

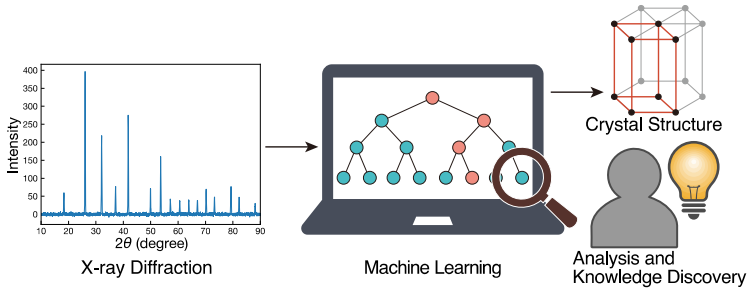

図 1　提案手法の概念図

Advanced Information and Communication Technology for Innovation
ACT-I

まず、結晶構造から XRD パターンを計算することで、約 20 万件の無機物質についての XRD データセットを作成した。このデータセットを用いて、結晶の最も重要な特徴である形（結晶系・空間群）と大きさ（格子定数）を予測する機械学習モデルを訓練した。機械学習アルゴリズムには、実際の実験に用いることを想定して、主に計算量の観点から、決定木を用いたアンサンブル法の一種である Extremely Randomized Tree（ExRT）を用い、物理的考察から、低角側からピーク 10 本の位置およびピーク本数を特徴量として用いた。この結果、それぞれの特徴について、ハイスループット測定データのスクリーニングには十分な予測性能を得ることができた。また実験的に測定した XRD パターンを用いて、実データにおいても提案手法が正しく予測できることを検証した。

学習済みの ExRT から得られる特徴量重要度の情報や、ExRT を近似した決定木の可視化結果を材料科学的な視点から考察することにより、機械学習モデルが獲得した結晶系の分類規則の一部について、物理的な意味合いを解釈することができた。すなわち、熟練者が XRD パターンを一目見るだけで結晶の形を推測できる直感を具体化することができた。

（2）詳細

- XRD データセット作成

  無機結晶構造データベース（ICSD）をデータソースに用い、典型的測定条件（波長：CuK_{α}、2 θレンジ 0°–90°）における XRD パターンを計算した。ICSD を精査したところ、データの欠損や有効数字が小さいなど、本研究に用いるには問題があるデータが含まれることがわかったため、約 3 万物質を除外し、約 16 万物質の XRD パターンを研究に用いた。

- XRD パターンから結晶構造を予測する機械学習モデル構築

  上記データセットを用いて、結晶の最も重要な特徴である形（結晶系・空間群）と大きさ（格子定数）を予測する機械学習モデルを訓練した。実際の XRD 計測では、都度実験セットアップ（データ長、測定範囲、X 線波長など）が変更されることから、この条件に合わせてモデルを再訓練する必要が見込まれた。そこでモデルの訓練が高速であり、かつハイパーパラメータのチューニングが容易であることが重要となる。さらに、クラス分類においては、人間に対して候補を提示するという使い方から、第 2 候補以降の予測結果も得られることが望ましい。そこで機械学習アルゴリズムには、決定木を用いたアンサンブル法の一種である Extremely Randomized Tree（ExRT）を用いた。特徴量には、物理的考察から、低角側からピーク 10 本の位置およびピーク本数の合計 11 変数を特徴量として用いた。この結果、結晶系（7 クラス）および空間群（230 クラス）については 92.2% および 90.8%（Top5）の Accuracy を、格子定数については Cubic の場合で R2=0.9892 を

得ることができ、スクリーニングには十分な精度と考えられる（図 2）。

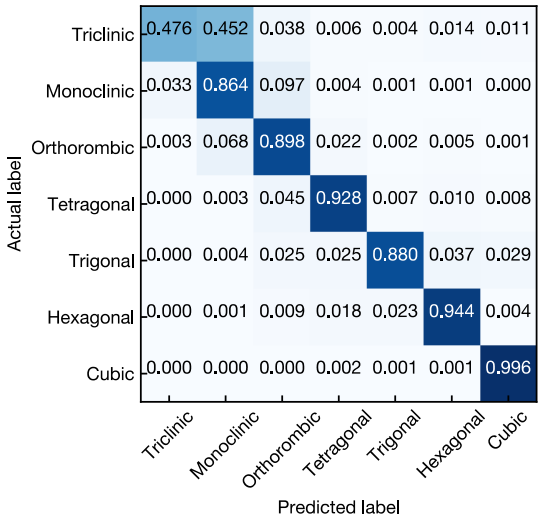

図 2　結晶系予測における confusion matrix

モデルの訓練は一般的ワークステーション（Intel i9 10core, 128GB RAM）を用いて 1 分程度で行うことができ、測定条件を変更する度にモデルを再訓練する場合でも十分に対応が可能である。また二酸化バナジウム（VO2）を例として XRD 測定を行い、実際の実権データにおいても提案手法が正しく予測を行えることを確認した。

- 知識発見

結晶系分類のため訓練した ExRT から特徴量重要度を調べると、ピーク本数および低角側のピーク位置が重要な特徴であった。ピーク本数の重要性は、物理的には反射の消滅則と呼ばれるルールに相当すると考えられ、複雑な結晶構造からは多数のピークが生じるという、ある程度自明な結果であった。しかし低角側ピーク位置についてはその意味は非自明であり、詳細に分析を行った。簡単のため、入力 XRD に対応する結晶が、Cubic 構造か、それ以外かを予測する 2 クラス分類を考える。深さ 2 の単一の決定木により ExRT を近似すると、非常にシンプルなモデルながら 83.4%の Accuracy を得ることができる（図 3）。このモデルでは低角側から 3 本目のピーク位置と、合計ピーク本数の 2 つの変数だけで予測を行っている。この変数を軸として、データの分布と決定木の識別境界を図示した（図 4）。

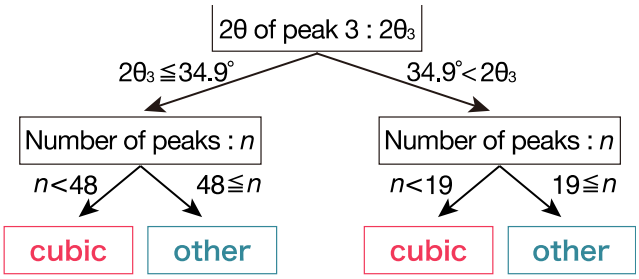

図 3 XRD から Cubic 構造を分類する決定木

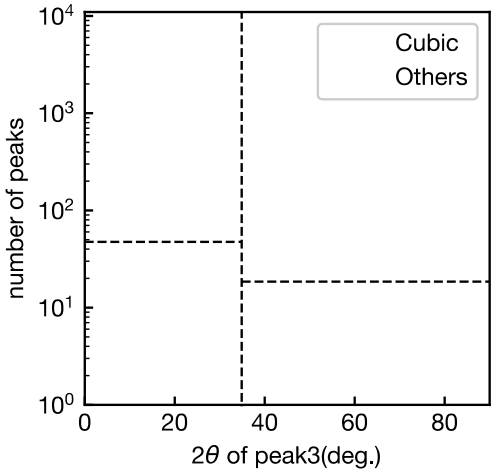

図 4 データの分布と、決定木の識別境界

図 4 から、低角側から数えて 3 本目のピーク位置（peak3）という変数の分布が、Cubic とそれ以外で大きく異なっていることが見て取れる。この結果を考察したところ、対称性の低い（≒複雑な）結晶構造であれば低角側にも多数のピークが生じるため、peak3 はかなり低角側に位置する一方、対称性の高い結晶構造ではピークが少ないため、peak3 は比較的高角側に位置することが考えられた。これはすなわち、ピーク位置の分布が間接的に結晶の複雑さを表現していることを意味する。ヒアリングの結果、この規則は XRD パターンを見た際の熟練者の直感とも整合していることがわかり、これまでは具体化が難しかった熟練者の直感の一部を、機械学習モデルの分析を通じて具体化することができたと考えられる。

3. 今後の展開

1 年 6 ヶ月の研究期間において、情報科学をはじめとする様々な研究者とのディスカッションを通じて、よりメタな視点から本研究の位置づけを俯瞰して捉えることができるようになった。本研究におけるコアコンセプトである、物理モデルを用いたデータ解析やシミュレーションを機械学習により置き換えることにより解析を高速化するというアイディアは、X 線回折データの解析のみならず材料科学一般において有効なアプローチであると考えられる。機械学習に通じた材料科学者という視点を最大限活かして、今後はこの方向性を深堀りし、現時点では機械学習の応用が難しいと考えられているものを含め、様々なデータ・タスクに関して高速かつ高度なデータ解析を実現する手法開発およびその応用を探っていく予定である。この際、各論と、多くの問題に適用できる普遍性とのバランスを取りながら問題設定に取り組む。また、これまでは解析が困難だった大量のデータから、経験則や熟練者の直感を具体化するようなデータマイニングについても引き続き取り組んでいく。これらを通じて、本研究が未来ビジョンとして提案する、人間と AI が協働する材料開発の実現に向け、引き続き尽力していく。

4. 自己評価

- 研究目的の達成状況

研究目的をほぼ達することができ、一部についてはより拡張する形でプロジェクト期間を終えることができたと考えている。

- 研究の進め方（研究実施体制及び研究費執行状況）

研究を進めるにあたって、所属組織において大学院生が JST の研究費を取得した例がなかったため、研究開始および所属機関の異動の際に若干の混乱があったが、各担当者の皆様に丁寧な調整をいただき、実施体制および研究費執行どちらの観点からも、円滑に研究を実施できた。

- 研究成果の科学技術及び学術・産業・社会・文化への波及効果

本研究は材料計測におけるデータ解析という普遍的なテーマを対象にした研究であり、成果は幅広い領域から興味を持たれるものであると考える。特に、学習した機械学習モデルの解析を通じた知識発見を目指す試みからは、これまで明らかにされてこなかった熟練者の直感や経験則を具体化することが見込まれ、Materials Informatics（MI）の分野において、今後ますます重要になるトピックであると考えられる。このような知識発見の取り組みには材料科学の知識とデータマイニングのスキルの両方が要求されるため、自分にしかできない、あるいは自分がやるべき研究であったと自負している。

- 研究課題の独創性・挑戦性

本研究開始時点において、計測データの解析において物理モデルのフィッティングを機械学習により代替する試みは発展途上であり、挑戦的試みであったと考えられる。特に実験に適用するという観点からモデルの性質や計算量に制約をおいて機械学習モデルを構成するというコンセプトは未だ独創的かつ実用上非常に重要であると考えられるため、引き続き研究に取り組み、コンセプトの普及に努めていく。一方、現時点で、古典的な教師あり学習を材料データ解析に応用する取り組み自体は、独創性・挑戦性の観点からは凡庸なアプローチと考えられることも認識している。これは ACT-I 研究を通じた自分自身の成長と、MI の急速な発展により、類似研究が複数発表されたことによると考えており、研究の価値が損なわれるものではないと考える。今後も引き続き、難しい問題を簡単に解くという理念は重視しつつ、具体的な問題解決と、新しい手法の開発のバランスを意識して研究に取り組んでいく。

5. 主な研究成果リスト

(1)論文(原著論文)発表

1. Suzuki, Y. et al. Machine Learning-based Crystal Structure Prediction for X-Ray Microdiffraction. Microscopy and Microanalysis. 2018, **24**, 144-145.

2. Suzuki, Y., Hino, H., Kotsugi, M. & Ono, K. Automated estimation of materials parameter from X-ray absorption and electron energy-loss spectra with similarity measures. *npj Computational Materials. 2019,* **5**, 39.

3. 鈴木雄太, 日野英逸 & 小野寛太. 機械学習を用いた X 線吸収スペクトル解析の自動化. 電気化学. 2020, **88**, 36-41.(招待記事)

(2)特許出願

研究期間累積件数:0 件

(3)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
  - Yuta Suzuki, Hideitsu Hino, Takafumi Hawai, Kotaro Saito, Kanta Ono, "Machine learning approach for on-the-fly crystal system classification from powder x-ray diffraction pattern", TMS 2020 Annual Meeting, 23-27, Feb. (2020)
  - 鈴木雄太, "物質計測における機械学習応用と知識発見", 第 4 回 統計・機械学習若手シンポジウム, 15-16, Nov. (2019), 招待講演
  - Yuta Suzuki, "Machine-learning-aided data analysis in X-ray diffraction and absorption for high-throughput measurement", International Young Researchers Workshop on Synchrotron Radiation Science 2019, 3-4, Sep. (2019), 招待講演
  - Yuta Suzuki, Hideitsu Hino, Takafumi Hawai, Masato Kotsugi, Kanta Ono, "Automated Lattice Constant Estimation of X-ray Diffraction by Ensemble Learning", The 5th International Conference on Electronic Materials and Nanotechnology for Green Environment (ENGE 2018), 11-14, Nov. (2018)

- プレスリリース:
  機械学習により X 線吸収スペクトル解析の自動化が可能に-データの類似度に着目し定量的なスペクトルの解析を実現-
  - https://www.kek.jp/ja/newsroom/2019/04/19/1400/

- 受賞
  - 東京理科大学 奨励賞 (数学・物理部門) (2019 年 3 月)
  - TUS Award 2018 (2019 年 3 月)