

氏 名 鈴木 雄太

学位(専攻分野) 博士(理学)

学位記番号 総研大甲第 2308 号

学位授与の日付 2022 年 3 月 24 日

学位授与の要件 高エネルギー加速器科学研究科 物質構造科学専攻
学位規則第6条第1項該当

学位論文題目 機械学習を用いた結晶構造の分析と物質空間の可視化

論文審査委員 主 査 佐賀山 基
物質構造科学専攻 准教授
小野 寛太
物質構造科学専攻 教授
木村 正雄
物質構造科学専攻 教授
熊井 玲児
物質構造科学専攻 教授
日野 英逸
複合科学研究科 統計科学専攻 教授

博士論文の要旨

氏名 鈴木 雄太

論文題目 機械学習を用いた結晶構造の分析と物質空間の可視化

材料の性質はその結晶構造に強く影響されることから、材料の結晶構造の同定や、材料の結晶構造の特徴を分析することは材料開発において不可欠なステップの一つである。結晶構造の測定手法として広く普及している粉末 X 線回折 (XRD) 法では、最新の放射光施設の XRD ビームラインでは 24 時間あたり 5000 件を超える試料の測定も可能となっている。一方で、データ解析においては、計測した XRD パターンに対して物理モデルに基づきシミュレーションした XRD をフィッティングする手法 (リートベルト解析) が広く用いられており、この解析に熟練者による試行錯誤が必要とされるために、データ解析が研究の時間的ボトルネックになるほか、測定した XRD の一部しか活用されないという状況を生み出している。

XRD パターンを高速かつ自動で分析することができれば、これらの問題が解決できるほか、解析コストの制約や人間の先入観によって十分に調べられてこなかったデータや解析条件を網羅することも可能になり、新たな発見が導かれる可能性も見込まれる。そこでこの問題に対して、リートベルト解析を用いずに結晶の特徴 (結晶系・空間群) を機械学習モデルにより予測する手法の開発 (アプローチ 1) および、数理最適化の技術を応用してリートベルト解析を自動化する (アプローチ 2) という 2 つの相補的なアプローチで取り組んだ。

また結晶構造データの分析の観点においても、機械学習を応用した分析手法の開発に取り組んだ。結晶構造と材料特性は非常に複雑に相関することから、結晶構造同士の類似性を材料の特性にまで踏み込んで定量化することは難しく、類似性の評価は研究者の主観的認識に留まっていた。また結晶構造を人間が精査する必要があるために、多数の構造の類似度を比較したり、結晶構造データベース全体を俯瞰したりといった分析は困難であった。データの持つ抽象的な意味を計算機上で取り扱おうとする問題は、自然言語処理や画像認識分野における重要なテーマであり、計算機で取り扱える固定長ベクトル (embedding) としてデータの抽象的情報を表現する手法が研究されている。結晶構造についてもそのコンセプト (構造の特徴や、構造に由来する物性) を捉えた embedding を得ることができれば、コンセプト類似性に基づく材料の検索や、多数の材料の類似性の可視化、すなわち地図を描くことが可能になるなど、これまでは難しいと思われた新たなデータ分析が可能となる。そこで、材料の抽象的な性質を主観によらず計算機上で定量的に取り扱えるようにすることを目指し、深層学習技術の一種である深層距離学習を用いて材料同士の類似性を学習するアプローチにより、材料の特徴や類似性を捉えた embedding を学習した。

本論文の目的と構成

本論文では、粉末 XRD パターンから自動で結晶構造の情報を得るデータ解析手法の開

発と、結晶構造のコンセプトを捉えた embedding の学習という 2 つのテーマに取り組む。

第 1 章では研究の位置づけについて述べた。

第 2 章および第 3 章では粉末 XRD パターンの自動解析手法の研究について述べた。第 2 章では、これまで蓄積されたデータを基に、XRD パターンから結晶構造の重要な特徴である対称性（結晶系・空間群）を直接予測する高精度な機械学習モデルを構築した（アプローチ 1）。また、高精度な予測が可能な機械学習モデルが構成できたならば、そのモデル内部には予測のための識別規則が存在するはずである。そこで、解釈可能な機械学習アルゴリズムの利用とその分析を通じて、単なる自動化に留まらず、XRD データ分析における「熟練者の勘」や、経験則を具体化することができた。

第 3 章では、XRD パターンへの物理モデルのあてはめを、「シミュレーション条件を与えると、あてはめ誤差を返す関数の最適化」とみなし、数理最適化の一種であるブラックボックス最適化の問題として定式化した（アプローチ 2）。具体的には、粉末回折パターンの分析法として広く普及しているリートベルト法を対象にこのアプローチを適用し、シミュレーションすべき条件の自動推定および、その結果をもとにした次の手の決定を自動化することができ、30-60 分程度の計算時間で熟練者を上回る性能を得ることができた。さらに、手動に比べて多数のパラメータを探索できるため、これまでの手動解析では発見されなかった構造候補の提案も実現した。

第 4 章では、結晶構造が物性を決定するという構造物性相関の原則から、結晶構造の学習を通じて「超伝導体」や「熱電材料」といった材料の機能的特徴（コンセプト）レベルの類似性も自然と現れるのではないかと、という仮説のもと、深層距離学習を用いて結晶構造の機能的類似性を捉えた embedding を得るための研究に取り組んだ。結晶構造の局所特徴（原子近傍の情報）および周期特徴（構造に対応する XRD パターン）を入力として、深層距離学習の枠組みのもとで 10 万件以上の結晶構造データから深層ニューラルネットワークを学習することにより、結晶構造 embedding を得た。この embedding の分析から、銅酸化物超伝導体や、Li イオン電池材料など、重要な物質のコンセプトを捉えた類似性が embedding に反映されていることが確認され、この embedding 可視化することで、10 万件以上の材料について、そのコンセプトレベルの類似性を可視化した材料地図を得ることができた。

まとめ

XRD データ解析に機械学習を応用した手法の開発に取り組み、結晶系および空間群の高速な予測や、リートベルト解析を自動かつ 30 分程度で完了することを可能にした。この過程で、熟練者の知見の具体化や、これまでの手動分析では見つけられない構造候補も発見できることが示された。さらに、深層距離学習を用いることで、結晶構造について材料のコンセプトを捉えた embedding を得る手法を開発し、この embedding を利用することで、結晶構造データのコンセプト類似性に基づく検索や、材料同士の類似性の可視化、すなわち物質空間の地図を描くことを可能にした。

博士論文審査結果

Name in Full
氏名 鈴木 雄太

Title
論文題目 機械学習を用いた結晶構造の分析と物質空間の可視化

本論文は、マテリアルズインフォマティクスを応用して高効率で新材料を探索することを目指して、高速でデータの分析や解析を行うための手法を開発するとともに、これまで定義が難しかった材料のもつ構造や特性の類似性を定量化するために、機械学習を用いた結晶構造の分析を行ったものである。

材料の特性を決める重要な知見である結晶構造を取得するために粉末 X 線回折(XRD)が広く用いられているが、今日では計測の自動化が進み、測定データが短時間で取得できるようになった一方で、データ解析がボトルネックとなっている。鈴木氏はこの状況を打開するために、データ解析を自動化し高効率で行うことに取り組んだ。回折データを高速で分析する技術として、原子構造を仮定せずに XRD パターンから空間群を予測する方法解析法の開発(アプローチ 1)と、XRD データと物理モデルとのフィッティングの自動化(アプローチ 2)という 2 種類の相補的なアプローチを行った。アプローチ 1 では、これまで蓄積されたデータをもとに、XRD パターンから結晶構造の重要な特徴である結晶の対称性(結晶系・空間群)を直接予測する高精度な機械学習モデルを構築した。また、アプローチ 2 では、XRD パターンの解析として広く用いられているリートベルト法による解析を対象に数理最適化の一種であるブラックボックス最適化の問題として定式化し、シミュレーションすべき条件の自動推定を行い、熟練者による解析と同等以上の良好な結果を得た。また、リートベルト解析においてしばしば見られる「フィッティングは良好であるが、物理的に不適切な解が得られる」という問題を解決するために、第一原理計算による構造安定性によって評価する手法を提唱し、XRD のフィッティングと構造安定性を同時最適化することで、より妥当な解を導く取り組みが可能であろうとの展望を述べた。これらのアプローチにおいて、XRD データの解析における、熟練者の勘や経験則を具体化すること成功した。さらに、あいちシンクロトロンを利用した実際の測定に本手法を適用し、1日に100試料の自動計測とデータ解析を行い、技術的可能性だけではなく、社会実装にむけて実用化できる技術であることも紹介があった。

続けて鈴木氏は、大量の結晶構造の情報を活用して、材料同士の類似性を定量化する研究に取り組んだ。材料の結晶構造の局所・周期構造をグラフ構造および XRD として表現し、これら 2 つのモダリティを用いて深層距離学習の枠組みのもとでディープニューラルネットワークを学習することによって、結晶構造の抽象的なベクトル表現(embedding)を得た。銅酸化物超伝導体やリチウムイオン電池材料など、材料として重要な機能をもつ類似性が反映されたベクトル表現が得られたことが確認され、材料のもつ機能の類似性が結晶構造の分析から可視化可能であることを示した。

本審査では、予備審査において指摘された事項(結晶系・空間群の決定における特徴量の

重要度について SHAP 値を用いた分析、物質の類似性の可視化に関して内在的な次元の検討、異なるモダリティへの応用の可能性など)にも適切に対応したことが確認された。

鈴木氏の研究は、これまで熟練者の勘や経験に頼った XRD データの解析を機械学習や深層学習を用いた自動化によって高速な処理を可能にしたのみならず、結晶構造の分析を通じて材料の類似性を可視化したという点で、材料開発におけるマテリアルズインフォマティクスへの応用に関する重要な研究であり、物質科学研究に大いに貢献したと判断される。また、これらの研究成果は、筆頭著者査付き英語論文 2 報が学術誌に採択となっている(さらに、投稿準備中が 1 報、共著論文が 1 報出版済み)。また、国内誌での解説記事が 3 報(うち 2 報が主著)出版されており、筆頭著者での国際会議発表が 2 件(うち 1 件が招待講演)、国内会議での招待講演 4 件がなされている。

以上のことから、本審査委員会では、博士論文審査は合格であると判断した。