

令和2年12月11日

報道関係者各位

大学共同利用機関法人 高エネルギー加速器研究機構
国立大学法人 総合研究大学院大学
大学共同利用機関法人 情報・システム研究機構 統計数理研究所
東京理科大学
国立研究開発法人 科学技術振興機構

X線回折パターンからの対称性予測における知識発見 －熟練者の勘・コツの定式化に成功－

本研究成果のポイント

- 機械学習により粉末X線回折パターンから結晶の対称性を予測する手法を開発
- 機械学習モデルの解析により、熟練者が経験を通して習得した勘やコツなどの経験的知識を定式化することに成功
- 研究者間で漠然と共有されている経験的知識を、計測データの中から発見し定式化することが可能となり科学研究の加速が期待される

【概要】

高エネルギー加速器研究機構（KEK）物質構造科学研究所 量子ビーム連携研究センターの小野 寛太 准教授、総合研究大学院大学 高エネルギー加速器科学研究科 鈴木雄太（博士後期課程2年）らの研究グループは、統計数理研究所、東京理科大学と共同で、物質・材料の構造評価に不可欠な計測データである粉末X線回折パターン※1から機械学習を用いて結晶の対称性※2を予測する手法を開発しました。

さらに機械学習モデルの解析を通じて、これまで明示されていなかった「粉末X線回折パターンを見ただけでおおよその対称性を推定できる熟練者の経験知」を定式化することに成功しました。この研究で用いた解釈可能な機械学習アプローチにより、熟練者の勘・コツのように研究者間で漠然と共有されていた経験的知識を、計測データの中から発見し定式化することが可能となります。それにより、定式化された知識を用いて誰でも熟練者なみの計測データ解析を効率的に行えるようになり、科学研究を加速することが期待されます。

この研究成果は、英国の学術誌「Scientific Reports」に12月11日オンライン掲載予定です。

【背景】

X線・中性子線・電子線などの量子ビームを用いて得られる粉末回折パターンは、物質・材料の機能と性質を支配する結晶構造の情報を得るためには欠かせない測定データであり、物質・材料研究において最も多く収集されている測定データの一つです。粉末回折パターンから結晶構造の情報を得るために必須のデータ解析は、手間と時間がかかる上に熟練した技術を要することが多く、結晶構造解析の課題となっていました。研究グループは、マテリアルズ・インフォマティクス (MI) ※3 に関する研究に取り組んでおり、これまでも、ブラックボックス最適化※4 を用いて結晶解析の自動化に成功し、熟練者を超える解析精度を実現するなどの研究成果を出しています。

【研究内容と成果】

粉末回折パターンの一例を図1に示します。研究グループは、熟練した研究者が回折パターンを見て対称性をおおまかに当てることができることに着目しました。例えば、図1(a)は対称性の高い結晶構造、(b)は対称性の低い結晶構造であることは、熟練した研究者であれば一目見ただけで判断することができます。

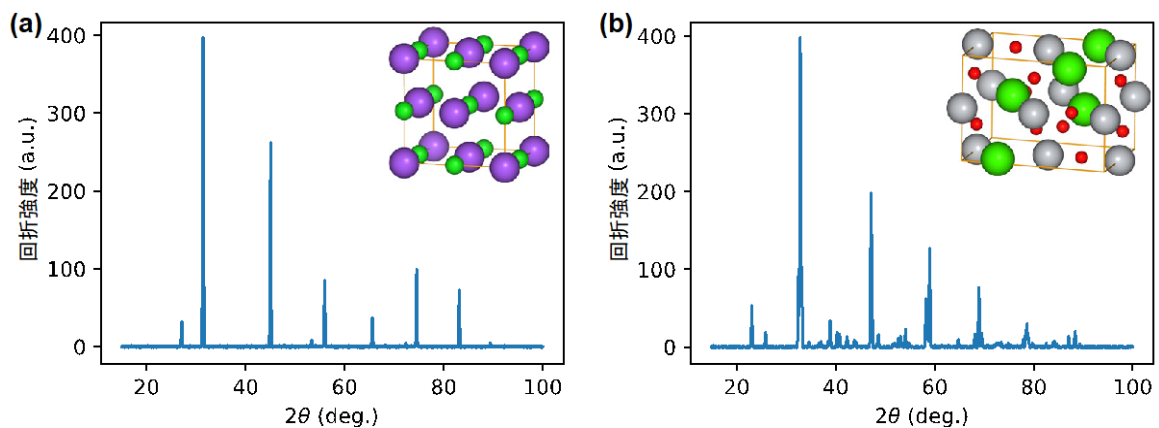


図1 (a) 対称性の高い結晶構造および (b) 対称性の低い結晶構造に対応する X 線回折パターンの例

しかしながら、こういった視覚的印象に基づく判断の基準は定量的に明らかにされていませんでした。研究グループは、これを熟練した研究者間で漠然と共有されている経験的知識であると考え、機械学習を通じてそのルールを明らかにすることを試みました。熟練した研究者が粉末回折パターンから対称性を予測できるなら、機械学習を用いて同様の予測を試み、そのルールを明らかにできるはずです。そのルールを明らかにできれば、熟練者の勘・コツなどの経験的知識を定式化し（知識発見※5）、これまでは関係がないと思われていた事柄に隠れた関係を見出すなど、新たな発見を導く可能性があります。また、ブラックボックスと考えられていた機械学習の予測モデルについて、既知の科学法則に基づいていることを示すことができれば、研究者はその予測モデルを信頼して使えるようになります。本研究の流れを図2に示します。

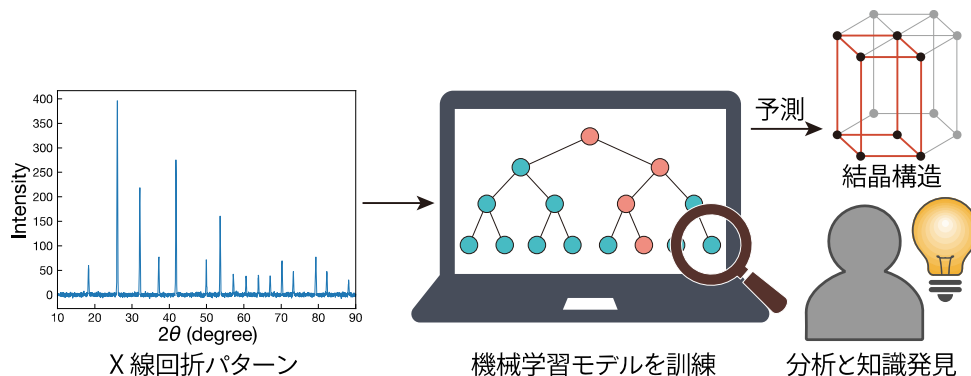


図2 本研究の模式図

X線回折パターンの大量データから機械学習モデルを構築し、機械学習モデルの解釈を通して、熟練者の経験に基づくルールを定式化した。

機械学習を用いて粉末回折パターンから対称性を予測するためには、高精度な機械学習モデルを構築する必要があります。機械学習モデル構築を行うために、結晶構造データベース※6に掲載された結晶構造をもとに20万件の粉末X線回折パターンを計算し学習データを作成しました。このデータをもとに、多数の決定木※7を用いた学習を行い、多数決により予測を行う機械学習モデルを構築しました。その結果、結晶構造の代表的な特徴である結晶系・空間群をいずれも90%を超える精度で自動かつ高速（一般的なノートPCを用いて1件あたり1000分の1秒以下）に予測することができました。

一般に、機械学習モデル内部は非常に多数の複雑なパラメータの集合から構成されており、予測に用いられているルールを人間が直接理解することは困難です。そこで、研究グループは人間でも比較的容易に理解できるシンプルな単一の決定木を用いて、粉末回折パターンから対称性を予測するのに重要なルールを具体的に書き下すことにしました。図3に立方晶※8を識別するための2本の独立した決定木の例を示します。

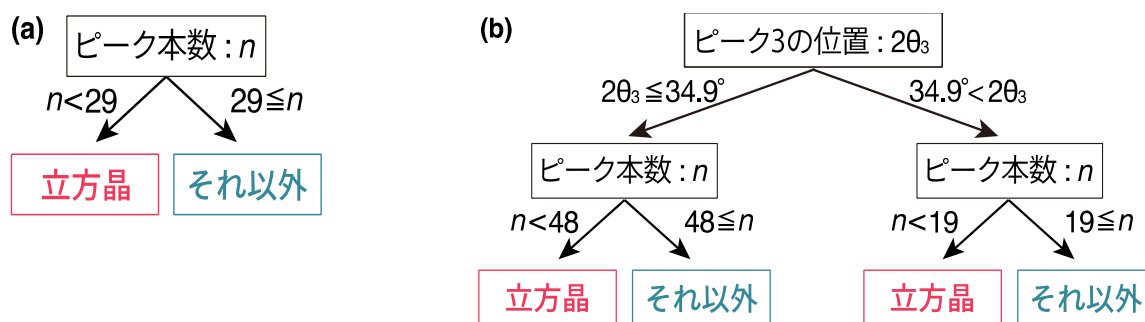


図3 研究により得られた立方晶を予測するための決定木
 予測精度は (a)83.42%、(b)83.20%

左側の非常にシンプルな決定木では、粉末回折パターンのピーク本数が28本以下であることという簡単なルールに従うだけで、83%以上の正解率で立方晶の構造であることを予測できます(図3(a))。大量のX線回折データから自動的に抽出された右側の決定木(図3(b))のルールは一見複雑に見えますが結晶学の基礎的な知識によって説明

が可能であり、熟練者の直感とも一致していることが確認されたことから、熟練者の勘・コツを定式化できたと結論づけました。シンプルなルールを複数組み合わせ多角的に判断することにより、より高い精度で予測できることが示唆されます。このようなルールは物理法則に基づく演繹的アプローチから導くことは困難であり、大量のデータに基づく帰納的アプローチを用いることで具体化されたものです。本研究で用いた解釈可能な機械学習手法により、これまで認識されていなかった実験データに内在するさまざまな経験則がデータ駆動アプローチにより発見されることが期待できます。

本研究はJST未来社会創造事業 探索加速型「共通基盤」領域の研究開発課題「数理学を活用したマルチスケール・マルチモーダル構造解析システム（研究開発代表者：小野 寛太）」（JPMJMI19G1）で実施されました。また、本研究の一部は、JST 戦略的創造研究推進事業 ACT-I研究領域「情報と未来」の研究課題「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（研究者：鈴木 雄太）」（JPMJPR18UE）により行いました。

< 論文情報 >

Symmetry prediction and knowledge discovery from X-ray diffraction patterns using an interpretable machine learning approach

著者・所属 鈴木雄太^{1,2} 日野英逸³ 羽合孝文¹ 斉藤耕太郎¹ 小嗣真人⁴ 小野寛太^{1,2}

1：KEK 物質構造科学研究所、2：総合研究大学院大学、3：統計数理研究所、4：東京理科大学

雑誌名 Scientific Reports

DOI：10.1038/s41598-020-77474-4

【本研究の意義、今後への期待】

物質・材料研究では熟練した研究者や技術者の間で経験的に取得され、暗黙的に共有される勘や経験則が多く、これらを効率的に獲得することは一般的に困難でした。本研究の手法を適用することにより、熟練者の勘・コツの定式化が可能となり、経験則の効率的伝承や研究の効率化につながることを期待されます。また、一連の成果を発展させることにより、熟練者の経験的知識の定式化を通じて新しい発見が導かれることが期待され、今後の物質・材料研究の加速や効率化と新たな物理現象に関する知識発見に貢献します。

【用語解説】

※1. 粉末X線回折パターン

物質にX線を照射すると、X線は物質を構成する原子の周りにおける電子により散乱されます。原子や分子が規則的に並んだ物質（結晶）の場合、散乱されたX線は原子や分子の並び方（結晶構造）に応じて回折パターンと呼ばれる独特な強度分布を示します。この現象を利用して結晶構造を調べることができます。X線回折（XRD）は、1つの結晶を対象とする場合と結晶を砕いた粉末試料を対象とする場合に大別されます。

後者は粉末X線回折(PXRD)と呼ばれ、試料の準備や測定が簡便なことから物質・材料研究において広く用いられる一般的な測定手法となっています。

※2. 結晶の対称性

結晶中における原子は空間的に規則的な繰り返しパターンに従い配列しているため、この周期性に着目して結晶構造を記述することができます。この周期性の特徴を結晶の対称性と呼び、いくつかの分類の仕方があります。本研究では、結晶構造の最小単位を切り出した単位構造の形について分類した結晶系(7種類)と、単位構造内の原子の対称操作(並進・回転・鏡映など)の集合である空間群(230種類)の2種類の対称性を予測の対象としました。

※3. マテリアルズ・インフォマティクス(MI)

情報科学や機械学習を材料科学に応用することにより材料の研究開発の効率化・高度化を目指す学問領域。

※4. ブラックボックス最適化

数理最適化問題のうち、目的関数や制約条件が解析的に与えられないようなものを、ブラックボックス最適化問題と呼びます。ブラックボックス最適化は、機械学習モデルのハイパーパラメータ最適化など、複雑なプログラムやシステムを対象とした実問題において実績があり、最近では結晶構造解析の自動化にも用いられています。

参考：プレスリリース2020/06/04

結晶構造解析の自動化～ブラックボックス最適化により熟練者を上回る解析精度を達成～
<https://www.kek.jp/ja/newsroom/attic/PR20200605.pdf>

※5. 知識発見

情報科学分野で用いられる用語で、データやルールから人間にとって意味のある知識を発見する学問領域および技術のこと。発見された知識が人間にとって新しいものなのか、既知のものかについては問いません。

※6. 結晶構造データベース

本研究では、結晶構造データベースとして、世界最大級の無機結晶構造データベースであるICSDを用いました。ICSDは科学文献等から収集され人力でキュレーションされた約20万件の結晶構造データベースであり、FIZ Karlsruhe(ドイツ)により作成されています。

※7. 決定木(けっぺいぎ)

決定木は機械学習アルゴリズムの一種で、入力データがある条件に当てはまるかどうかの条件分岐を繰り返すことで予測を行います。識別ルールは分岐条件の集合であるため、条件を列挙することで人間による解釈が可能です。

※8. 立方晶

7つの結晶系の一つ。結晶構造の各辺の長さが等しいサイコロ状で、もっとも対称性が高い結晶系です。鉄をはじめとする多くの金属や、シリコンなどの半導体材料など、実用上重要な材料が多く属します。

【本件に関するお問い合わせ先】

大学共同利用機関法人 情報・システム研究機構 統計数理研究所

運営企画本部 企画室 URAステーション

Tel: 050-5533-8580

Fax: 042-526-4348

e-mail: ask-ura@ism.ac.jp