

# 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

## 『材料開発を加速する自動データ解析』

鈴木 雄太<sup>1,2</sup> 1. 総合研究大学院大学, 2. 高エネルギー加速器研究機構



### 解決したい課題と背景

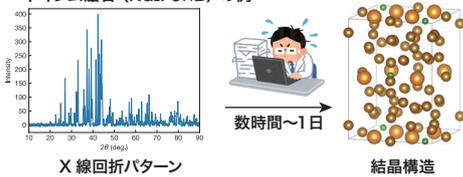
- 研究のボトルネックを解決し、材料開発を加速するデータ解析の効率化は、材料開発の加速に直結



- 優れた材料は豊かな社会の基盤  
高性能磁石、リチウムイオン電池、青色LEDの実現など、優れた材料は社会を一変させるインパクトがある。材料開発の効率化は、社会全体に波及する基盤技術

- X線回折 (XRD) : 結晶構造を調べる基本  
物質の性質は結晶構造 (原子の並び方) に支配される。XRDを使った結晶構造解析は、最も普及している材料解析手法の一つ。材料開発、創薬、美術品鑑定まで、様々な分野で活用されている。

- 生のデータを見てもよくわからない  
計測データの解析には、試行錯誤と経験が必要  
ネオジム磁石 (Nd<sub>2</sub>Fe<sub>14</sub>B) の例



### 研究の目的

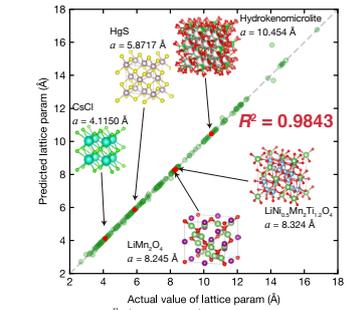
- 機械学習を用いて、XRD データから高速に結晶構造を推定する手法を開発する。
- さらに、構築した機械学習モデルの解析やデータマイニングにより、「熟練者の勘」の具体化や、材料科学における新しい知識の発見を目指す。

### 結果 : XRD からの結晶構造予測

| Actual label \ Predicted label | Triclinic | Monoclinic | Orthorhombic | Tetragonal | Trigonal | Hexagonal | Cubic |
|--------------------------------|-----------|------------|--------------|------------|----------|-----------|-------|
| Triclinic                      | 0.476     | 0.452      | 0.038        | 0.006      | 0.004    | 0.014     | 0.011 |
| Monoclinic                     | 0.033     | 0.864      | 0.097        | 0.004      | 0.001    | 0.001     | 0.000 |
| Orthorhombic                   | 0.003     | 0.068      | 0.898        | 0.022      | 0.002    | 0.005     | 0.001 |
| Tetragonal                     | 0.000     | 0.003      | 0.045        | 0.928      | 0.007    | 0.010     | 0.008 |
| Trigonal                       | 0.000     | 0.004      | 0.025        | 0.025      | 0.880    | 0.037     | 0.029 |
| Hexagonal                      | 0.000     | 0.001      | 0.009        | 0.018      | 0.023    | 0.944     | 0.004 |
| Cubic                          | 0.000     | 0.000      | 0.000        | 0.002      | 0.001    | 0.001     | 0.996 |

格子系予測における混同行列

- 結晶系 (7クラス) の分類精度は **92.24%** (上図)、空間群 (230クラス) について **92.42%** (Top-5 Acc.) を得た。
- 測定中の on-the-fly 解析に用いるためには十分な精度が得られた。



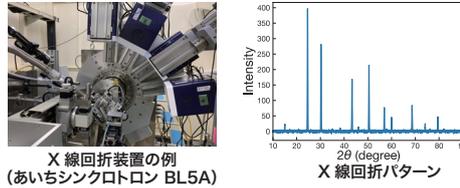
格子定数 a の予測結果

- Cubic system の格子定数予測では、mean percentage error (MPE) は **0.83%**、**R<sup>2</sup> = 0.9843** となり、高い精度を得た。
- 70% 以上のデータでは MPE < 0.01%

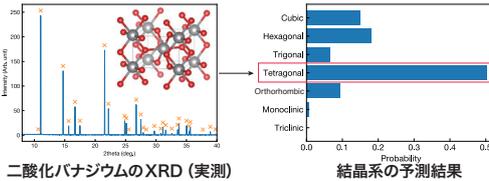
### ACT-I 期間中の成果

XRD 測定 → 機械学習モデルを用いた予測までの一連のワークフローを実証した

#### X線回折実験



#### 1. 高精度な結晶構造予測モデル構築

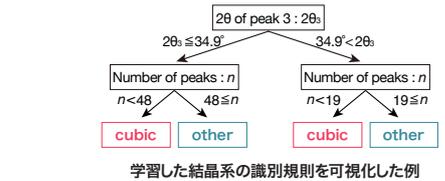


- 数時間かかる解析が1秒未満に  
既存手法を上回る精度を得たほか、自動解析は難しいとされてきた実際の実験データについても、高速かつ適切な予測ができた。

#### 機械学習を用いた予測と知識発見



#### 2. 機械学習モデルの解析



- 熟練者の経験則を具体化  
学習した機械学習モデルを解析し、熟練者の直感 (XRD をひと目見ただけでおおよその構造がわかる) の一部を具体化でき、物理的にも理に合った結果を得た。

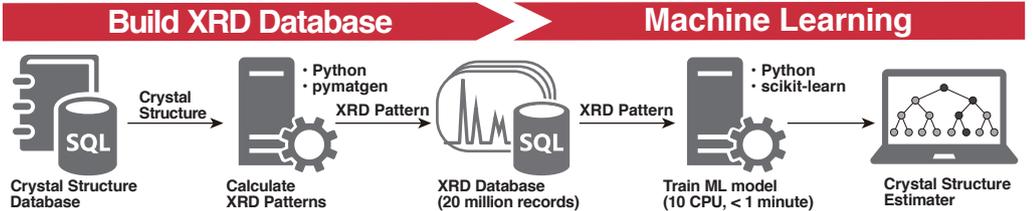
### 今後の展望

- 結晶構造・材料特性のより高度な推定  
形や大きさだけでなく、原子位置まで含めた構造推定の自動化、様々な計測モダリティを統合した材料特性推定の実現を目指し、引き続き研究に取り組む。

### 目指す未来ビジョン

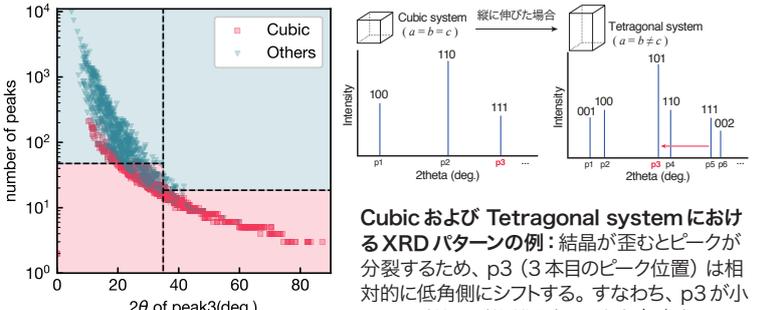
- 人間とAI、ロボットが協働する材料開発の実現  
研究者の発想やデータ分析を支援するAI、自動実験を行うAIとロボットなどにより、より優れた材料を迅速に生み出し世に届けられる世界の実現を目指す。

### 研究の流れと手法



- データソースとして、結晶構造データベース (ICSD 2018.1) を用いた。約20万件のデータのうち、情報に欠損や誤りがあるものを除き、15万件を実験に用いた。
- XRD パターンの計算には材料解析用ライブラリ pymatgen を用い、X線波長はCuのKa特性X線 (1.5418 Å)、2θ範囲0° - 90° の典型的な条件でXRDを計算し、データセットとした。
- XRD の特徴量として、低角側から順に10本のピークの位置 (2θ) および0° - 90° における合計ピーク本数を用いた。
- 機械学習アルゴリズムには、決定木のアンサンブル法の一つである Extremely Randomized Tree (ExRT) を用いて、結晶系、空間群、および格子定数の予測モデルを構築した。

### 結果 : 結晶系分類規則の解析と知識発見



XRD から、結晶が cubic system かそれ以外かを予測する2クラス分類の決定木の可視化 : 予測精度は 83.32%

- 識別規則を調べるため、近似としてシンプルな決定木を用いて、入力 XRD に対応した結晶が Cubic 構造か、それ以外かを予測するモデルを訓練した。

- 可視化した決定木の解釈から、3本目のピーク位置が結晶の複雑さを表現しているという物理的に非自明な発見があった。
- 「熟練者の勘」の一部が具体化できた  
熟練した研究者は、XRD パターンを見ただけで直感的におおよその結晶構造がわかる場合がある。