

研究終了報告書

「高効率な新物質発見のための合成手法推薦システムの構築」

研究期間：H29 年 10 月～R3 年 3 月

研究者：林 博之

1. 研究のねらい

本研究対象とした無機物質の場合、既知物質の機能を大きく凌駕する新規物質が発見できれば、物性発現機構の理解を深化するだけでなく、産業のイノベーションにも繋がるのが期待される。計算材料学の長足の進歩により多数の候補物質の熱力学的安定性を平衡論的に評価することが可能になった。しかし、それは合成可能性の必要条件を与えるだけであり、合成手法や条件については何ら情報を与えないため、実際の新物質合成に応用するには依然として大きなギャップがある。現在においても未発見な物質は合成条件がシビアな場合も多い。その段階で試行錯誤的な合成プロセスを採ることになれば、発見効率を大きく損ねてしまうのである。結果的に合成できないケースが大半でも、その判断には月から年単位の時間を要してしまう。このような個々の研究者による失敗データは、公開はおろか整理されることもなく、他者が活用することも全くなかった。物質ごとに合成条件とその成否が整理されている合成条件データベースは現在皆無である。データ科学の観点に立つと、その構築こそが効率的な物質探索を行う上での急務となる。これまで各研究者により蓄積されてきた実験データをうまく整理したうえで、合成が成功する可能性の高い条件をピンポイントで指定し、合成プロセスを高効率化するシステムが開発できれば、その効果は極めて大きいと期待できる。そのため、本研究では以下の二つの研究項目に注力する。

研究項目 1 並列合成手法をもちいた実験条件データベースの作成

研究項目 2 合成可能性を定量的に評価する合成条件推薦システムの構築

並列合成実験により獲得した多数の既知物質の合成実験データを整理し、物質ごとに合成条件(手法、組成、熱処理温度や時間、雰囲気や水溶液の pH など)と合成結果(目的物質単相であれば高評価、未反応や相分離していれば低評価)をデータベースとして構築する。データベースを用いて推薦システムにより合成に成功する可能性が高いと予測した合成条件を実際に試行し、その結果を再度データベースへフィードバックする。このシステムでは対象が未知物質であっても、指定した原料や熱処理条件等での程度合成が成功するかを定量的に評価することが可能である。機械学習により、人力では扱えない多数の物質及び合成条件データから、合成条件ごとに合成成功確率を定量的に評価できることが本手法の最も重要な点である。

2. 研究成果

(1) 概要

本研究では、まず擬二元系酸化物の合成可能条件の予測を目的とし、無機結晶データベースにある既知組成およびデータベースにない未知の組成に対して並列合成手法により収集した合成結果を合成条件データベースとして整理した。次に、原料 2 種とその混合比、前駆体合成手法と焼成温度の 5 つを軸としたテンソルデータに変換し、各要素に合成結果に対して点数を付与した。この際、目的とする物質ができた場合高い点数を与えるようにした。このテンソ

ルデータに対し、テンソル分解を用いた推薦システムにより未実験条件の点数を予測し、点数の高い未実験条件を優先して合成実験を行うことで予測能力の検証を行うと共に、新規の擬二元系酸化物を探索した。

以下、研究項目ごとの概要を記す。

研究項目 1 並列合成手法をもちいた実験条件データベースの作成

酸化物の合成手法として一般に用いられている手法のうち、固相反応法、錯体重合法、噴霧共沈法、環状エーテルゾルゲル法の4つの合成手法を並列化して計 3000 件の合成データの収集を行った。合成結果は X 線回折法と粉末回折データベースを用いた相同定により評価した。過去の文献にある合成データを収集する方法に比べ、焼成条件等をそろえた均質なデータを得ることができるだけでなく、研究項目 2 で行った予測結果に対し自ら検証実験を行いデータベースにフィードバックさせることも可能である。取得したデータの一部は公開した。

研究項目 2 合成可能性を定量的に評価する合成条件推薦システムの構築

研究項目 1 で作成したデータベースを基にしたテンソルデータをテンソル分解の一種である Tucker 分解することで未実験条件の点数を予測する推薦システムを構築した。推薦システムの精度を既知データの交差検定で評価した結果、予測得点は合成の成功率と比例関係にあることが示唆され、得点が高いグループでは合成に成功し、逆に低いグループでは目的組成の酸化物が得られないことが分かった。また、学習データに含まれていない未実験の既知酸化物 17 種類を予測した合成条件で実際に合成を行ったところ、得点が高い合成条件では 8 割以上の合成条件で目的組成の酸化物を得ることができた。

(2) 詳細

研究項目 1 並列合成手法をもちいた実験条件データベースの作成

本研究項目では、酸化物の合成手法として一般に用いられている手法のうち、固相反応法、錯体重合法、噴霧共沈法、環状エーテルゾルゲル法の4つの合成手法を並列化して計 3000 件の合成データの収集を行った。固相反応法は原料となる酸化物粉末をあらかじめよく粉砕し、指定の陽イオン比率になるように混合した粉末を樹脂製カプセルにジルコニアボールとともに封入してボールミルで攪拌した。このような作業は 100 個程度の試料を同時に処理することが可能であり、攪拌後の試料は自作の簡易プレスで圧粉した後に燃焼ボートに乗せ、多数の試料を同時に焼成した。錯体重合法では、あらかじめ調製した原料のストック溶液をピペットを用いてるつぼに直接分注し、その後超音波振動による攪拌とホットプレートによる重合および乾燥過程を 100 試料程度同時に進行できるようにした。乾燥後の前駆体はるつぼを電気炉内に並べてまとめて焼成を行った。噴霧共沈法は所定量混合した原料溶液をエアブラシに通して、沈殿剤が含まれる溶液に噴霧したのち、沈殿物を遠心分離器により回収し、上と同様にまとめて焼成を行った。環状エーテルゾルゲル法では、所定量を混合した原料溶液に酸化プロピレンを滴下し、生成した沈殿物を遠心分離により回収し、同様にまとめて焼成を行った。このようにして作製した多数の試料の X 線回折測定は、本研究課題で購入した 120 試料を自動的に交換して測定を行える X 線回折装置により効率的に評価した。測定した回折プロフ

ファイルを、既知の粉末回折データベースである International Centre for Diffraction Data (ICDD)にある回折プロファイルと照合することで結晶相の同定を行った。この解析も自動化を行い、人間が解析した場合とほぼ同様の結果が得られることを確認した。本研究で取得した合成データの一部は研究項目 2 の結果とともに公開した[研究成果(論文発表)1]。このような均質で大規模な酸化物の合成条件データベースは世界的に例がなく、機械学習を用いた合成研究分野に先鞭をつけることができた。

研究項目 2 合成可能性を定量的に評価する合成条件推薦システムの構築

合成データベースにある既知の実験条件から、未実験の実験条件における目的物質の合成成功確率を予測するシステムを提案した。研究項目 1 で作成したデータベースを原料 2 種とその混合比、前駆体合成手法と焼成温度の 5 つを軸としたテンソルデータに変換し、各要素に合成結果に対して点数を付与した。目的物質の組成と同じ仕込み組成で合成できた場合に高い点数を与えた。これをテンソル分解の一種である Tucker 分解することで未実験条件の点数を予測する推薦システムを構築した。合成結果が全くの未知である物質に対する予測能力を評価するために、ある組成に対する合成結果をすべてテストデータとし、それ以外のデータを学習データとする組成ベースの交差検定を行った。

その予測得点は合成の成功率と比例関係にあることが示唆され、得点が高いグループでは合成に成功し、逆に低いグループでは目的組成の酸化物が得られないことが分かった(図 1)。また、学習データに含まれていない未実験の既知酸化物 17 種類を予測した合成条件で実際に合成を行ったところ、図 2 に示すように得点が高い合成条件では 8 割以上の合成条件で目的組成の酸化物を得ることができた。本手法では、図 3 に例示するように原料間の類似性を評価することも可能であり、合成研究者の経験則と矛盾しない結果が得られている[研究成果(論文発表)1]。本手法を用いて新規の擬二元系酸化物の合成に成功しており、その合成可能条件と相関が強い既知物質の合成条件を定量的に評価する手法の構築も行った[論文執筆中]。

上記手法の発展形として、以下の研究も行った。擬二元系酸化物と、擬三元系酸化物などより多くの元素を含む系では、上述の手法ではテンソルの次元が異なるためデータを流用することはできない。しかし、多元系酸化物は元素および混合比の組み合わせにより探索範囲が飛躍的に大きくなるため、より単純な系のデータから予測モデルをた

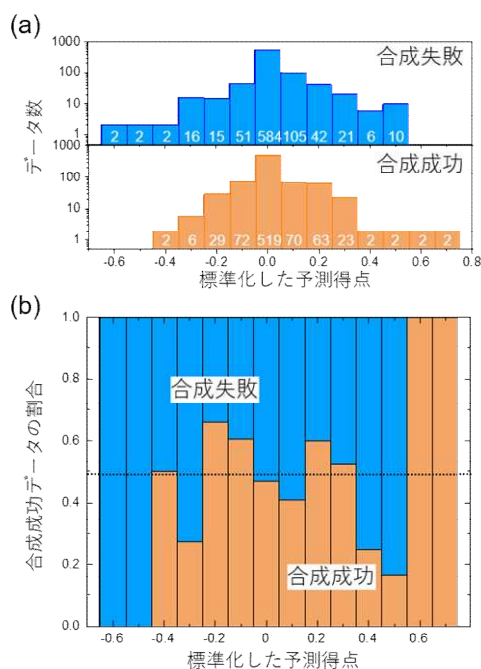


図 1 交差検定による推薦システムの予測能力評価。

てること効率的に探索ができる手法が望まれる。そこで、擬二元系酸化物データの Tucker 分解で得た原料の類似性を用いて合成データを系の構成元素数に依らない数値ベクトルで表現し、擬二元系酸化物の合成データのみから擬三元系酸化物の合成結果を予測する学習モデルの構築を行った[論文執筆中]。

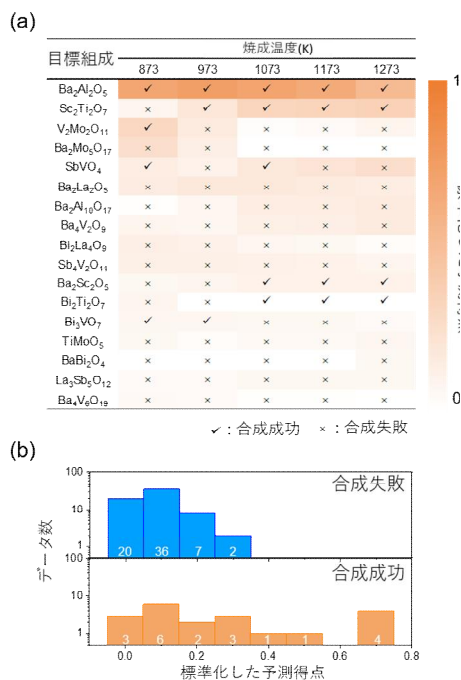


図 2 未実験物質の予測得点と実証実験結果。

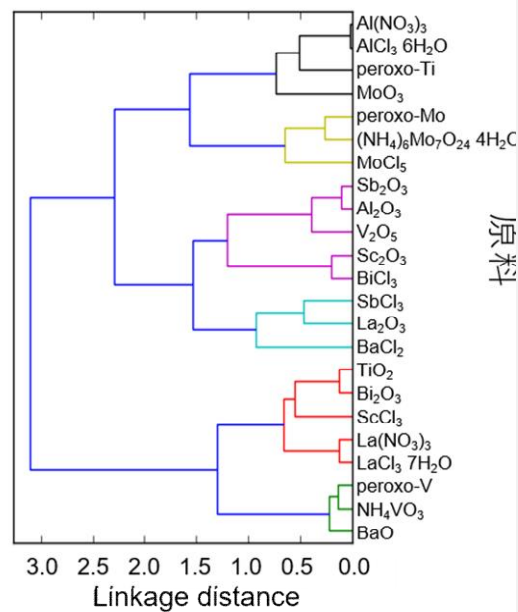


図 3 本研究で得た原料間類似性のデンドログラム。

また、同様の手法を用いて実験条件ではなく化学組成の存在確率を大規模な既知物質データベースである ICSD から予測する研究も行った[研究成果(論文発表)3 および 4]。その他にも、第一原理計算により求めた比誘電率と実験値との差を機械学習し、高い比誘電率を有する物質を予測する研究も行った[研究成果(論文発表)2 および 5]。

3. 今後の展開

研究項目 1 並列合成手法をもちいた実験条件データベースの作成

本研究で獲得した合成条件データベースは、推薦システムの実装スクリプトと併せて随時論文上で公開していくことを検討している。同じ合成手法でも、微妙な合成条件の違いにより研究者によって合成結果が異なる可能性があるというのは合成研究分野ではよく知られた問題である。現在公開しているのは原料や混合比および焼成温度といった限られたパラメータのみであるが、実際にはより詳細に記録されている合成条件も公開していくことで、合成研究分野で利用しやすいデータを提供できると考えている。

研究項目 2 合成可能性を定量的に評価する合成条件推薦システムの構築

今後も合成条件データベースが拡充されるたびに、新規物質の合成可能性予測は更新されていくため、予測結果を随時公開できるような仕組みを検討したい。新規物質だけでなく、

既知物質であってもこれまで知られていない条件で合成できることがわかれば、粒径や表面形状の制御を目的とした合成研究など物質探索以外にも波及効果を見込めると考えている。

4. 自己評価

研究目的の達成状況

本研究で設定した目的は、物質ごとに合成条件とその合成成否が整理された合成条件データベースの整理と、それを活用して未実験物質であってもその合成可否を予測できる合成条件推薦システムの構築であった。研究項目1においては、様々な合成手法及び焼成温度における1000件以上の合成データベースを世界に先駆けて公開することができた。また研究項目2では、未実験物質であってもその合成可能条件の予測にテンソル分解を適用した推薦システムが有用であることを示した。よって、本研究目的は達成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)

並列合成装置の構築や、X線回折プロファイルの自動解析スクリプトの作成、および合成条件データベースの整理とテンソル分解を利用した合成条件推薦システムの構築の大部分を代表研究者が単独で実行した。

本研究で導入した物品は多検体粉末X線回折測定装置と雰囲気制御用ボックス炉である。多検体粉末X線回折測定装置では、連続して120個の試料を測定することが可能であり、雰囲気制御用ボックス炉等で合成した試料のX線回折プロファイルを効率よく測定し、合成条件データベースを拡充した。

研究成果の科学技術及び社会・経済への波及効果(今後の見込みを含む)

無機物質の合成研究はこれまで合成研究者の試行錯誤に基づく経験に依っていた部分があったが、大量の合成データから機械学習手法によりオンライン学習的に次の候補の合成候補を逐次選択する合成プロセスは物質探索分野において革新的な取り組みであり、本研究の並列合成手法と合成条件推薦システムはそれを達成するための基礎となるものである。現状のデータベースは酸化物の合成可否を対象としたものであるが、窒化物や硫化物の合成を得意とする研究者や企業が本手法を同様に適用することは可能であり、合成可否ではなく物性値を目的とした最適な合成条件を推薦するように本方法を応用することで、産業界への波及効果も見込める。本研究で取得したデータの一部および開発したコードは公開済みであり、学会発表等を通して周知を図りたい。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 5件

1. H. Hayashi, K. Hayashi, K. Kouzai, A. Seko, and I. Tanaka, "Recommender System of Successful Processing Conditions for New Compounds Based on a Parallel Experimental Data Set", Chem. Mater. 31, 9984-9992 (2019).

本研究では並列合成実験による合成条件とその結果を紐づけた1000件規模の合成

実験データベースの作成と、そのデータを活用して未知物質であってもその合成条件を予測することが可能である合成条件推薦システムの構築を行った。交差検証および予測に基づいた実際の検証実験により、1000件程度の合成実験結果から約20万件の探索範囲において合成可能な条件を予測可能であることを示し、特に推薦上位では高い成功率になることを示した。

2. A. Seko, H. Hayashi, H. Kashima and I. Tanaka, “Matrix- and tensor-based recommender systems for the discovery of currently unknown inorganic compounds”, Phys. Rev. Materials 2, 013805 (2018).

本研究では既知物質の化学組成を構成元素とその構成比というカテゴリカルデータに変換してテンソル型データベースを構築し、行列分解及びテンソル分解手法を用いて未知化学組成の存在を予測した。予測結果の検証として、多数の第一原理計算による相安定性評価と比較を行い、予測した組成は第一原理計算においても熱的安定な結晶構造を有することを示した。

3. A. Seko, H. Hayashi, I. Tanaka, “Compositional descriptor-based recommender system for the materials discovery”, J. Chem. Phys. 148, 241719 (2018).

本研究では既知物質の化学組成を構成元素の原子番号や電気陰性度といった特徴量の統計量というベクトルに変換し、ランダムフォレスト等の分類器を用いて未知化学組成の存在を予測した。予測結果の検証として、多数の第一原理計算による相安定性評価と比較を行い、予測した組成は第一原理計算においても熱的安定な結晶構造を有することを示した。

(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)

主要な学会発表

- H. Hiroyuki, A. Seko, and I. Tanaka, Recommender system of processing conditions for inorganic compounds based on a parallel experimental dataset, The Materials Research Society 2019 Fall meeting (Boston, Massachusetts), 2019/12/5.
- H. Hiroyuki, A. Seko and I. Tanaka, Similarity between synthesis conditions by tensor factorization of synthesis database, PRESTO International Symposium on Materials Informatics (The University of Tokyo, Tokyo), 2019/2/9.

著作物

- 田中 功、世古敦人、林 博之、東後篤史、” マテリアルズインフォマティクスによる新材料の開拓”、現代化学、590号、(2020).