

令和 4 年 5 月 30 日現在

機関番号：14301

研究種目：挑戦的研究（萌芽）

研究期間：2019～2021

課題番号：19K22054

研究課題名（和文）データ科学に基づく新奇な無機化合物探索の実践

研究課題名（英文）Data-driven exploration of novel inorganic compounds

研究代表者

田中 功（TANAKA, ISAO）

京都大学・工学研究科・教授

研究者番号：70183861

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：本研究は、無機の『未知化合物』を系統的に発見するための方法論を構築し、実際に合成・解析実験によって存在を確認することを目的に、様々な物質に対して多岐にわたるパラメータを系統的に変えた合成実験を行い、成功データだけでなく失敗データをも系統的に収集した。この結果をスコア化し、効率的なテンソル分解法を活用して、成功確率が高いと予測された実験条件で検証実験を行い、2つの新物質を発見した。このように、データ科学に基づいて新規な無機化合物発見を志向し機械学習と実験を組み合わせた研究は世界的にも例を見ず、新物質や材料機能探索にマテリアルズ・インフォマティクスを活用した重要な試金石を与えることができた。

研究成果の学術的意義や社会的意義

本研究の観点は、従来広く行われてきた物質機能を向上させるための研究とは方向性が異なるものであり、世界的にもほとんど類を見ない挑戦的なものである。これら要素技術を未知化合物発見の目的に合うように組み合わせ、具体的な方法論として確立できると、新しい材料科学の流れを創り出すことができる。これは材料科学の未踏領域に挑戦する大きなテーマであり、目的が達成できれば、材料科学全般に大きな波及効果を持つ成果になると期待される。

研究成果の概要（英文）：This study developed a data-driven methodology systematically to discover inorganic "yet-to-be-discovered compounds." We carried out the synthesis and analytical experiments to verify it. A tensor decomposition method was adopted to predict successful synthesis conditions. We found two novel compounds based on the prediction. We believe that this unprecedented research combining machine learning and synthesis experiments to discover new inorganic compounds can provide an essential touchstone for the use of materials informatics in the search for new materials and material functions.

研究分野：材料基礎科学

キーワード：無機化合物探索 データ科学 合成条件推薦システム 並列合成実験

1 . 研究開始当初の背景

新物質発見 , あるいは既知物質の新機能発見に端を発して物質・材料科学の一分野が大きく進歩した例は枚挙に暇がない . 一旦 , 新物質が発見されると , 多くの研究者が周辺物質の探索に集まり , 詳細な実験や理論計算を基に機能発現メカニズムを議論し , その結果として新しい学術分野と技術が構築される . したがって新物質の発見は , 物質・材料科学に大きな飛躍を与える重要な契機となる . しかし従来の多くの研究では , 新物質の発見は , 膨大な網羅的探索の結果 , あるいは偶発的に成されたものが多かった . これは極めて効率の低いプロセスであり , 換言すれば効率改善の余地が大きいプロセスである .

本研究代表者は , この問題意識のもとで 10 年以上前からデータ科学の活用に興味を持ち情報収集を行っていた . 2012 年度には , 領域代表として科学研究費・新学術領域研究「ナノ構造情報のフロンティア開拓-材料科学の新展開」を立案したが , その際には材料科学と情報科学の両方が理解できる研究者は , 未だ国内はおろか , 国際的に見てもほとんど存在しなかった . しかし領域研究の中で研究者間の意思疎通を図り , 研究目標や専門用語を共有することに努めた結果 , データ科学を活用した材料科学研究のあるべき姿を認識することができた . また , 重要な要素技術となる推薦システムなどの利用方法や , 並列実験についても着想を得た . 本研究は , このアイデアを具体的に新物質探索のための一連の流れとして実践するものとして立案された .

2 . 研究の目的

固体化学の長い歴史の中で発見され , 無機結晶データベース(ICSD)などのデータベースに結晶構造が収録されている無機の『既知化合物』は約 10 万件あるが , これらは熱力学的に安定な化合物のごく一部と考えられる . 周期表から 66 種の陽イオン元素と 11 種の陰イオン元素を選び , 3 元から 5 元系までの単純な組成比を考えるだけでも 231 億とおりの組み合わせがある . このうち熱力学的に安定・準安定な化合物で『未知化合物』が多数存在すると想定できる . 本研究はこのような新奇な無機化合物を系統的に発見するための方法論を構築し , 実際に発見することを目的としている . 発見するというのは , 予測や計算に留まるのではなく , 実験によって存在を確認するという意味である . 我々のこれまでの経験で , 第一原理計算によって熱力学的に安定であることが保証された化合物であっても , 合成困難なものが多いことが分かってきた . その理由は適切な合成方法・条件が選択できていないと纏めることができる . 出発原料に固相 , 液相 , 気相の何れを用いるのか , 反応温度・時間 , 圧力などの合成条件をどう選択するのか , 多くの経験を有する合成研究者のノウハウを駆使しなければ , 合成が困難な化合物が多いのである . 本研究ではデータ科学の手法に則り , こ

のノウハウを計算機上で構築し、新しいアプローチとして実証することを目指した。

3. 研究の方法

本課題では主に擬二元系酸化物を探索対象とし、大きく分けて 2 つの課題に対して研究を行った。一つ目は多数の合成結果を効率的に得るための並列合成手法の開発である。未知物質の合成条件を予測には、擬二元系酸化物を合成可能な条件だけではなく、合成できなかったという失敗データも重要な役割を果たす。そのため、まず錯体重合法において多数の試料を同時並列に調製し効率的に多くの合成結果を得るための装置開発を行った。次に、蓄積した約 1,000 件の合成データを様々な手法でスコリングし、合成条件パラメータを各軸としたテンソルデータに整形し、Tucker 分解法などのテンソル分解法を活用して未実験実験条件におけるスコアを予測した。交差検定法により、この予測スコアと合成成功確率に相関があることを評価した。

もう一つは、成功確率が高いと予測された実験条件での新奇化合物合成の検証であり、数百件の追加実験を行った。また、新規物質の合成成功条件の発見が偶発的でないことを検証すべく、機械学習手法を用いずランダムに選んだ条件においても合成実験を行い、合成成功数の比較を行った。

4. 研究成果

約 1,000 件の合成実験結果をテンソル型データに整形し、Tucker 分解を行った。全く未実験の化学組成においても、その合成成功確率を的確に推薦することが本研究の目的であるため、全データから各化学組成に対応する合成データを未実験とし、残りの合成データでそれらの結果を予測する交差検証を行った。その予測スコアの分布を、合成成功条件および失敗条件に対してプロットしたものを図 1(a)に示す。今回用いた学習データは全テンソル要素の 1%程度しかないため、合成成功条件も失敗条件も類似した分布を示すが、合成成功(失敗)条件の分布は高(低)スコア側に偏ることがわかった。また、これらの分布が最も分離する際の Tucker 分解のハイパーパラメータを図 1(b)に示す。例えば原料の場合、用いた原料全 28 種は 9 次元のベクトルで表現する場合に良い予測結果を示すことを示し、Tucker 分解で行っている低ランク性の仮定が有効に働く問題設定であることを示唆している。

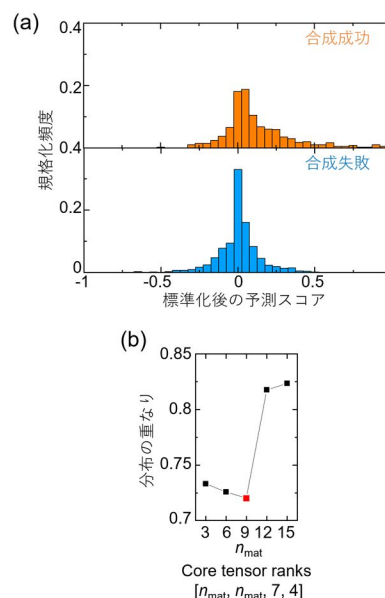


図 1 (a)交差検証の予測スコア分布。(b)Tucker 分解におけるハイパーパラメータの最適化結果。

このような予測モデルに基づいて、予測スコアの高い順に追加の合成実験を行ったところ、2件の新規物質を発見した。1つは $\text{La}_4\text{V}_2\text{O}_{11}$ であり、図2に各合成温度で焼成した後の試料のX線回折プロファイルを、各合成温度における予測スコアとともに示す。赤の矢印で示したピークが新規物質である $\text{La}_4\text{V}_2\text{O}_{11}$ の回折ピークであり、予測スコアが高い高温側ほどその回折強度が高く、また不純物相である LaVO_4 の回折ピークに対しての相対強度も強くなっていることがわかる。もう一つの新規物質である $\text{La}_7\text{Sb}_3\text{O}_{18}$ においても同様の傾向を示したことから、本手法は未発見の物質であってもその合成成功条件をピンポイントで予測可能であることを示している。

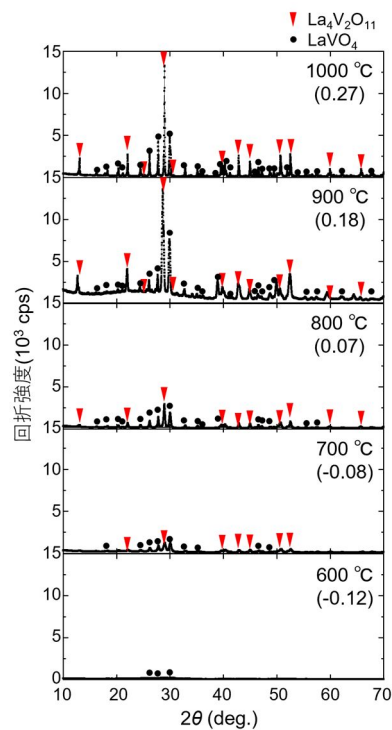


図2 新規物質 $\text{La}_4\text{V}_2\text{O}_{11}$ の合成温度ごとの合成確率の予測値と実際のX線回折プロファイル。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Takayuki Nishiyama, Atsuto Seko, and Isao Tanaka	4. 巻 4
2. 論文標題 Application of machine learning potentials to predict grain boundary properties in fcc elemental metals	5. 発行年 2020年
3. 雑誌名 Phys. Rev. Materials	6. 最初と最後の頁 123607-1-5
掲載論文のDOI（デジタルオブジェクト識別子） 10.1103/PhysRevMaterials.4.123607	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Hiroyuki Hayashi, Keita Kouzai, Yuta Morimitsu and Isao Tanaka	4. 巻 105(2)
2. 論文標題 Synthesis-condition recommender system discovers novel inorganic oxides	5. 発行年 2021年
3. 雑誌名 Journal of the American Ceramic Society	6. 最初と最後の頁 853-861
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/jace.18113	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------