

令和 6 年 6 月 7 日現在

機関番号：82108

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K05287

研究課題名（和文）マテリアルキュレーションのための物性間関係性データ作成・検索技術

研究課題名（英文）Techniques for data compile and retrieval on material property relationship for Materials Curation

研究代表者

吉武 道子（YOSHITAKE, Michiko）

国立研究開発法人物質・材料研究機構・電子・光機能材料研究センター・NIMS特別研究員

研究者番号：70343837

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：物性間に関する情報を、収集&データベース化して、検索する技術を開発した。収集源として、既に関係性に関する科学的根拠が確立したものを収録している教科書的な文献のみを対象とすること、つまり論文は含めないことが重要である。また、教科書的な文献の文言（自然言語）から、係り受けなどの文脈を利用する自然言語処理技術を用いて物性間に関する関係性のみを抽出する技術を開発した。その際に必要となる辞書も作成した。さらに、物性間の定量的関係性を抽出するために、文献中から関係性を記述した数式を抽出し、数式中に含まれる記号を、本文中に記載の物性名と紐づける技術も開発した。

研究成果の学術的意義や社会的意義

数値データのみではなく、文字で表記された情報（=データ）をコンピュータ処理して材料科学分野において利用することは、数値データの取得に時間とコストのかかる材料分野において非常に重要である。例えば、既存のデータがない添加元素が多く含まれる合金などにおいて、測定の困難な熱伝導率の値を、科学原理に基づいて、測定の容易な電気伝導率の値から推定できるなどの利点をもたらす。また、機械学習により最適化ができたとしても、その材料がよい理由がブラックボックスでは考慮外の特性に予期せぬ劣化の可能性があるため、科学原理に基づいた物性間に関する関係性データベースを用いることで、そのような考慮外の事態を防ぐことができる。

研究成果の概要（英文）：Techniques to collect and compile information on the relation between various materials properties, to make a database on such relationships, and to retrieve relationships have been developed. As a source of information, it is important to use textbook-like literature that contains only already established relationships based on scientific principles. Papers should be excluded. Techniques to extract materials property relationships from textbook-like literature (written by natural language) using natural language processing techniques to utilize contexts have been developed. A dictionary that enables such automatic extraction has also been developed. Furthermore, to extract quantitative relationship, techniques to extract mathematical equations describing relations among materials properties, and to connect the parameters in the extracted mathematical equations to the corresponding materials properties have also been developed.

研究分野：マテリアルズインフォマティクス

キーワード：物性間関係性 自然言語処理 ネットワーク型データベース 科学原理

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

## 様式 C-19、F-19-1 (共通)

### 1. 研究開始当初の背景

マテリアルズインフォマティクスが勃興し、多くの数値データを統計的に解析して、予測したり分類したりする技術が盛んに研究されるようになった。ただ、材料科学の分野においては、数値データが少なく、データベースもデータの欠損が多い。また、数値データの機械学習により得られる結論が、定性的には既に知られていることが多く、文字データの活用も必要と感じられた。

### 2. 研究の目的

ある程度材料の種類を問わずに成立する、科学原理に立脚した物性同士の関係性、例えばバンドギャップと価数で規格化した酸化物の生成エンタルピーの間には非常に良い線形関係がある、というような分子軌道法的な化学結合論から導かれる知識を、物性間（この場合、バンドギャップと生成エンタルピー）の関係性としてデータベース化して、関係性を検索できる技術を開発して、数値データのみではなく、科学原理の知識も利用したマテリアルズインフォマティクスを確立する。

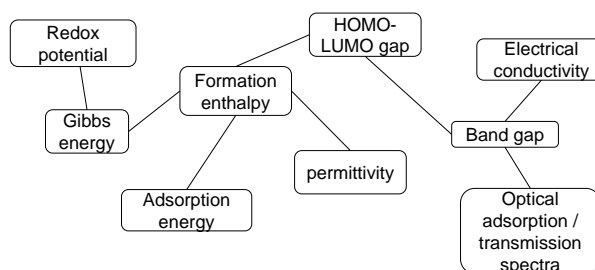
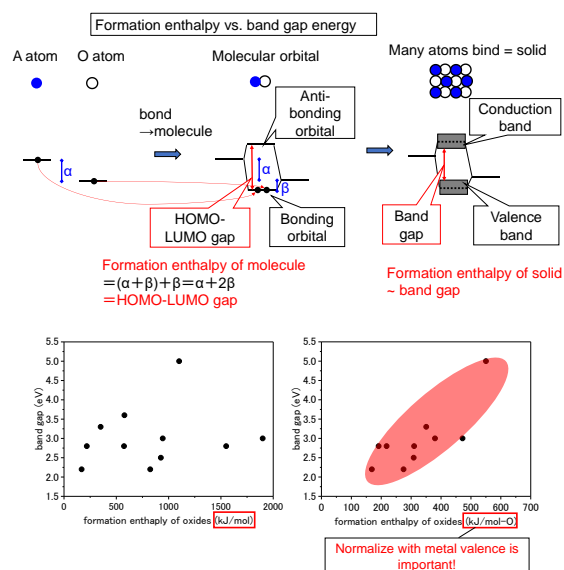
### 3. 研究の方法

自然言語処理というコンピュータで自動的に文章を解析する技術を用い、科学原理の記述された材料科学の教科書的な文献から、物性間の関係性を自動抽出する。そのために必要な各種の技術や、解析時に用いる当該目的に適した用語辞書を開発した。また、定量的な関係性を抽出するために、教科書的な文献から物性間の定量的関係性を記述した数式を抜き出し、数式中に用いられている物性を表す記号（例えばバンドギャップは  $E_g$ 、電気伝導率は  $\sigma$  など）と、その記号とその記号が表す物性名（ $E_g$ -bandgap、 $\sigma$ -electrical conductivity など、文献中の数式の前後の文章中に出現）とを、コンピュータによって自動的に結びつける技術を開発した。物性間の関係性を数学の「グラフ」というネットワーク型のデータ形式で格納し、グラフ理論に基づく関係性の探索ができる技術を開発した。

### 4. 研究成果

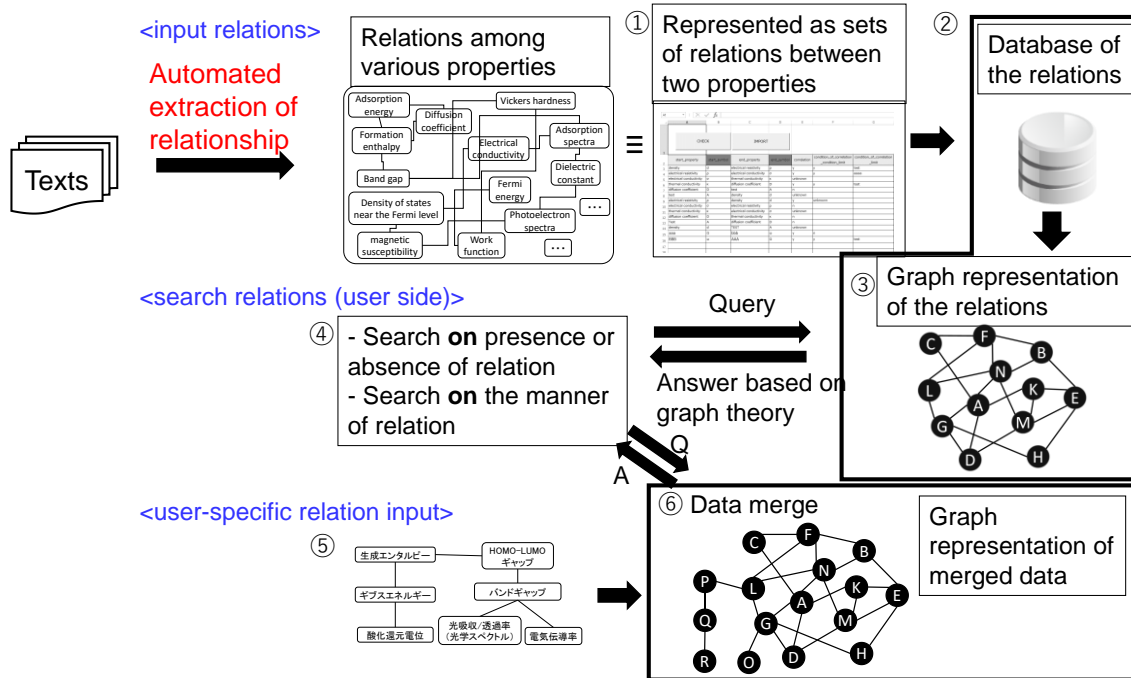
右に示したのは、バンドギャップと生成エンタルピーとが関係する原理を示した図である。この図からわかるように、基本的に材料の種類を問わず、原理的に関係性が成立する。実際に、金属酸化物について、バンドギャップと生成エンタルピーの値をプロットしてみたのが右下図である。生成エンタルピーの値として、普通にデータベースに格納されている酸化物 1 モル当たりのエンタルピーでプロットすると（図の左側）関係がなさそうに見えるが、生成エンタルピーの値を酸化物の価数で規格化してプロットすると（図の右側）、かなり良い線形の関係が存在していることがわかる。すなわち、先のバンドギャップと生成エンタルピーとが関係する原理が成立していることがわかる。

このようにして、様々な物性間の関係性をつないでいくと、右図のような、物性同士が繋がっているネットワーク構造ができる。このようなネットワーク=つながりの情報をデータベース化した。注意すべきは、この情報は、関係性の情報であり、それぞれの物性の各材料における数値データは入っていないことである。例えば、鉄道網の情報で、「乗換案内」のようなもので、どの駅からどの駅へ行くのに、どのような鉄道に乗ってどの駅で乗り換えればよいかわかるようなものである。物性間の関係性の場合、ある物性と別の物性が、その経路上にある様々な物性を介して影響を受けていることがわかる。この経路をたどることで、ある材料について数値データが存在しない場合でも、例えば生成



エンタルピーの値からバンドギャップの値を推測するということが可能になる。

このような物性間の関係性を教科書的な文献から開発した自然言語処理技術を駆使して自動抽出してデータベース化し、その検索が可能なシステム（プロトタイプ）を構築した。その概要を下図に示す。<input relations>が関係性の抽出・データベース化部分、<search relations>が関係性の探索部分である。このデータベースが数学のグラフ理論に基づく構造をしていることで、<user-specific relation input>に書かれているように、一般的な物性間の関係性ではなく、ユーザーに特化した狭い範囲の物性間の関係性を追加して入力し、そのユーザーのみ追加されたデータベース内を探索するということが簡単にできる構造になっている。

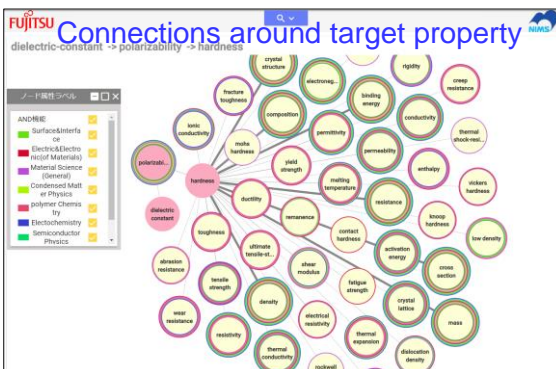
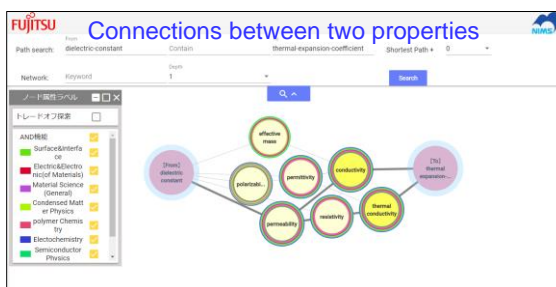


データベースの探索は、基本的に二つの異なる物性間にどのような関係が存在しているかを調べるモードA（右上図）と、特定の物性が他のどのような物性と関係しているかを調べるモードB（右下図）がある。

モードAは、ある物性を向上させようと材料を変化させたときに思いもかけず別の物性が悪化した場合などに、ある物性と悪化した別の物性との間に関係があるかを調べて、悪化の原因を探ることを可能にする。現状ではプロトタイプに実装されていないが、2つの物性がトレードオフの関係にある場合に、トレードオフを回避する方法を示すアルゴリズムも開発済みである。

モードBは、ある物性を向上させようとしたときに、思いもかけない物性とのつながりを見つけて、従来の延長線上にない材料候補の示唆に役立つ。また、ある物性を向上させようと材料を変化させたときに、関係している物性をあらかじめ知っておくことで、思いもかけない別の特性の悪化を防ぐことが可能である。

先に例を挙げた、バンドギャップと生成エンタルピーの関係について、このシステムを用いて関係性探索を行った結果を下図に示した。

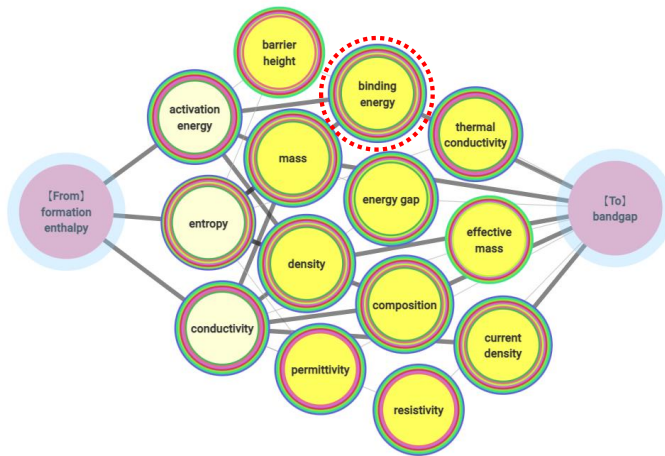


ノード属性ラベル

トレードオフ探索

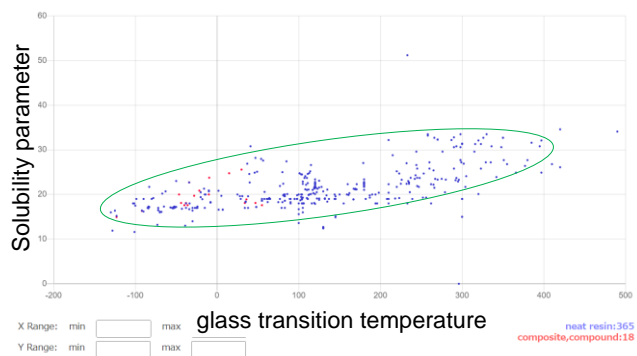
AND機能

- Surface&Interface
- Electric&Electronic(of Materials)
- Material Science(General)
- Condensed Matter Physics
- Polymer Chemistry
- Electrochemistry
- Semiconductor



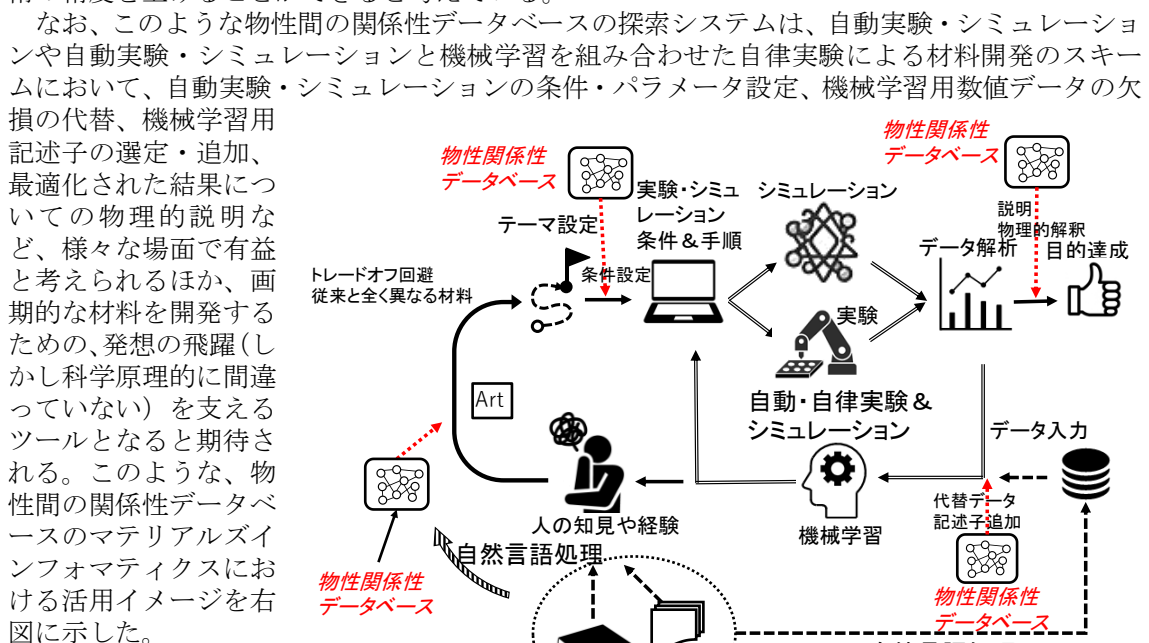
原理のところでも示したように、formation enthalpy と bandgap の間に binding energy が表示され、結合エネルギーが formation enthalpy と bandgap の起源になっていて両者に関係が発生していることがわかる。

先の例では、無機化合物を例に出したが、有機化合物・高分子化合物においても基本的な科学原理は同じで、本システムで関係性が見つかった、solubility parameter と glass transition temperature について、高分子のデータベース PolyInfo[1]の数値データを用いて関係性を検証してみたのが右図である。このように、きれいな線形の関係性があることが見て取れる。



定量的関係性については、今後、生成

AI を利用することで、数式中に用いられている物性を表す記号と、その記号とその記号が表す物性名（文献中の数式の前後の文章中出现）とをコンピュータによって自動的に結びつける技術の精度を上げることができると考えている。





5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件／うち国際共著 0件／うちオープンアクセス 5件）

1. 著者名 吉武道子	4. 巻 30
2. 論文標題 科学原理を分野横断的に俯瞰・活用する～マテリアルキュレーション	5. 発行年 2023年
3. 雑誌名 Journal of the Society of Inorganic Materials, Japan	6. 最初と最後の頁 291-296
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 吉武道子	4. 巻 72
2. 論文標題 データキュレーションの現状と展望：材料科学分野を例に	5. 発行年 2022年
3. 雑誌名 情報の科学と技術	6. 最初と最後の頁 164-169
掲載論文のDOI（デジタルオブジェクト識別子） 10.18919/jkg.72.5_164	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 吉武道子	4. 巻 27
2. 論文標題 マテリアルキュレーション支援システムの教育・研修用展開および計測法とのリンク拡張	5. 発行年 2021年
3. 雑誌名 Journal of Surface Analysis	6. 最初と最後の頁 95～103
掲載論文のDOI（デジタルオブジェクト識別子） 10.1384/jsa.27.95	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Yoshitake Michiko	4. 巻 14
2. 論文標題 Tool for Designing Breakthrough Discovery in Materials Science	5. 発行年 2021年
3. 雑誌名 Materials	6. 最初と最後の頁 6946～6946
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/ma14226946	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 YOSHITAKE Michiko	4. 巻 72
2. 論文標題 Materials Curation: Transdisciplinary Utilization of Scientific Principles	5. 発行年 2021年
3. 雑誌名 Journal of The Surface Finishing Society of Japan	6. 最初と最後の頁 79 ~ 83
掲載論文のDOI (デジタルオブジェクト識別子) 10.4139/sfj.72.79	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 吉武道子	4. 巻 19
2. 論文標題 マテリアルキュレーションR-科学法則を俯瞰的に利用した材料探索法-	5. 発行年 2020年
3. 雑誌名 J. Comput. Chem. Jpn	6. 最初と最後の頁 36, 42
掲載論文のDOI (デジタルオブジェクト識別子) 10.2477/jccj.2020-0004	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計6件 (うち招待講演 1件 / うち国際学会 1件)

1. 発表者名 吉武 道子, 長田 貴弘
2. 発表標題 マテリアルキュレーション支援システム: 事例研究
3. 学会等名 応用物理学会秋季講演大会
4. 発表年 2023年

1. 発表者名 吉武 道子, 河野 敬
2. 発表標題 文書中の数式の変数及び変数の意味を抽出する試み
3. 学会等名 情報知識学会
4. 発表年 2022年

1. 発表者名 吉武 道子、河野 敬
2. 発表標題 専門書の数式とテキストから材料物性間の定量的関係性を抽出する
3. 学会等名 人工知能学会
4. 発表年 2022年

1. 発表者名 吉武 道子
2. 発表標題 マテリアルクキュレーションRとその支援システム：科学原理を分野横断的・俯瞰的に探索
3. 学会等名 PSA-20
4. 発表年 2020年

1. 発表者名 Michiko Yoshitake
2. 発表標題 Natural Language Processing for Materials Science
3. 学会等名 VITC-2020（招待講演）（国際学会）
4. 発表年 2020年

1. 発表者名 吉武 道子
2. 発表標題 文書ファイルからの数式抽出と式変形の試み
3. 学会等名 応用物理学会
4. 発表年 2020年

〔図書〕 計1件

1. 著者名 吉武道子（分担執筆）、他57名	4. 発行年 2021年
2. 出版社 技術情報協会	5. 総ページ数 500
3. 書名 マテリアルズインフォマティクスのためのデータ作成とその解析、応用事例	

〔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
---------	---------