

科学研究費助成事業 研究成果報告書

令和 5 年 6 月 8 日現在

機関番号：14401

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19805

研究課題名（和文）グラフと点群を特徴量とした深層学習による結晶性質予測の高精度化

研究課題名（英文）Improving Accuracy of Crystal Property Prediction via Neural Network for Graphs and Points

研究代表者

佐々木 勇和（Sasaki, Yuya）

大阪大学・大学院情報科学研究科・助教

研究者番号：40745147

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）：本研究は、高精度な結晶の物性値予測の達成のために、(1)マテリアルインフォマティクスにおける深層学習技術のサーベイ、(2)データセットの構築、(3)点群とグラフを用いた結晶予測深層学習技術の開発、および(4)深層学習技術に基づく物質生成の4つの項目を実施した。データセットおよび物質生成では、太陽光電池材料に着目し、既発表文献から分子構造と物性値を抽出し、開発した深層学習技術に基づいて新たな物質生成に取り組んだ。

研究成果の学術的意義や社会的意義

情報学と材料工学を融合させたマテリアルインフォマティクスの研究が盛んに行われており、各国において様々な取組みが行われている。マテリアルインフォマティクスの目的のひとつは、新材料の開発の効率化である。従来の材料工学では、開発者や研究者の直感に頼って、新材料候補に対する実験を試行錯誤しながら実施していたが、マテリアルインフォマティクスは蓄積された過去の材料実験データやシミュレーションデータを活用し、実験候補の絞り込みを行い、実験回数を減らすことができる。これにより、新たな有用な物質探索の発見を効率的に実施することができるため、新たな産業の発展が期待できる。

研究成果の概要（英文）：This project conducted four main tasks in order to achieve accurate prediction of the physical properties of crystals: (1) a survey of deep learning techniques in materials informatics, (2) construction of a dataset, (3) development of deep learning techniques for crystal prediction using point clouds and graphs, and (4) material generation based on deep learning techniques. For the dataset and material generation, the focus was on solar cell materials. Molecular structures and physical properties were extracted from existing literature, and novel material generation was attempted using the developed deep learning techniques.

研究分野：データベース

キーワード：マテリアルズインフォマティクス 深層学習 グラフデータ 点群データ 太陽光電池

1. 研究開始当初の背景

情報学と材料工学を融合させたマテリアルインフォマティクスの研究が盛んに行われており、各国において様々な取組みが行われている。例えば、米国では、2011年に発表した「Materials Genome Initiative」においてマテリアルインフォマティクスの推進を打ち出しており、その予算総額は2億ドルに上る。欧州では「Novel Material Discovery Laboratory」が2015年に、中国では「北京マテリアルズ・ゲノム・エンジニアリング・イノベーション連盟」が2016年に設立されている。日本においても、2015年には物質・材料研究機構に材料統合型物質・材料研究拠点が創設され、多くの研究成果が生み出されている。このように各国がマテリアルインフォマティクスに関する研究強化を図っており、今後益々マテリアルインフォマティクスの重要性が増すと予想される。

マテリアルインフォマティクスの目的のひとつは、新材料の開発の効率化である。従来の材料工学では、開発者や研究者の直感に頼って、新材料候補に対する実験を試行錯誤しながら実施していたが、マテリアルインフォマティクスは蓄積された過去の材料実験データやシミュレーションデータを活用し、実験候補の絞り込みを行い、実験回数を減らすことができる。新材料の探索では、合成元の候補の列挙、合成物および合成成否の予測、そして合成物の性質予測の3つの研究課題が大きくある。それぞれの課題を高精度に予測することが重要であり、半導体開発やイオン生成など様々な分野でそれぞれの物質に合わせた高精度化が行われている。特に、化学分野はオープンデータの量や合成物質の探索範囲の広さといった観点から研究が進んでいる。一方、本研究が対象とする結晶におけるマテリアルインフォマティクスはまだまだ発展途上であり、高精度な予測モデルが求められている。

2. 研究の目的

本研究はマテリアルインフォマティクスの研究課題のうち、結晶の性質予測に取り組む。結晶の性質予測は、形成エネルギーやバンド幅などの物性値を実験により確認することができるが、実験のコストが高く、予測により実験の量を減らすことが可能である。また、合成元の物質から生成が予測される結晶に対して、物性値の予測を行うことで、結晶の生成そのものを行うかの判断もすることができる。本研究の目的は、結晶の性質の予測において最高性能を達成することである。

3. 研究の方法

本研究は、高精度な結晶の物性値予測の達成のために、(1)マテリアルインフォマティクスにおける深層学習技術のサーベイ、(2)データセットの構築、(3)点群とグラフを用いた結晶予測深層学習技術の開発、および(4)深層学習技術に基づく物質生成の4つの項目を実施した。

(1) マテリアルインフォマティクスにおける深層学習技術のサーベイ論文

マテリアルインフォマティクスにおける深層学習技術の研究はまだ広く行われていない。まず、技術開発の前段階として、現在の性質予測における深層学習技術を化学や応用物理学などの分野に関わらず、どのような技術が開発され、どの程度の精度が達成されているかをまとめ、サーベイ論文を執筆する。これにより、本研究のプレゼンスを明らかにする。

(2) データセット構築

応用に特化してモデルを構築する場合、対象とする物質のデータセットを構築する必要がある。本研究では、文献情報から物質情報を抽出し、応用に特化したデータセットを構築する。

(3) 点群とグラフを活用した深層学習技術の開発

点群およびグラフを対象とした深層学習技術の開発は盛んに行われている。さらなる結晶予測における高精度を達成するために、点群とグラフ両方の性質を入力として予測を可能とする深層学習技術を開発する。ハイブリッドに用いる場合は二つの課題がある。まず一つ目は、一つの結晶からどのように入力となる点群とグラフを構築するかである。次に、点群とグラフ両方を入力とするどのような深層学習モデルを構築するかも重要である。例えば、それぞれの学習結果を深層学習の最後の層で連結して予測することや、最初から連結して学習するといった方法も考えられる。そのため、点群とグラフからのモデルは多種多様なモデル構築が可能である。アテンションや敵対的生成ネットワークなどの技術を取り入れることも必要である。本研究では、様々な入力方法とモデル構築を試し、最高性能の予測が可能な深層学習技術を開発する。

(4) 深層学習に基づく物質生成

上記の研究項目(2)と(3)の成果を活用し、物性値を予測し物質生成に取り組む。

4. 研究成果

それぞれの項目に対する研究成果をまとめる。

(1) マテリアルインフォマティクスにおける深層学習技術のサーベイ

本研究において、新物質探索における深層学習の調査として化学物質と結晶における深層学習技術の違いおよび識別と生成の技術の発展についてまとめた。さらに、結晶探索に特化させ既存技術の調査と今後の方向性を議論した。これからの成果はサーベイ論文として出版されている。さらに、この調査に基づき書籍におけるチャプターの執筆も行った。

(2) データセットの構築

データセットの構築では、共同研究者の大阪大学 陣内青萌助教の協力のもと、ヘテロ接合型有機太陽電池に着目した。物質生成の可否と実用化の観点から P2HT に注目し、既発表文献 500 件程度の論文からし、600 件超の分子構造とヘテロ接合系の機能を抽出した。このデータセットは既存の公開データベースでは未だ存在しないデータであり、今後の太陽光電池開発において有用なものである。

(3) 点群とグラフを用いた結晶予測深層学習技術の開発

原子間のねじれを枝の属性として保持するモデルを構築した。位置情報を節点の属性として保持する既存のモデルと比較して、位置情報の原点を決める必要がなく、相対的な原子間の情報のみでよいという特徴がある。また既存モデルと同等の情報、例えば、物質間の距離、を枝に付与することができるため、既存モデルの機能を包含するような一般化モデルとなっている。本モデルを公開されているデータセットおよび項目(3)で作成したデータに対して性能評価を行い、既存手法よりも高精度であることを確認した。

(4) 深層学習技術に基づく物質生成

研究項目(3)に基づいて太陽電池開発を行っている。現在も物質開発実験を実施中であり、今後も継続して行う。

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 佐々木 勇和	4. 巻 49
2. 論文標題 深層学習技術による結晶探索の現状と今後	5. 発行年 2022年
3. 雑誌名 日本結晶成長学会誌	6. 最初と最後の頁 1 - 7
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 奥野智也, 佐々木 勇和, 鈴木 雄太	4. 巻 13
2. 論文標題 深層学習を用いた新物質探索に関するサーベイ	5. 発行年 2020年
3. 雑誌名 情報処理学会論文誌データベース (TOD)	6. 最初と最後の頁 22 - 31
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 -

〔学会発表〕 計2件（うち招待講演 1件 / うち国際学会 0件）

1. 発表者名 佐々木 勇和, 奥野智也, 鈴木雄太
2. 発表標題 深層学習と物質探索
3. 学会等名 第1回インフォマティクス応用研究グループ 研究会（招待講演）
4. 発表年 2020年

1. 発表者名 奥野 智也, 佐々木 勇和, 原田 圭, 吉村 政志
2. 発表標題 結晶の物性値予測における点群深層学習の応用
3. 学会等名 第12回データ工学と情報マネジメントに関するフォーラム (DEIM 2020)
4. 発表年 2020年

〔図書〕 計1件

1. 著者名 大野かおる等, 58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
研究協力者	陣内 青萌 (Jinnai Seihou)	大阪大学・産業科学研究所・助教	

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
---------	---------