

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 1 日現在

機関番号：12601

研究種目：新学術領域研究(研究領域提案型)

研究期間：2016～2017

課題番号：16H00736

研究課題名(和文)材料インフォマティクスに適した機械学習法の開拓

研究課題名(英文)Development of machine learning methods for materials informatics

研究代表者

津田 宏治(Tsuda, Koji)

東京大学・大学院新領域創成科学研究科・教授

研究者番号：90357517

交付決定額(研究期間全体):(直接経費) 9,300,000円

研究成果の概要(和文):本計画研究は、新学術領域「ナノ構造情報のフロンティア開拓 - 材料科学の新展開」において、情報科学の研究を加速させるため、2016年度に新しく追加された。他班との協力を進めた結果、以下のような成果を得ることができた。(1)ナノ構造の網羅的な分光計測データ解析のための統計的機械学習法。(2)モンテカルロ木探索を用いた材料自動設計アルゴリズム。(3)推薦アルゴリズムを用いた安定な化合物の予測。

研究成果の概要(英文):This research started from 2016 to boost the development of informatics methods in materials development. As a result of collaborations with other groups, we obtained the following achievements. (1) Statistical machine learning methods for spectrum analysis to discovering nanostructures. (2) Automatic design of materials using Monte Carlo tree search. (3) Discovery of stable compounds with recommendation algorithms.

研究分野：機械学習、バイオインフォマティクス

キーワード：マテリアルズインフォマティクス

1. 研究開始当初の背景

新学術領域「ナノ構造情報のフロンティア開拓—材料科学の新展開」は、材料科学と情報科学の融合を目指している。2015年度まで、本研究の代表者である津田は、計画研究の分担者として参加してきたが、より情報学の研究に注力するという目的で、津田を代表とする新たな計画研究を2016年度から開始することに決定した。

2. 研究の目的

新設された本計画研究では、ナノ構造情報の活用を目指した材料科学と情報学の融合研究を推進する。『材料探索』、『構造探索』、『物理法則探索』の3つを材料科学と情報学との融合新領域におけるターゲットとし、データの解析に必要な機械学習アルゴリズムの開発を進める。具体的には、1)原子の空間配置を捉える記述子の開発、2)データに基づく記述子の導出、3)高性能な実験計画法の開発という三課題を掲げる。そしてコモンサブジェクト(CS)課題に該当する材料テーマに貢献できるように研究を進める。さらに他班との連携研究において、新たな手法・アルゴリズム開発が必要となった場合には、迅速かつ柔軟に対応する。このように本研究は領域内の学際的研究を加速させ、材料研究にインフォーマティクス手法を深く根付かせることを目指す。また、そのための若手人材育成にも貢献する。

3. 研究の方法

他班とのコミュニケーションを通じて、研究目標の共有を行ったあと、本班にて機械学習アルゴリズムの設計を行い、他班から送付されるデータに適用する。また、他班が保持している第一原理計算の技術を利用して、機械学習に必要なデータを作り出す。機械学習で得た結果を考察して、論文として発表する。

4. 研究成果

(1)計画研究班 A01(ウ)(代表者:武藤俊介教授・名古屋大学)との連携によって、ナノ構造の網羅的な分光計測データ解析のための統計的機械学習法を開発している。本年度は、計測点の局所構造を反映するスペクトルの2次元計測(スペクトル・イメージ)データから、成分毎のスペクトルおよび空間分布を同定する汎用的な手法を開発した。開発手法によるデータ解析の自動化により微細構造の評価コストが下がり、また、客観的かつ定量的な解析の実現が期待される。この開発法に関して、STEM-EELSやSTEM-EDXなどの様々な実計測データを用いて手法の性能を検証し Ultramicroscopy にて発表を行った。

(2)金属触媒の表面では反応物の吸着や表面

反応が起きており、触媒作用はその電子状態により規定される。計画研究班 A03(ケ)(代表者:高草木達准教授・北海道大学)との連携により、金属触媒の特性を系統的に表す量の一つである d-band center の機械学習による正確な予測に成功した。本研究では、金属や二元合金の電気陰性度やイオン化陰性度を記述子として用い、Gradient boosting regression などの機械学習手法を用いることによって、第一原理計算を用いずに、d-band center の高精度の予測が可能であることを示した。複雑な表面のモデリングを伴う第一原理計算には通常大きな計算コストがかかるため、機械学習による簡便で高速な代替により、網羅的な触媒探索の高効率化や試行錯誤過程の合理化が期待できる。

(3)計画班 01(ウ)との連携により、走査透過型電子顕微鏡のエネルギー分光スペクトル解析におけるデータ解析法の開発、特に、バックグラウンドの取り扱いや成分数の統計的推論に関して取り組み、国際ワークショップ EDGE 2017 および複数の国内学会で発表した。また、3D ラマン分光スペクトル解析の研究にも新規に取り組み、機械学習によるスペクトル分解法を開発し、学術雑誌 Journal of Physical Chemistry C に発表した。

(4)計画班 A02(エ)および A03(キ)との連携により、プロトンイオン伝導経路探索を機械学習に基づき高速計算する手法に組み込み、学術雑誌 Physical Review B で発表した。また、計画班 A02(エ)との連携により、点欠陥の周辺の原子構造配置に関する記述子を用いたエネルギー予測により点欠陥のある結晶構造の構造緩和計算を高速化する手法の開発に取り組んだ。また、公募班の山崎教授らとの連携により、プロトン輸送に関する物性値を機械学習に基づき予測する手法の研究に取り組んだ。

(5)化合物はいくつかの組成の組み合わせによって表現されるが、存在が確認されているものの情報をもとに、未知の化合物を予測する問題は、推薦システムと同様の穴埋め問題として捉えることができる。そこで、推薦システムで用いられる行列分解・テンソル分解といった手法を用いることで、未知の化合物を予測することが可能であることを示した。その研究成果は Physical Review Materials 誌上で発表された。

(6)A01(イ)班との共同研究において、転移学習を用いた粒界構造決定のアルゴリズムを開発した。これまで、ある界面、例えば、 $\Sigma 3[110](111)$ に関する粒界エネルギーの計算結果が多数揃っていても、それを、異なる界面、例えば、 $\Sigma 3[110](112)$ の粒界構造決定に用いることはできなかった。本研究では、

様々な粒界に関して共通に定義できる記述子を考案し、それらが表す特徴空間上で転移学習を行う事によって、ベイズ最適化による構造決定を大幅に加速できることを示した。本成果は、Journal of the Physical Society of Japan に出版された。

(7) 望みの物性を持つ有機化合物を自動的に発見することは、これまで難しいこととされてきたが、近年、Variational Autoencoderなどの深層学習を用いた手法が続々と提案されており、実現可能性が高まっている。本研究では、Recurrent neural network とモンテカルロ木探索を用いた手法 ChemTS を開発し、github 上で一般に公開した。ChemTS は、従来法に比べ、単位時間により多くの分子を生成することができ、高速に分子の最適化を行うことができる(図1)。本成果は、Science and Technology of Advanced Materials に出版された。

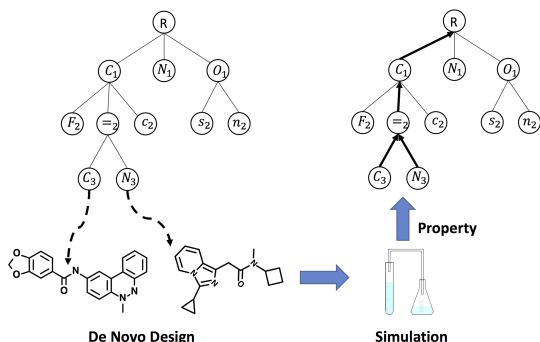


図1 : ChemTS による新規化合物の設計

(8) ベイズ最適化では、候補点の数に比例した計算量がかかるため、大規模な問題には適用できなかった。そこで、コンピュータ囲碁で用いられる最適化アルゴリズムに基づくpython ライブラリ MDTs を開発し公開した。MDTs を Si-Ge 合金設計問題に適用し、ベイズ最適化が適用できない大規模な問題で良好な結果を得た。本成果は、Science and Technology of Advanced Materials に出版された。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 9 件) 全て査読有り

[1] H. Wang, H. Zhang, B. Da, M. Shiga, H. Kitazawa, D. Fujita, Informatics-Aided Raman Microscopy for Nanometric 3D Stress Characterization, Journal of Physical Chemistry C, 122, 7187-7193, 2018.
DOI: 10.1021/acs.jpcc.7b12415

[2] K. Kanamori, K. Toyoura, J. Honda, A. Seko, M. Karasuyama, K. Shitara, M. Shiga,

A. Kuwabara, I. Takeuchi, Exploring a potential energy surface by machine learning for characterizing atomic transport, Physical Review B, 97, 125124, 2018.

DOI: 10.1103/PhysRevB.97.152124

[3] A. Seko, H. Hayashi, H. Kashima, I. Tanaka, Matrix- and Tensor-based Recommender Systems for the Discovery of Currently Unknown Inorganic Compounds, Physical Review Materials, 2, 013805, 2018.

DOI: 10.1103/PhysRevMaterials.2.013805

[4] T.M. Dieb, S. Ju, K. Yoshizoe, Z. Hou, J. Shiomi, K. Tsuda, MDTs: automatic complex materials design using Monte Carlo tree search, Science and Technology of Advanced Materials, 18, 498-503, 2017.

DOI: 10.1080/14686996.2017.1344083

[5] H. Oda, S. Kiyohara, K. Tsuda and T. Mizoguchi, Transfer Learning to Accelerate Interface Structure Searches, Journal of the Physical Society of Japan, 86, 123601, 2017.

[6] X. Yang, J. Zhang, K. Yoshizoe, K. Terayama and K. Tsuda, ChemTS: an efficient python library for de novo molecular generation, Science and Technology of Advanced Materials, 18, 972-976, 2017.

[7] M. Shiga, K. Tatsumi, S. Muto, K. Tsuda, Y. Yamamoto, T. Mori and T. Tanji, Sparse Modeling of EELS and EDX Spectral Imaging Data by Nonnegative Matrix Factorization, Ultramicroscopy, 170, 43-59, 2016.

DOI: 10.1016/j.ultramic.2016.08.006

[8] T. Ueno, T.D. Rhone, Z. Hou, T. Mizoguchi and K. Tsuda, COMBO: An Efficient Bayesian Optimization Library for Materials Science, Materials Discovery, 4, 18-21, 2016.

DOI: 10.1016/j.md.2016.04.001

[9] I. Takigawa, K. Shimizu, K. Tsuda and S. Takakusagi, Machine-learning prediction of d-band center for metals and bimetals, RSC Advances, 6, 52587-52595, 2016.

DOI: 10.1039/C6RA04345C

[学会発表] (計 1 件)

[1] M. Shiga, S. Muto, Automatic Spectral Imaging Analysis Based on Machine Learning,

The 8th International Workshop on Electron Energy Loss Spectroscopy and Related Techniques (EDGE 2017: Enhanced Data Generated by Electrons), 2017.

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

○取得状況（計 0 件）

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

津田 宏治 (TSUDA, Koji)
東京大学・大学院新領域創成科学研究科
教授
研究者番号：90357517

(2) 研究分担者

鹿島 久嗣 (KASHIMA, Hisashi)
京都大学・情報学研究科・教授
研究者番号：80545583

志賀 元紀 (SHIGA, Motoki)
岐阜大学・工学部・准教授
研究者番号：20437263

(3) 連携研究者

竹内 一郎 (TAKEUCHI, Ichiro)
名古屋工業大学・工学（系）研究科・教授
研究者番号：40335146

(4) 研究協力者

()