様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ５ 月 ２２ 日現在

機関番号：６２６０３
研究種目：若手研究
研究期間：2018～2021
課題番号：１８Ｋ１８０１７

研究課題名（和文）Inverse materials design by integrating transfer learning techniques into a Bayesian framework

研究課題名（英文）Inverse materials design by integrating transfer learning techniques into a Bayesian framework

研究代表者
Ｗｕ　Ｓｔｅｐｈｅｎ（Wu, Stephen）

統計数理研究所・データ科学研究系・准教授

研究者番号：７０８０４１８６

交付決定額（研究期間全体）：（直接経費）　　3,100,000円

研究成果の概要（和文）：ベイズ推定と転移学習という2つの重要な機械学習技術に基づき、私はPythonでXenonPyというオールインワン材料情報学プラットフォームを開発し、目的の特性を持つ材料構造の候補を作り出すことができるようになった。一般に公開されているXenonPyは、データサイエンスの活用し、新しい機能性材料の発見を加速するための基盤になる。このプラットフォームを使って、3つの新しい高熱伝導性ポリマーを発見し、合成することに成功した。最も優れた設計では、従来の市販ポリマーよりも80%高い熱伝導率を達成した。さらに、新しい液晶ポリイミドや高格子熱伝導性結晶など、産業的価値の高そうなものをいくつか発見した。

研究成果の学術的意義や社会的意義
材料科学と機械学習の融合を図るマテリアルズインフォマティクス呼ばれる学際領域に期待が高まっているが、産業的に重要かつ新規な機能性材料の合成に成功した研究はまだ少ない。実験コストが高いためにデータが不足していることが、マテリアルインフォマティクスの実際の産業価値を実現するための重要なボトルネックになっている。同技術を用いることで，データが豊富な重要度の低い材料特性から、データが少ない重要度の高い材料特性にも関連する有用な情報を抽出することができる。今までデータ不足のためマテリアルインフォマティクスの応用が失敗した材料科学問題を突破する鍵になる研究である。

研究成果の概要（英文）：Based on two key machine learning technologies, Bayesian inference and transfer learning, I have developed an all-in-one materials informatics platform in Python, called XenonPy, that can produce candidates of materials structures with desired properties. The openly available platform serves as the foundation to promote using data science for accelerating discovery of new functional materials suitable for various industrial needs in our daily life. Using this platform, three new high thermal conductivity polymers were discovered and successfully synthesized. The best performing design reached 80% higher thermal conductivities than conventional commercial polymers. Furthermore, several new liquid-crystal polyimide and high lattice thermal conductivity crystals were discovered that seem to have high industrial values.

研究分野：統計

キーワード：polymer informatics　transfer learning

1. 研究開始当初の背景 (Research background)

Materials informatics is an interdisciplinary field of materials science and machine learning that aims at accelerating materials discovery through rational strategies for acquisition, management and analysis of material data. This is an emerging field with great opportunities for researchers to challenge with various new methodologies in data science, especially after the Materials Genome Initiatives launched by the US federal government back in 2011. However, despite the raising expectation on materials informatics, successful studies of actually synthesizing truly novel functional materials have been rarely reported in this area. While a small set of applications of machine learning has been established on materials science problems with abundant data, we speculated that the lack of data for industrially important properties due to high experimental cost is the key bottleneck to realize the actual industrial value of materials informatics. For example, only a few tens of thermal conductivity (an important thermal management related property for many industrial applications) data points could be extracted from the largest polymer database, PoLyInfo, in 2018[1], which is hardly suitable for applying existing methods in materials informatics. Therefore, I proposed to use the transfer learning technique in machine learning to extract useful information from less important materials properties with abundant data that is also relevant to more important materials properties with little data. Such statistical methodology is expected to bridge the gap between academic studies of materials informatics and the actual needs from materials industry.

2. 研究の目的 (Research goal)

The objective of this study is to identify novel functional materials that achieve any given requirements on physical, electronic and mechanical properties driven by actual industrial needs. The problem is a difficult multiobjective optimization with candidate materials spanning an infinitely high-dimensional search space. Some initial success to address this issue within the framework of Bayesian machine learning has been reported for some specific materials applications with abundant data[2]. In order to extend the method to real industrial problems that often have little data for training machine learning models, a transfer learning approach could be combined with Bayesian machine learning to extract relevant knowledge to important industrial problems with little data from other materials problems with large data. I aimed to develop open source software for the use of general computational materials design with transfer learning and Bayesian machine learning, which will serve as a foundation for promoting materials informatics in important industrial applications.

3. 研究の方法 (Research method)

By combining the Bayesian inverse design algorithm with machine learning models built from transfer learning methods, I developed a complete workflow of autonomous material design targeting specific properties using only a small amount of available data. The full design algorithm has been implemented in an open source Python package, called XenonPy, that serves as an all-in-one materials informatics

platform. The impact of my proposed method on real industrial applications has been demonstrated through design of new polymers for thermal management purposes. I briefly explain the details of the research workflow separated in three parts:

1. Bayesian inverse material design with transfer learning based surrogate model

From experimental or computational data that shows examples of different material structures having different material properties, one can learn a mathematical model that maps any structures to specific property values. Modern machine learning technology is efficient in learning such model from a large amount of data, but not with a small dataset. Similar to how human can learn how to play guitar faster after experienced playing other musical instruments, transfer learning is a technique to extract relevant knowledge from a task with large data that is applicable to another new task with little. As a result, the model learning for the new task could be speeded up.

After having a model that predicts materials properties of a structure using only little data, one can apply a Bayesian inverse design algorithm to search for structures with desired properties. As an example in this study, I began with some known molecular structure represented in SMILES, which is a string representation of molecules. Using existing natural language processing models, I could modified the SMILES, thus, making slight changes to the original molecular structure. The properties of the modified structure is then predicted by the learned model and given a score based on how close the predicted properties are to the desired ones. By repeatedly modifying and scoring the structures, one can eventually obtain candidates of desired molecule design.

2. Implementation of design algorithm in a fully open Python package, XenonPy

The above algorithm involves integration of multiple machine learning methods. In order to build a unified platform that works for different material design problems, the design workflow is broken down into multiple modules, where each can be further customized depending on specific applications. All modules are standardized under the open source Python package, called XenonPy. First, material structures are converted into numerical information used as input of the learned predictive models through a descriptor calculator that can handle multiple classes of materials. Second, a scoring function is used to evaluate how close a structure is to the desired design based on predicted property values. Third, a generator is available for modifying different types of material structures. Finally, an integrating module that wraps the previous three modules into a continuous design loop, called iQSPR-X, which can produce design candidates autonomously.

3. Validation of proposed method on designing polymers for thermal management applications

The XenonPy platform is being deployed to multiple materials science problems, including search of high thermal conductivity polymer, new liquid-crystal polyimide, and high thermal conductivity inorganic crystal. After proposing candidates to a design target, synthesis of the materials is performed by my collaborators at The Tokyo Institute of Technology.

4. 研究成果 (Research results)

I have developed an all-in-one materials informatics platform in Python, called XenonPy, that can produce candidates of materials structures with desired properties (i.e., inverse design). The open source platform includes modules of machine learning algorithms, called transfer learning, that can learn mathematical models to predict material properties using only a small set of data. In particular, the transfer learning modules take a pre-trained neural network, which is a mathematical model that is specialized in learning from large datasets, as input. Assuming the neural network has successfully extracted certain useful knowledge from a prediction task with large data (source task), the modules feature multiple algorithms to transfer the knowledge to a new learning task with little data (target task). In order to cover a wide variety of design targets and material types, having a broad range of source tasks would increase the probability of finding relevant knowledge to a target task. Therefore, I have also developed XenonPy.MDL, a library of pre-trained models covering various properties of small molecules, polymers, and inorganic crystals, by comprehensively extracting data from public databases and literatures. As an analogy learning with human brain, the implementation of a large and diverse set of pre-trained models is equivalent to the construction of a collection of memories from different experiences in life. The more diverse the models are, the more powerful the memory aggregate is, the more powerful the transfer learning may be. As a result, two methodological papers have been published related to the development of XenonPy, each includes validation of inverse design for polymers and successful transfer learning cases for different materials science problems, respectively.

An important aspect in this project is to promote materials informatics for practical industrial uses. Three end-to-end studies of polymer design have been conducted starting with data collection, to learning machine learning models from data, to proposing candidates for a design target, to finally synthesizing the design for verification. The first study targeted new high thermal conductivity polymers, for which the available data is significantly small (tens of data points only). A virtual library of 1,000 polymers targeting high thermal conductivity was created by combining PoLyInfo, the world's largest polymer database, and the iQSPR-X algorithm in XenonPy. Three aromatic polyamides were synthesized from the library by my collaborators at The Tokyo Institute of Technology. In total, we synthesized three aromatic polyamides and discovered a new polymer with a thermal conductivity of 0.41 W/mK. This corresponds to a performance improvement of up to 80% compared to typical polyamide polymers. In addition, these materials have several characteristics required at the stage of industrial applications, such as high heat resistance, solubility in organic solvents, and ease of film processing. Our results have been published and created a high impact on other research works (more than 100 citations in around two years of publication). The second study targeted high lattice thermal conductivity crystals. In this study, transfer learning model learning from data up thermal conductivity of 500 W/mK was able to predict for materials with over 2000 W/mK thermal conductivity. Such kind of extrapolation behavior has never been reported before. Yet, the final synthesis of materials in this study has been proponed to

be done by another collaborator in the near future. The third study targeted new liquid-crystal polyimide, which has high industrial values, such as being used in display devices. Hundreds of potential candidates have been discovered by the algorithms in XenonPy. However, due to the COVID-19 pandemic, synthesis of the candidates is being delayed and the results are still under preparation for publication.

On top of the planned work on transfer learning and autonomous material design, the proposed transfer learning framework has been extended to multiple engineering applications, outside of the scope of XenonPy. For example, transfer learning technique is being used to (1) improve microscopic image processing related to understanding crystal materials, and (2) increase compression rate of crack image compression used for damaged material detection, producing two publications. The XenonPy platform has also been extended for different related materials science problems, such as retrosynthesis of molecules and understanding intrinsic hydrodynamic thermal transport, producing another two publications. While transfer learning has become a core technique in materials informatics, the underlying mechanism of when it will be successful is still unclear. The results from this project has pointed us to a bigger challenge of aiming to control the behavior of transfer learning, which will lead to more efficient algorithms for discovery of novel materials for industrial uses.

Reference:

[1] PoLyInfo, http://polymer.nims.go.jp/, last access: October 16, 2017.

[2] H. Ikebata, K. Hongo, T. Isomura, R. Maezono, R. Yoshida (2017) Bayesian molecular design with a chemical language model, Journal of Computer-Aided Molecular Design, 31(4): 379-391.

５．主な発表論文等

〔雑誌論文〕　計10件（うち査読付論文　9件／うち国際共著　10件／うちオープンアクセス　6件）

１．著者名 Ma Boyuan、Wei Xiaoyan、Liu Chuni、Ban Xiaojuan、Huang Haiyou、Wang Hao、Xue Weihua、Wu Stephen、Gao Mingfei、Shen Qing、Mukeshimana Michele、Abuassba Adnan Omer、Shen Haokai、Su Yanjing	４．巻 6
２．論文標題 Data augmentation in microscopic images for material data mining	５．発行年 2020年
３．雑誌名 npj Computational Materials	６．最初と最後の頁 125
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41524-020-00392-6	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 該当する

１．著者名 Stephen Wu, Hironao Yamada, Yoshihiro Hayashi, Massimiliano Zamengo, Ryo Yoshida	４．巻 NA
２．論文標題 Potentials and challenges of polymer informatics: exploiting machine learning for polymer design	５．発行年 2020年
３．雑誌名 arXiv (in press at Proceedings of the Institute of Statistical Mathematics (2021 special issue))	６．最初と最後の頁 NA
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 無
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 該当する

１．著者名 Huang Yong、Zhang Haoyu、Li Hui、Wu Stephen	４．巻 146
２．論文標題 Recovering compressed images for automatic crack segmentation using generative models	５．発行年 2021年
３．雑誌名 Mechanical Systems and Signal Processing	６．最初と最後の頁 107061～107061
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.ymssp.2020.107061	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 該当する

１．著者名 Shunya Minami, Song Liu, Stephen Wu, Kenji Fukumizu, Ryo Yoshida	４．巻 NA
２．論文標題 A general class of transfer learning regression without implementation cost	５．発行年 2020年
３．雑誌名 Proceedings of the AAAI Conference on Artificial Intelligence	６．最初と最後の頁 NA
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 該当する

1．著者名 Guo Zhongliang、Wu Stephen、Ohno Mitsuru、Yoshida Ryo	4．巻 60
2．論文標題 Bayesian Algorithm for Retrosynthesis	5．発行年 2020年
3．雑誌名 Journal of Chemical Information and Modeling	6．最初と最後の頁 4474～4486
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jcim.0c00320	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 該当する

1．著者名 Wu Stephen、Kondo Yukiko、Kakimoto Masa-aki、Yang Bin、Yamada Hironao、Kuwajima Isao、Lambard Guillaume、Hongo Kenta、Xu Yibin、Shiomi Junichiro、Schick Christoph、Morikawa Junko、Yoshida Ryo	4．巻 5
2．論文標題 Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm	5．発行年 2019年
3．雑誌名 npj Computational Materials	6．最初と最後の頁 66
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41524-019-0203-2	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 該当する

1．著者名 Yamada Hironao、Liu Chang、Wu Stephen、Koyama Yukinori、Ju Shenghong、Shiomi Junichiro、Morikawa Junko、Yoshida Ryo	4．巻 5
2．論文標題 Predicting Materials Properties with Little Data Using Shotgun Transfer Learning	5．発行年 2019年
3．雑誌名 ACS Central Science	6．最初と最後の頁 1717～1730
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acscentsci.9b00804	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 該当する

1．著者名 Wu Stephen、Lambard Guillaume、Liu Chang、Yamada Hironao、Yoshida Ryo	4．巻 39
2．論文標題 iQSPR in XenonPy: A Bayesian Molecular Design Algorithm	5．発行年 2019年
3．雑誌名 Molecular Informatics	6．最初と最後の頁 1900107～1900107
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/minf.201900107	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 該当する

１．著者名 Ju Shenghong、Yoshida Ryo、Liu Chang、Wu Stephen、Hongo Kenta、Tadano Terumasa、Shiomi Junichiro	４．巻 5
２．論文標題 Exploring diamondlike lattice thermal conductivity crystals via feature-based transfer learning	５．発行年 2021年
３．雑誌名 Physical Review Materials	６．最初と最後の頁 53801
掲載論文のDOI（デジタルオブジェクト識別子） 10.1103/PhysRevMaterials.5.053801	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 該当する

１．著者名 Torres Pol、Wu Stephen、Ju Shenghong、Liu Chang、Tadano Terumasa、Yoshida Ryo、Shiomi Junichiro	４．巻 34
２．論文標題 Descriptors of intrinsic hydrodynamic thermal transport: screening a phonon database in a machine learning approach	５．発行年 2022年
３．雑誌名 Journal of Physics: Condensed Matter	６．最初と最後の頁 135702～135702
掲載論文のDOI（デジタルオブジェクト識別子） 10.1088/1361-648X/ac49c9	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 該当する

〔学会発表〕　計16件（うち招待講演　7件／うち国際学会　7件）

１．発表者名 Stephen Wu
２．発表標題 Recovering compressed images for auto-segmentation of building cracks using deep generative models
３．学会等名 The 163rd TCU-ARL Seminar: International Workshop on Data-driven Infrastructure Maintenance and Risk Management
４．発表年 2020年

１．発表者名 Stephen Wu
２．発表標題 Recovering compressed images for auto-segmentation of building cracks using deep generative models
３．学会等名 応用力学講演会 2020（招待講演）
４．発表年 2020年

１．発表者名
Stephen Wu

２．発表標題
A case study of machine-assisted polymer design and other transfer learning applications in materials informatics

３．学会等名
第4 回FRIS 若手研究者学際 融合領域研究会（招待講演）

４．発表年
2019年

１．発表者名
Stephen Wu

２．発表標題
Discovery of new polymers using machine learning models and a Bayesian molecular design algorithm

３．学会等名
The 3rd Forum of Materials Genome Engineering（招待講演）（国際学会）

４．発表年
2019年

１．発表者名
Stephen Wu

２．発表標題
Engineering applications of hierarchical Bayesian modeling

３．学会等名
ISI World Statistics Congress（国際学会）

４．発表年
2019年

１．発表者名
Stephen Wu

２．発表標題
Potential of transfer learning with uncertainty quantification for materials informatics

３．学会等名
Big data and uncertainty quantification: statistical inference and information theoretic techniques applied to computational chemistry conference（国際学会）

４．発表年
2019年

１．発表者名
Stephen Wu

２．発表標題
A Bayesian molecular Design Framework for Searching High Thermal Conductivity Polymers

３．学会等名
統数研・流体研・AIMR合同Workshop

４．発表年
2018年

１．発表者名
Stephen Wu

２．発表標題
Accelerated Discovery of High Thermal Conductivity Polymers with a Bayesian Molecular Design Method

３．学会等名
第67回高分子学会年次大会

４．発表年
2018年

１．発表者名
Stephen Wu

２．発表標題
機械学習に基づくポリマー設計: 合成の壁を超えるには ~ 高熱伝導率ポリマーの設計事例

３．学会等名
The 39th Japan Symposium on Thermophysical Properties

４．発表年
2018年

１．発表者名
Stephen Wu

２．発表標題
Bayesian Inverse Material Design for High Thermal Conductivity Polymers

３．学会等名
I-URICフロンティアコロキウム

４．発表年
2018年

１．発表者名
Stephen Wu

２．発表標題
Engineering applications of transfer learning

３．学会等名
Asia-Pacific-Euro Summer School on Smart Structures Technology 2018（招待講演）（国際学会）

４．発表年
2018年

１．発表者名
Stephen Wu

２．発表標題
Engineering applications of the Bayesian problem solving framework

３．学会等名
Workshop on the Frontier of Applied Bayesian Inference and Computation（招待講演）（国際学会）

４．発表年
2018年

１．発表者名
Stephen Wu

２．発表標題
Potential of transfer learning in engineering applications

３．学会等名
Big data forum on Big Data in Civil Engineering（招待講演）

４．発表年
2018年

１．発表者名
Stephen Wu

２．発表標題
Applications of transfer learning in materials science

３．学会等名
ISI-ISM-ISSAS Joint Conference 2019（国際学会）

４．発表年
2018年

１．発表者名 Stephen Wu
２．発表標題 Potential of applying transfer learning to engineering applications: an example in materials science
３．学会等名 信頼性工学分野におけるデータサイエンス技術の活用に関する講演会（招待講演）（国際学会）
４．発表年 2018年

１．発表者名 Stephen Wu
２．発表標題 高分子のバーチャルスクリーニング
３．学会等名 統計関連学会連合大会
４．発表年 2021年

〔図書〕　計0件

〔出願〕　計1件

産業財産権の名称 ベイズ推論による高熱電導高分子の設計と合成	発明者 森川淳子，吉田亮，Stephen Wu，等	権利者 同左
産業財産権の種類、番号 特許、Ａ００３８６ＪＰ０１	出願年 2018年	国内・外国の別 国内

〔取得〕　計0件

〔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	森川　淳子 (MORIKAWA JUNKO)	東京工業大学・教授 (12608)	
研究協力者	吉田　亮 (YOSHIDA RYO)	統計数理研究所・教授 (62603)	

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