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　６ 年　 ４ 月 ３０ 日現在

機関番号：６２６０３
研究種目：若手研究
研究期間：2021～2023
課題番号：２１Ｋ１４６７５

研究課題名（和文）Building foundation of polymer informatics: ensemble of generators and virtual libraries of diverse functional polymers

研究課題名（英文）Building foundation of polymer informatics: ensemble of generators and virtual libraries of diverse functional polymers

研究代表者

Ｗｕ　Ｓｔｅｐｈｅｎ（Wu, Stephen）

統計数理研究所・先端データサイエンス研究系・准教授

研究者番号：７０８０４１８６

交付決定額（研究期間全体）：（直接経費）　　4,900,000円

研究成果の概要（和文）：本研究では、人工知能（AI）を活用して新しいポリマー（高分子材料）を設計し、発見する方法を開発した。XenonPyというオープンソースのソフトウェアを使用し、様々なポリマーの生成器を作成した。化学者たちはこれらの生成器を用いて、産業用途に適した新しいポリマー構造をAIから提案され、より迅速に新材料の設計が可能となった。社会実装例として、我々はこの技術を用いて、新しい液晶ポリイミドという材料を発見したが、その中のいくつかは高い熱伝導性を示している。これらの材料は将来のエレクトロニクスやその他の産業用途に役立つ可能性がある。

研究成果の学術的意義や社会的意義
本プロジェクトの目的は、データサイエンスの専門知識がない材料科学者でも最新の機械学習技術を活用して、新しいポリマー材料の発見を加速できる基盤技術を提供することである。XenonPyというオープンソースソフトウェアを用いて、液晶ポリイミドなどの新材料を発見した。この実証結果が、実際の産業応用でのこの技術の利用への関心を高めることを期待している。この取り組みにより、材料設計のプロセスが効率化され、新材料の迅速な開発が可能になるとともに、持続可能な技術開発に貢献することができる。

研究成果の概要（英文）：In this study, we developed a method to design and discover new polymers using artificial intelligence (AI). We created various polymer generators using an open-source software called XenonPy. Chemists can use these generators to quickly design new materials based on new polymer structures suggested by AI for any specific industrial applications. As a practical example of this technology, we targeted new material based on liquid crystalline polyimide, which is typically very rare at the moment. We successfully discovered six new material using our technology and all of them have demonstrated high thermal conductivity. These materials have the potential to be useful in future electronics and other industrial applications.

研究分野：統計学

キーワード：polymer informatics　generative models

様　式　Ｃ－１９、Ｆ－１９－１（共通）

１．研究開始当初の背景

In the past decade, there was growing intersection of polymer science and machine learning, known as polymer informatics. This interdisciplinary field aims to accelerate the discovery of functional polymers with specific physical and chemical properties by leveraging data extracted from existing polymer databases. The main challenge lies in the complex, high-dimensional search space of polymers due to their intricate hierarchical structures.

The momentum for this research is further propelled by the historical context provided by the Materials Genome Initiative launched in 2011, which catalyzed the data-driven approach to materials design. Despite significant advances, the scarcity of comprehensive, open-access polymer databases has been a critical bottleneck, as highlighted by notable studies in the field.

Addressing this, the research builds on foundational work from the JST CREST project "Polymer Thermophysical Properties Materials Informatics", which began in 2019 under Prof. Junko Morikawa at the Tokyo Institute of Technology. This project's aim to establish a substantial open database for polymer thermal properties has laid the groundwork for the current study's objectives. The researcher's participation in this project has identified essential needs and opportunities within polymer informatics to employ machine learning for efficient polymer design and discovery.

The current proposal intends to develop an ensemble of polymer generators and virtual libraries tailored to various material properties, employing machine learning ensemble methods to navigate the vast polymer search space. This approach is designed to integrate these tools into a user-friendly platform, making them accessible to polymer scientists without a background in data science, thus democratizing the advanced capabilities of machine learning in polymer research.

２．研究の目的

The goal of this research project is to leverage polymer informatics, blending polymer science with machine learning, to significantly expedite the discovery of novel functional polymers. By developing ensembles of polymer generators and establishing virtual libraries, the project aims to provide a robust framework that allows for the exploration of diverse functional polymers tailored for various industrial applications. This approach is intended to overcome the limitations posed by the vast and complex search space of polymers. Additionally, the project seeks to make these generators and virtual libraries accessible to the polymer science community, facilitating a more efficient and data-driven approach to polymer design. This is envisioned to enhance the design process by integrating expert knowledge with advanced computational tools, thereby fostering a more systematic exploration of new polymer materials with desired properties.

３．研究の方法

The research methods designed to achieve the goal of accelerating the discovery and design of new functional polymers using polymer informatics can be broken down into six aspects, covering both the data science aspect and the open science aspect of the project. Details are explained as follows:

Development of Polymer Generators:

The research will utilize ensemble machine learning methods to navigate the vast and complex search space of polymers, focusing on the generation of structures with specific properties. The approach includes the development of generators using a variety of methods, including N-gram (a Markov Chain probabilistic model for chemical language) and rule-based generators grounded on the empirical knowledge of polymer scientists. These generators are designed to cover different types of polymers, which enable the production of polymers tailored to diverse functional requirements, addressing the challenge of predicting novel polymer structures from high-dimensional data.

Establishment of Virtual Libraries:

Virtual libraries of diverse functional polymers will be constructed, emphasizing the application of unsupervised learning techniques such as UMAP to categorize and organize polymers based on molecular structure similarities and properties. These libraries will be dynamically updated and expanded using new data from ongoing research, forming a comprehensive resource for industrial applications.

Integration and Validation:

The polymer generators and virtual libraries will be integrated into the XenonPy platform, making them accessible to the global polymer science community. Validation of the polymers' properties and functionalities will be conducted through both computational simulations and experimental methods, in collaboration with experts in the field. This phase ensures that the polymers meet the required specifications and real-world applicability.

Collaboration and Resource Utilization:

This project will harness the extensive polymer data from open databases and utilize high-performance computing resources, including GPU clusters, to support the development and validation processes. Collaborations with specialists in computational simulations and experimental polymer science are crucial, enhancing the project's scope and effectiveness. This project is closely related to the RadonPy project, which aims to develop the largest open database for polymer simulations.

Public Accessibility and Community Engagement:

All developed tools, libraries, and data will be made available through the open-source XenonPy platform, promoting accessibility and collaborative improvement. The project aims to engage the broader polymer science community, encouraging the use of these innovative tools and fostering an environment of shared knowledge and continuous innovation.

Feedback and Continuous Improvement:

The project will employ an iterative approach to refine the generators and virtual libraries, using new data and community feedback. Advanced machine learning techniques, such as transfer learning, will be adapted to enhance the capabilities of the polymer generators, particularly in data-scarce scenarios, ensuring the platform remains at the cutting edge of polymer informatics and material science.

These methods collectively aim to revolutionize how polymer materials are designed and discovered, harnessing the power of data science and machine learning to create a more efficient, creative, and collaborative approach in the field of polymer science.

４．研究成果

Over the past three years, our research initiative has significantly advanced the field of polymer informatics by integrating machine learning with traditional polymer science to enhance the speed and efficiency of discovering new functional polymers. Utilizing our open-source materials informatics software, XenonPy, developed by our group, we have successfully created a suite of sophisticated NGram-based generators, as well as a set of rule-based polymer generators that provide a proposed synthetic path along with the generated polymer candidates. We continue to expand our collection, including deep generative models.

In order to demonstrate and promote our design approach, we collaborated with Prof. Junko Morikawa and Prof. Teruaki Hayakawa’s research group in Tokyo Institute of Technology in order to make a proof-of-concept study that discover and synthesis new liquid crystalline polyimide using our generative approach. This diverse array of tools has enabled the discovery of six novel liquid crystalline polyimides. More importantly, all of them demonstrated promising high thermal conductivity, which highlights the potential of our approach for industrial applications.

Furthermore, our project has worked hand-in-hand with RadonPy, a project aimed at developing the largest open calculation database for polymers, which showcases the potential of our foundational technology to spur further research and collaboration within the industry. The virtual libraries have been rigorously validated both computationally and experimentally, confirming the effectiveness of our machine learning models in identifying polymers with desirable properties. These libraries not only assist in the rational design of polymers but also facilitate the rapid prototyping and testing of new polymer materials.

In conclusion, our research has transformed the approach to polymer design, making it more data-driven and efficient. By providing powerful tools and a robust platform through XenonPy, we are enabling polymer scientists to quickly harness the potential of vast datasets for the development of advanced materials. This integration of machine learning into polymer science represents a significant step forward in the field, offering new avenues for innovation and application in various industrial sectors. Our work will continue to evolve and expand even after the end of this research project.

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　6件／うちオープンアクセス　2件）

１．著者名 Iwayama Megumi、Wu Stephen、Liu Chang、Yoshida Ryo	４．巻 62
２．論文標題 Functional Output Regression for Machine Learning in Materials Science	５．発行年 2022年
３．雑誌名 Journal of Chemical Information and Modeling	６．最初と最後の頁 4837～4851
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jcim.2c00626	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 Zamengo Massimiliano、Wu Stephen、Yoshida Ryo、Morikawa Junko	４．巻 218
２．論文標題 Multi-objective optimization for assisting the design of fixed-type packed bed reactors for chemical heat storage	５．発行年 2023年
３．雑誌名 Applied Thermal Engineering	６．最初と最後の頁 119327～119327
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.applthermaleng.2022.119327	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 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

1．著者名 Zhang Qi、Liu Chang、Wu Stephen、Hayashi Yoshihiro、Yoshida Ryo	4．巻 3
2．論文標題 A Bayesian method for concurrently designing molecules and synthetic reaction networks	5．発行年 2023年
3．雑誌名 Science and Technology of Advanced Materials: Methods	6．最初と最後の頁 2204994
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/27660400.2023.2204994	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 Aoki Yuta、Wu Stephen、Tsurimoto Teruki、Hayashi Yoshihiro、Minami Shunya、Tadamichi Okubo、Shiratori Kazuya、Yoshida Ryo	4．巻 56
2．論文標題 Multitask Machine Learning to Predict Polymer-Solvent Miscibility Using Flory-Huggins Interaction Parameters	5．発行年 2023年
3．雑誌名 Macromolecules	6．最初と最後の頁 5446～5456
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.macromol.2c02600	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

〔学会発表〕　計13件（うち招待講演　0件／うち国際学会　7件）

1．発表者名 Stephen Wu
2．発表標題 Challenges and opportunities in polymer informatics from a statistical perspective
3．学会等名 CONFERENCE ON A FAIR DATA INFRASTRUCTURE FOR MATERIALS GENOMICS（国際学会）
4．発表年 2022年

1．発表者名 Stephen Wu
2．発表標題 Data-driven design and synthesis of new liquid crystal polyimides
3．学会等名 ACS Fall Meeting（国際学会）
4．発表年 2022年

１．発表者名
Stephen Wu

２．発表標題
専門知識と機械学習の融合に基づく液晶性ポリイミドの設計

３．学会等名
統計関連学会連合大会

４．発表年
2022年

１．発表者名
Stephen Wu

２．発表標題
データ駆動型の液晶ポリイミド設計

３．学会等名
第71回高分子討論会

４．発表年
2022年

１．発表者名
Stephen Wu

２．発表標題
Design of high thermal conductivity liquid-crystalline polyimides based on interplay of expert knowledge and machine learning

３．学会等名
The 13th Asian Thermophysical Properties Conference（国際学会）

４．発表年
2022年

１．発表者名
Stephen Wu

２．発表標題
高熱伝導性液晶性ポリイミドのデータ駆動型設計

３．学会等名
第43回日本熱物性シンポジウム

４．発表年
2022年

１．発表者名
Stephen Wu

２．発表標題
Design of liquid-crystalline polyimides by integrating expert knowledge to a data-driven machine learning framework

３．学会等名
The 17th Pacific Polymer Conference（国際学会）

４．発表年
2022年

１．発表者名
Stephen Wu

２．発表標題
Data-Driven Design of Liquid-Crystalline Polyimides by Interplay with Expert Knowledge

３．学会等名
ISI-ISM-ISSAS Joint Conference 2023（国際学会）

４．発表年
2023年

１．発表者名
Stephen Wu

２．発表標題
XenonPy: an open source platform for data-driven materials design with small data

３．学会等名
TMS 2023 152nd Annual Meeting & Exhibition（国際学会）

４．発表年
2023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

１．発表者名
Stephen Wu

２．発表標題
Polymer Design by Interplay of Machine Learning, Computer Simulation and Expert Knowledge

３．学会等名
64th World Statistics Congress（国際学会）

４．発表年
2023年

１．発表者名
白鳥 和矢，青木 裕太，Stephen Wu，釣本 輝希，林 慶浩，南 俊匠，大久保 忠利，吉田 亮

２．発表標題
マルチタスク学習を用いたポリマー・溶媒系における Flory-Huggins χ パラメータの予測

３．学会等名
日本高分子討論会

４．発表年
2023年

１．発表者名
中川琹，Felix Jiang，前田颯，Stephen Wu，林慶浩，劉芽久哉，Xuan Thang Vu，Sven Ingebrandt，吉田亮，早川晃鏡，森川淳子

２．発表標題
集積センサーアレイとフォトサーマルロックイン法によるポリイミド薄膜の面内熱拡散率測定と高速熱分析

３．学会等名
第59回熱測定討論会

４．発表年
2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
XenonPy Open software: https://xenonpy.readthedocs.io/en/latest/

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) (12608)	東京工業大学・教授	
研究協力者	早川　晃鏡 (Hayakawa Teruaki) (12608)	東京工業大学・教授	
研究協力者	林　慶浩 (Hayashi Yoshihiro) (62603)	統計数理研究所・助教	
研究協力者	吉田　亮 (Yoshida Ryo) (62603)	統計数理研究所・教授	

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