

Starrydata web システムによる熱電材料の Materials Informatics

Materials Informatics of Thermoelectric Materials by using Starrydata Web System

東大新領域¹, NIMS², 理研³, さくらインターネット(株)⁴

○桂 ゆかり^{1,2,3}, 熊谷 将也^{3,4}, 小谷 拓史¹, 佐藤 陸¹, 木村 薫¹, 津田宏治^{1,2,3}

○Yukari Katsura^{1,2,3}, Masaya Kumagai^{3,4}, Takushi Kodani^{1,2}, Riku Sato^{1,2},

Kaoru Kimura^{1,2,3}, Koji Tsuda^{1,2,3}

The Univ. of Tokyo¹, NIMS², RIKEN³, SAKURA Internet Inc.⁴

【緒言】近年、データ科学を材料科学に活用する試みとして、Materials Informatics (MI)の研究が発展している。MI 研究の主流は計算データを使うものであるが、この一因として、MI に利用できる規模のまとまった実験データが手に入りにくいという問題がある。そこで我々は、論文中のグラフ画像から大規模実験データを効率的に集められる独自の論文データ収集 Web システム Starrydata を開発してきた (図 1)。人の手を借りた高品質なデータ収集が特徴であり、論文単位でデータを格納・共有できる。本 Web システムとデータは <https://www.starrydata2.org> において公開しており、営利・非営利を問わず、論文[1]の引用だけで無償で利用できる。これまでの3年で、研究業務員と外部ユーザーの力を借りて、6000 本以上の論文から 30,000 試料、90,000 本以上のカーブのデータを収集してきた[1]。現在は、熱電材料と磁性材料についてデータ収集を進めており、今後、機能材料の種類を増やしていく予定である。そこで本研究では、本システムで集めた熱電材料のデータを機械学習することによって、熱電特性の予測を行った結果を報告する。

【方法】Starrydata 上で集めた熱電材料のデータについて、各試料の化学組成と温度から熱電特性 (ゼーベック係数 S , 電気伝導率 σ , 熱伝導率 κ , 無次元性能指数 ZT) の予測を行った。学習はニューラルネットワーク(NN)およびランダムフォレスト(RF)法によって行い、学習済みの機械学習モデルを用いて、5 元素系の複雑組成 (数万点) の PbTe 系熱電材料の ZT の予測を試みた。また、7000 種類の既知半導体の ZT を予測し、有望な化合物のランキングを作製した。

【結果と考察】NN と RF のいずれの予測においても、複雑な組成領域において高い ZT が予測された。既知半導体の ZT の予測では、NN と RF で異なる結果となり、RF では既知熱電材料に近い材料において高い

ZT が予測される傾向があった。図 1. Starrydata におけるグラフデータの収集のコンセプト[1]。

参考文献 [1] Y. Katsura, M. Kumagai, T. Kodani *et al.*, *Sci. Technol. Adv. Mater.* **20**, 511 (2019).

謝辞 本研究は自動車用内燃機関技術研究組合 (AICE), 科研費, JST-CREST の支援を受けたものである。

