Hybrid Modeling of Mixed Perovskite Oxides with Gaussian Process Regression: Overcoming Data Scarcity

Franz M. Rohrhofer[1,*], Jean-François Paul[2], Bernhard C. Geiger[1], and Elise Berrier[2]

[1] Know Center Research GmbH, Sandgasse 34, 8010 Graz, Austria
[2] Université Lille, CNRS UMR8181, Unité de Catalyse et de Chimie du Solide, UCCS, F-59655 Villeneuve d'Ascq, France
* Corresponding author: frohrhofer@know-center.at

Key Words: *Hybrid Modeling, Mixed Perovskite Oxides, Gaussian Process Regression, Machine Learning, Data Scarcity.*

Abstract

Mixed perovskite oxides present a diverse class of potential electrode materials for chemical-to-power devices, enabling the reduction of critical raw materials while maintaining high performance and economic viability. However, multi-scale modeling of these materials remains computationally challenging due to the vast number of possible elemental compositions in perovskite-based structures[1]. First-principles methods, such as Density Functional Theory (DFT), can only be applied to a limited subset of materials, providing highly accurate yet data-scarce estimates of target properties.
Hybrid modeling approaches that integrate machine learning offer a promising solution by leveraging available data to predict material properties with significantly lower computational cost compared to conventional methods. However, the prevalent data scarcity necessitates careful selection of machine learning models that can effectively handle the complexity of mixed perovskite oxides while ensuring accuracy, efficiency, and reliability.
In this work, we employ Gaussian Process Regression (GPR) for this task, leveraging its demonstrated potential in materials modeling[2]. Our approach utilizes a well-defined and interpretable set of descriptors/features that encode the structural characteristics of mixed perovskite oxides, enabling an effective representation for GPR with a custom kernel. Experiments on predicting DFT-based energies, using data from the Materials Project[3] and an in-house generated database, validate the effectiveness of our approach. Results show that our model can efficiently and accurately predict target properties while being trained on only a limited number of examples.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation program through the KNOWSKITE-X Project under grant agreement no. 101091534.

References

[1] Tao, Qiuling, et al. "Machine learning for perovskite materials design and discovery." *Npj computational materials* 7.1 (2021): 23.
[2] Bartók, Albert P., et al. "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons." *Physical review letters* 104.13 (2010): 136403.
[3] Jain, Anubhav, et al. "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation." *APL materials* 1.1 (2013).