Data as Code - Sharing Machine Learned Interatomic Potentials with MLSuite and ZnTrack

Fabian Zills[1]

[1]) University of Stuttgart, Institute of Computational Physics, Allmandring 3, 70569 Stuttgart, Germany
fzills@icp.uni-stuttgart.de

Machine learned potential energy surfaces (ML-Pes) provide a cost-effective and accurate way to run molecular dynamics (MD) simulations. ML-Pes are typically trained on energies on forces from first principle methods such as density functional theory and almost reproduce their accuracy[1, 2].

In our research, we applied ML-Pes to study room temperature ionic liquids, which have high viscosity and therefore, require long simulation times to investigate dynamic properties. However, constructing and analyzing ML-Pes involves multiple steps and parameters, that necessitate specialized research software. Our workflow typically consists of many steps, such as

- data generation.
- model training.
- model deployment.

We developed MLSuite, a python package that provides a comprehensive suite of tools for ML-Pes research. MLSuite includes state-of-the-art models like MACE and Nequip, as well as analysis and data generation tools[3, 4]. By applying MLSuite, we were able to streamline our workflow and easily generate, analyze and train models on data for our research.

MLSuite is built on top of ZnTrack, an open-source package that provides a flexible framework for managing data-focused workflows[5]. ZnTrack is built on the concept of Data as Code (DaC), which means that data and metadata are treated as code, making them versionable, shareable, and reusable. For this purpose we developed ZnTrack to work with GIT and Data Version Control.

Moreover, ZnTrack is a flexible tool that can be used for many different types of data-focused workflows. It was important to us to provide ZnTrack as an open source tool so that others can easily use and modify it for their own research. We believe that open-source software is crucial for advancing scientific research and hope to contribute to it.

References

[1] J. Behler, "Constructing high-dimensional neural network potentials: A tutorial review", International Journal of Quantum Chemistry **115**, 1032–1050 (2015).

[2] A. P. Bartók and G. Csányi, *Gaussian Approximation Potentials: a brief tutorial introduction*, Feb. 5, 2020.

[3]S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials", Nature Communications **13**, 2453 (2022).

[4]I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner, and G. Csányi, *MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields*, June 15, 2022.

[5]F. Zills and S. Tovey, *Zincware/ZnTrack: ZnTrack 0.5.1*, Zenodo, Jan. 31, 2023.