AKPIK 2: Machine Learning & Physics

Time: Wednesday 15:00–16:15 Location: MAR 0.002

AKPIK 2.1 Wed 15:00 MAR 0.002

Bringing long-ranged interactions to the JAX ecosystem with the multilevel summation method — •Florian Buchner[1], Johannes Schörghuber[1], Jesús Carrete[2,1], and Georg K. H. Madsen[1] — [1]Institute of Materials Chemistry, TU Wien, 1060 Vienna, Austria — [2]Instituto de Nanociencia y Materiales de Aragón (INMA), CSIC-Universidad de Zaragoza, 50009 Zaragoza, Spain

Despite the tremendous success of machine-learned force fields (MLFFs), their extension beyond the locality approximation remains a field of active research. In constructing such MLFFs, the efficient (ideally with linear scaling) evaluation of pairwise long-ranged interactions is a ubiquitous requirement. It is routinely solved by well-established algorithms such as Ewald summation.

While implementations of such algorithms are readily available, they tend not to interface well with modern machine-learning environments and workflows. This includes Google's JAX framework, which is proving transformative to machine-learning research by providing high performance and general-purpose automatic differentiation.

Here, we present a JAX-based implementation of the multilevel summation method (MSM) [D. J. Hardy *et al.*, J. Chem. Phys. 144, 114112 (2016)], a powerful linearly scaling algorithm for pairwise long-ranged interactions. Its notable features include support for mixed boundary conditions and freedom from artefactual force discontinuities encountered in competing methods. We introduce the basics of the MSM, discuss our design and implementation strategy, and highlight example applications.

AKPIK 2.2 Wed 15:15 MAR 0.002

Atomic Graph-based Symmetry Search for Enhancing Machine Learning Force Fields Architectures — •Anton Charkin-Gorbulin[1,2], Igor Poltavsky[1], Alexandre Tkatchenko[1], Claudio Quarti[2], and David Beljonne[2] — [1]University of Luxembourg, Luxembourg, Luxembourg — [2]University of Mons, Mons, Belgium

Machine-learning force fields (MLFF) show high accuracy and efficiency for modeling the potential energy surfaces of molecules, materials, and interfaces. However, the performance of MLFFs greatly depends upon incorporating the physical symmetries. Finding all relevant symmetries becomes a challenging task for large system sizes. Here, we develop a data-driven method based on molecular graphs to reveal relevant permutational symmetries and distinguish atoms with different chemical environments in molecules and materials.

The method was applied to improve the performance of the kernel ridge regression (KRR) model and the message-passing neural network (MPNN). KRR model, enhanced with extracted symmetries, demonstrates superior accuracy, enabling comprehensive investigations of complex systems like the 1,8-naphthyridine/graphene interface at finite temperatures. MPNN was enhanced by expanding atomic species using the extracted distinctive chemical environments, resulting in improved accuracy for $CsPbI_3$ slab systems, particularly notable with large training sets. Overall, this research underscores the critical role of symmetries in advancing MLFFs for complex systems, enabling further advances in atomistic simulations.

AKPIK 2.3 Wed 15:30 MAR 0.002

Combining genetic algorithm and compressed sensing for features and operators selection in symbolic regression — •Aliaksei Mazheika[1], Sergey V. Levchenko[2], and Luca M. Ghiringhelli[3,4] — [1]Technische Universitaet Berlin, DE — [2]Moscow, RU — [3]The NOMAD Laboratory at the Fritz Haber Institute and Humboldt University, Berlin, DE — [4]Friedrich-Alexander University, Erlangen, DE

The symbolic inference method SISSO (Sure-Independence Screening and Sparsifying Operator) has recently found a broad application in materials science. It performs regression or classification by adopting compressed sensing for the selection of an optimized subset of features and mathematical operators out of a given set of candidates. However, SISSO becomes computationally unpractical when the set of candidate features and operators exceeds the size of few tens. Here we combine SISSO with a genetic algorithm (GA) for the global search of the optimal subset of features and operators. We test GA-SISSO for the search of predictive models of perovskites lattice parameters, and demonstrate that our method efficiently finds more accurate models than the original SISSO. GA-SISSO was also applied for the search of the model for prediction of CO_2 adsorption energies on semiconductor oxides. The model learned by GA-SISSO has much higher accuracy compared to previously discussed models based on the O 2p-band center. The statistical analysis of contributions of all features to the learned models shows that, besides the O 2p-band center, the electrostatic potential above adsorption sites and surface formation energies are key features.

AKPIK 2.4 Wed 15:45 MAR 0.002

PSeudocode Projective Simulation (PS^2) — •Marius Krumm and Hans J. Briegel — University of Innsbruck, Institute for Theoretical Physics, Technikerstr. 21a, A-6020 Innsbruck, Austria

The rise of deep learning has enabled significant progress in technology and science. However, the opaque oracle-like nature of artificial neural networks severely limits their potential to discover new qualitative scientific insights. This motivates the exploration of methods from eXplainable AI (XAI) that allow to understand the reasoning process of accurate predictions. In this talk, I present a new XAI method called PSeudocode Projective Simulation (PS^2) which represents chains-of-thought in the form of pseudocodes. Here, a thought is modeled as a small data processing module acting on the agent's memory, which can be a trainable neural network. These subroutines are selected in a trainable random walk, making our method an extension of Projective Simulation. On a technical level, methods from hierarchical and safe reinforcement learning are adapted and integrated into our setting. These modified methods help to model domain knowledge about the nature of the thoughts and the environment. The framework is applied numerically to the Highway-Env.

AKPIK 2.5 Wed 16:00 MAR 0.002

Humans in the loop for more trustworthy Bayesian optimization of materials — •Armi Tiihonen[1], Louis Filstroff[2], Petrus Mikkola[1], Emma Lehto[1], Samuel Kaski[1,3], Milica Todorović[4], and Patrick Rinke[1] — [1]Aalto University, Espoo, Finland — [2]ENSAI, CREST, Rennes, France — [3]The University of Manchester, Manchester, United Kingdom — [4]University of Turku, Turku, Finland

Bayesian optimization (BO) is a machine learning method for global optimization of black-box functions, e.g. the composition of perovskite materials for more stable solar cells. BO can be coupled to automated sample preparation and characterization pipelines, which introduces the challenge of ensuring sufficient sample quality during the optimization. Low quality samples are hard to detect automatically, but could obscure the optimization process. To make automated materials optimizations more robust and trustworthy, we add humans into the BO loop (HITL) as an additional data source. We implemented three HITL schemes, two based on data fusion and one on multifidelity BO. They query human opinion on sample quality only when necessary and guide the sampling away from composition regions with a lot of low-quality samples. We tested them in simulations based on previously obtained experimental perovskite data. Our data fusion HITL BO queries from humans on average 7% of the samples when the run is repeated 25 times. This leads to only 2% of low quality samples, in contrast to the 25% with the reference method without humans. Thus, HITL ensures more consistent sample quality during BO.