Advances in active learning for materials research

Milica Todorović[1],

[1]University of Turku, Yliopistonmäki, Turku, Finland

milica.todorovic@utu.fi

Materials research can be accelerated with active machine learning methods like Bayesian optimization (BO), where datasets are collected on-the-fly in the search for optimal outcomes. We encoded this probabilistic algorithm into the Bayesian Optimization Structure Search (BOSS) Python tool [1] and deployed it to study complex phenomena such as surface and interface chemistry, gas and biosensors [2], magnets [3] and in AI-guided experimental protocols (see Figure 1). Active learning is now a well-established method in both computational and experimental materials research, achieving solutions previously inaccessible through human intuition.

However, materials data acquisitions are often costly and datasets small. Further algorithm development and transfer of knowledge from computer science is necessary to focus and accelerate active learning. In materials science, we demonstrate how data gradients and materials symmetry [4] can be leveraged to fit better models with less data. We used human-in-the-loop strategies to direct the search towards desired areas of parameter space [5]. Multi-task Gaussian Process models [6] were deployed to integrate information from different simulation types into multi-fidelity BO, so that less accurate data can inform the search at a higher level of accuracy. Innovative and complementary algorithms can deliver more sophisticated BO decision making and faster discoveries to boost materials development.

References

[1] M. Todorović et al., npj Comput. Mater., 5, 35 (2019).
[2] S. B. Malik et al, ACS Appl. Mater. Interfaces 16, 48585–48597 (2024).
[3] J. Baumsteiger et al., Dig. Disc. 4, 1639-1650 (2025).
[4] J. Li et al., Small Struct. 2400268 (2024).
[5] A. Tiihonen et al., AI for Accelerated Materials Design NeurIPS 2022 Workshop (2022)
[6] E. V. Bonilla et al., Adv. Neural Inf. Process. 20 (2007).

Figures

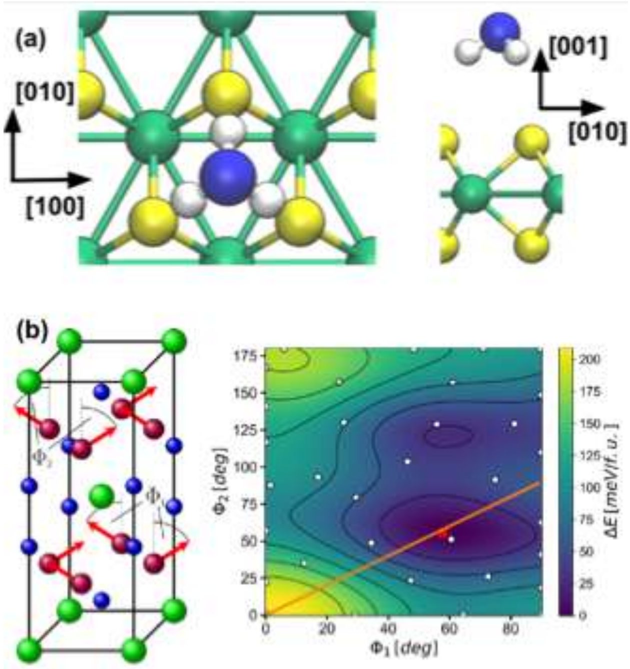

Figure 1. Bayesian optimization for a) atomistic structure search of NH_3 gas sensing via WS_2 2D material [2] and b) studies of bulk $LaMn_2Si_2$ spin-spin interactions via magnetic energy landscapes [3].