Bayesian Optimization for the Inverse Problems in Materials Science

Yixuan Zhang[1], Xiankang Tang[1], Mian Dai[1], Hongbin Zhang[*1]
[1]*Institute of Materials Science, Technical University of Darmstadt, 64287 Darmstadt, Germany*

Correspondence to: [Hongbin Zhang] hongbin.zhang@tu-darmstadt.de

1. Introduction

Focusing on mapping the many-to-one composition – processing – structure – property (CPSP) relationships in materials design, the applications of machine learning approaches have empowered the fourth (data-driven research) and fifth (autonomous discovery) paradigms, unlocking the inverse design of materials. In particular, Bayesian optimization (BO) offers a universal framework, where uncertainty is explicitly leveraged in sequential exploration of the design space. However, how physical knowledge can be incorporated, how to overcome the dimensionality curse, as well as how to guarantee convergence towards global minima are three representative challenges applying BO for inverse design. In this work, going beyond the standard closed-loop adaptive design, we demonstrate how such challenges can be tackled by showcasing the applications of BO in structural and spectroscopic characterizations as well as effective sampling integrated with ensemble fusion, Monte Carlo simulations and reinforcement learning.

2. Methodology

In conventional BO, a closed-loop optimization is usually done on the existing databases in order to identify the optimal surrogate model + acquisition function combinations. Going beyond it, we suggest an ensemble-empowered approach, where various base learners from multiple algorithmic families (such as regressions, tree-based approaches, Gaussian processes, and neural networks) are constructed, in order to (1) avoid the numerical problem upon evaluating the covariance matrices for high-dimensional design spaces and (2) enable multi-model fusion.

3. Results and discussion

For spectroscopic characterizations in Materials science, the inverse problem is the challenge. Focusing on those cases where simulations can be performed to map either the crystal structures or the electronic structure to experimentally measured spectroscopies, it is demonstrated that BO can be applied to solve the inverse problem for x-ray absorption spectroscopy[1], x-ray diffraction, transmission electron microscopy (TEM), low-energy electron diffraction, and inelastic neutron scattering[2]. For instance, for high-resolution TEM, by embedding physical constraints directly into the probabilistic sampling process, our approach leverages a symbiotic loop between image simulation and inference, where the BO engine efficiently explores the coupled parameter space while ensuring physical consistency, enabling the three-dimensional crystal structure reconstruction from a single experimental high-resolution TEM image.

Furthermore, taking the high-entropy alloys as an example, where the chemical space is huge, we show how Monte Carlo sampling[3] and reinforcement learning can be applied for multi-objective optimization to recommend compositions on the Pareto front. In the former case, combined with on-the-fly density functional theory calculations, Monte Carlo sampling enables controlled efficient screening for the optimal compositions, starting with limited (~50) initial random compositions. In the latter cases, based on a surrogate model trained on 400,000 equimolar compositions, reinforcement learning can give rise to reasonable suggestions on the new compositions to improve the accuracy of the surrogate model. On the other hand, it is demonstrated that reinforcement learning is an effective approach to search for the global minima.

Lastly, we present an information-theoretic multi-model information fusion BO framework[4], where multiple information channels including diverse surrogate models, bootstrap ensembles, and multi-level mutual information between models and data are integrated to establish complementary insights and to formulate a dynamic sampling strategy. Such a framework is applied on 14 single- and multi-objective materials design tasks spanning datasets with 600 – 4,000,000 samples, feature dimensions of 10 – 1024, and one to four target properties. Starting from 20 randomly selected initial samples, such an information-theoretic approach manages to identify the top-10 candidates within fewer than 10 iterations for most cases. We benchmarked such an implementation using three representative 20-

dimensional analytical models, i.e., Ackley, Schwefel, and Rastrigin, where robust performance is obtained.

In short, it is demonstrated that BO approaches with advanced sampling strategies can be applied to tackle the inverse problems of materials science, unlocking an important element for closed-loop design in future autonomous experimentation.

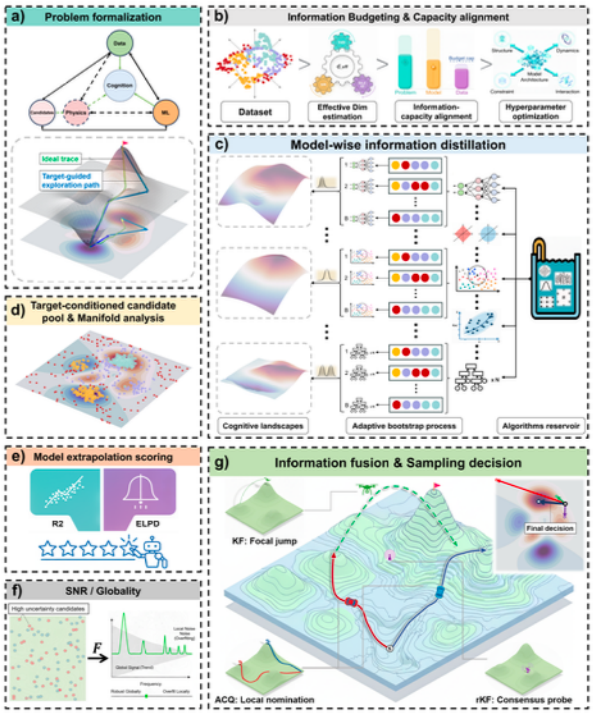

Fig. 1: Information-theoretic multi-source fusion pipeline for target-oriented adaptive sampling, cf. [4] for more details.

Acknowledgments

This work is financially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project ID No. 405553726, TRR 270

References

[1] Y. Zhang, R. Xie, T. Long, D. Günzing, H. Wende, K.J. Ollefs, and H. Zhang. Autonomous atomic Hamiltonian construction and active sampling of X-ray absorption spectroscopy by adversarial Bayesian optimization. npj Computational Materials 9, 46, 2023.

[2] N. Abuawwad, Y. Zhang, S. Lounis, and H. Zhang. Kalman filter enhanced active learning sampling for inelastic neutron scattering: The case of CrSBr. Phys. Rev. B 111, 054404, 2025.

[3] M. Dai, Y. Zhang, W. He, C. Shen, X. Li, S. Schönecker, L. Han, R. Xie, T. Zhou, and H. Zhang. Accelerated Design of Mechanically Hard Magnetically Soft High-entropy Alloys via Multi-objective Bayesian optimization. arXiv:2509.05702

[4] Y. Zhang, Z. Li, W. He, M. Dai, C. Shen, T. Long, and H. Zhang. Information-theoretic multi-model fusion for target-oriented adaptive sampling in materials design. arXiv:2602.03319