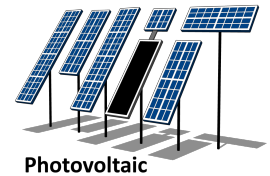
Benchmarking the Performance of Bayesian Optimization across Multiple Experimental Materials Science Domains

Harry Qiaohao Liang[1], Aldair E. Gongora[2], Zekun Ren[3], Zhe Liu[1], Armi Tiihonen[1], Shijing Sun[1], Flore C.L. Mekki-Berrada[4], Saif A. Khan[4], Daniil Bash[5], Kedar Hippalgaonkar[5], James R. Deneault[6], Benji Maruyama[6], Keith A. Brown[2], John Fisher III[1], Tonio Buonassisi[1]

[1]Massachusetts Institute of Technology, Cambridge, USA. [2]Boston University, Boston, USA. [3]Singapore-MIT Alliance for Research and Technology, Singapore. [4]National University of Singapore, Singapore. [5]Agency for Science, Technology and Research (A*STAR), Singapore. [6]Air Force Research Laboratory, USA.



Motivation

- In the field of machine learning (ML) for materials optimization, active learning algorithms, such as Bayesian Optimization (BO), have been extensively used to guide high-throughput autonomous experimentation systems.
- Previous studies [1] benchmarked BO's performance within specific electrocatalyst composition space, but whether their observation of ML models and acquisition functions are applicable to broader array of materials research remains unanswered.
- Benchmarking many off the shelf BO algorithms' performance across five diverse experimental materials science domains allow us to observe different their performance across a broader range of materials systems.

Scope of BO

BO: Global minimization problem

$$\vec{x}^* = \underset{\vec{x} \in X}{\operatorname{argmin}} f(\vec{x})$$

Surrogate models

(1) Random Forest (RF)
n_{tree} = 50, bootstrap = True.
(2) Gaussian Process (GP)
kernel = Matérn52 / Matérn32 / Matérn12 / RBF / MLP
(3) GP with ARD
Automatic relevance detection (ARD): assigns independent lengthscales l_i for each input dim i

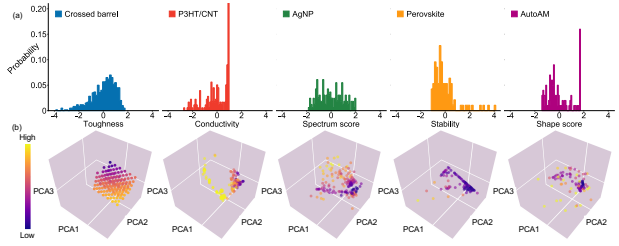
Acquisition functions

$$\vec{x}^* = \underset{\vec{x}}{\operatorname{argmax}}\, \alpha(\vec{x}) \quad \text{at each learning cycle}$$

(1) Lower confidence bound (LCB)
Tunable exploration and exploitation
$\mathrm{LCB}_\lambda = \mathrm{LCB}_{\omega_1\omega_2} = -\omega_1 \cdot mean + \omega_2 \cdot std, \lambda = \frac{\omega_1}{\omega_2}$
(2) Expected improvement (EI)
(3) Maximum probability of improvement (MPI)

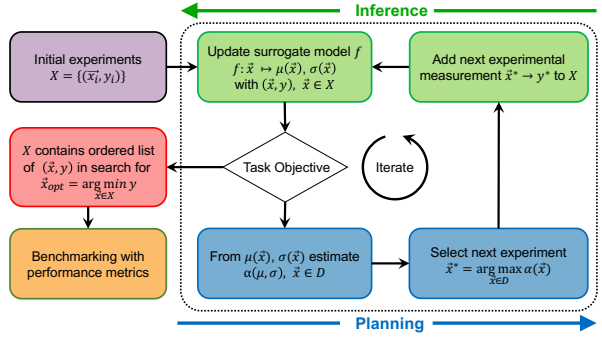
Experimental Datasets

Dataset	Domain	Synthesis	Size	n_{dim}	Optimization
P3HT/CNT[*]	Composites	Drop cast	178	5	Conductivity
AgNP[2]	Nanoparticles	Flow synthesis	164	5	Spectrum
Perovskite[3]	Perovskites	Spin coating	94	3	Stability
Crossed Barrel [4]	3D Printed Structures	Extrusion printing	600	4	Toughness
AutoAM[*]	Manufacturing	Extrusion printing	100	4	Shape score

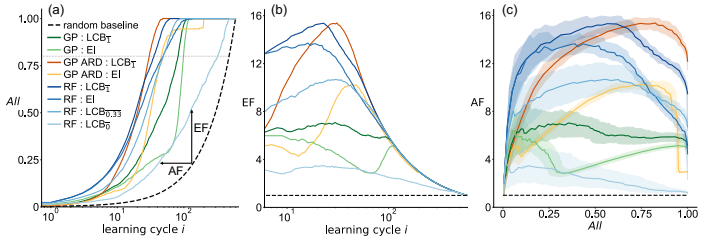


Dataset manifold complexity analysis. a) Histogram of normalized objective values. Each dataset's objective values are independently centered to their mean and scaled to unit variance. b) Design space after dimension reduction to 3D via principal component analysis (PCA).

Benchmarking Framework



Results



Task oriented metric [1]
$All\ (i)$: fraction of top 5% performing sample by cycle i

Enhancement factor
$EF(i) = \frac{All_{BO}(i)}{All_{random}(i)}$
at cycle i

Acceleration factor
$AF(a) = \frac{i_{BO}(a)}{i_{random}(a)}$
at $All = a$

(1) Under same acquisition function, RF's performance as surrogate model is on par, if not slightly worse, compared to that of GP equipped with ARD.
(2) Both RF and GP (ARD) outclass GP w/o ARD as surrogate model.
(3) Acquisition functions: balanced > overly greedy or exploratory.
In the context of high-throughput experiments for materials optimization,
(4) RF has advantage in time complexity, robustness as ensemble learning method, and allows one to make fewer structural assumptions about unfamiliar domain manifolds at initial hyperparameter selection.
(5) GP with ARD provides individual characteristic lengthscales for each dimension, which provide "weights" for understanding the objective's sensitivity to each input dimension.

Conclusion

(1) Under same acquisition function, RF's performance as surrogate model is on par, if not slightly worse, compared to that of GP equipped with ARD. Both RF and GP with ARD outclass GP without ARD as surrogate model.
(2) Besides GP, RF also warrants consideration as surrogate model for future materials optimization campaigns.
(3) It would be good practice for researchers in the field to emphasize their use of ARD in GP as surrogate model.

Reference

[1] Rohr, et al. "Benchmarking the acceleration of materials discovery by sequential learning."Chemical Science 11.10(2020)
[2] Mekki-Berrada, et al. "Two-Step Machine Learning Enables Optimized Nanoparticle Synthesis." ChemRxiv preprint(2020).
[3] Sun, et al. "A Physical Data Fusion Approach to Optimize Compositional Stability of Halide Perovskites." ChemRxiv (2020).
[4] Gongora, et al. "A Bayesian experimental autonomous researcher for mechanical design." Science Advances 6.15 (2020).
[*] Datasets will be published by original authors in journal or on arxiv in the future.