Adaptive Experimental Design to Accelerate Scientific Discovery and Engineering Design

Aryan Deshwal

adeshwal@umn.edu
University of Minnesota

One important class of problems that lies at the heart of accelerating scientific discovery and engineering design is that of adaptive experimental design where we need to select a sequence of experiments in order to achieve a goal in a resource-efficient manner. For example, in materials/-chemical design, scientists are interested in finding the best material where making and evaluating a candidate material involves performing an expensive physical lab experiment. Similarly, in hardware design, engineers are interested in finding high-performing design where evaluating each candidate design involves performing a computationally-expensive simulation to emulate the real hardware.

In this talk, we will focus on this general problem setting where the goal is to find the best design over a given input design space in order to optimize an objective of interest guided by expensive experiments. Commonly, evaluating each candidate input design requires performing an expensive physical lab-experiment or computational simulation. We will first formally define the overall problem setting and then introduce novel adaptive experiment design algorithms to tackle this challenge.

The talk will be divided into two parts. First, we will survey a subset of probabilistic modeling approaches that we developed over high-dimensional combinatorial structures. We introduce a novel dictionary embeddings-based Gaussian Process (GP) model for high-dimensional binary/-categorical spaces (Deshwal et al. 2023). We will present theoretical analysis showing that the regret bounds for GP bandits trained on the embeddings is a function of a notion of orthogonality of the dictionary. To the best of our knowledge, this is the first result showing a direct connection between GP modeling of combinatorial structures and compressed sensing techniques. We will also show experiments results demonstrating its performance on both synthetic and real-world domains including maximum satisfiability, communications engineering, pest contamination control, and Auto ML. Subsequently, we plan to discuss our LADDER approach (Deshwal and Doppa 2021) that combines kernels over latent spaces of deep generative models with expert-designed structured kernels, demonstrating improved optimization performance on challenging molecule design benchmarks.

In the second part, we will discuss new information theoretic decision-making policies that we developed for real-world design optimization challenges. We will introduce the output space entropy (OSE) based decision-making framework and effective instantiations (Belakaria, Deshwal, and Doppa 2021) which are based on the principle of maximizing the information gain per unit resource cost about the optimal Pareto front. These information-theoretic policies significantly improved the state-of-the-art in multi-fidelity multi-objective design optimization problems (Deshwal, Simon, and Doppa 2021).

The impact of these methodologies is evident across multiple high-impact real-world applications. We plan to cover a representative subset of applications during the talk. For example, in nanoporous materials discovery, our approaches identified optimal covalent organic frameworks for methane storage and noble gas separation, evaluating only a small fraction of candidates. For electronic design automation, we enabled efficient design space exploration of 3D heterogeneous manycore systems. In additive manufacturing, our AI-driven approach improved the shape precision and porosity of 3D-printed organ models. We conclude with open challenges and future directions, including cost-aware optimization with black-box constraints, and exploration of grey-box optimization leveraging side information from experiments.

References

Belakaria, S.; Deshwal, A.; and Doppa, J. R. 2021. Output space entropy search framework for multi-objective Bayesian optimization. *Journal of Artificial Intelligence Research*, 72: 667–715.

Deshwal, A.; Ament, S.; Balandat, M.; Bakshy, E.; Doppa, J. R.; and Eriksson, D. 2023. Bayesian Optimization over High-Dimensional Combinatorial Spaces via Dictionary-based Embeddings. In *International Conference on Artificial Intelligence and Statistics*, 7021–7039. PMLR.

Deshwal, A.; and Doppa, J. 2021. Combining latent space and structured kernels for Bayesian optimization over combinatorial spaces. *Advances in Neural Information Processing Systems*, 34: 8185–8200.

Deshwal, A.; Simon, C. M.; and Doppa, J. R. 2021. Bayesian optimization of nanoporous materials. *Molecular Systems Design & Engineering*, 6(12): 1066–1086.