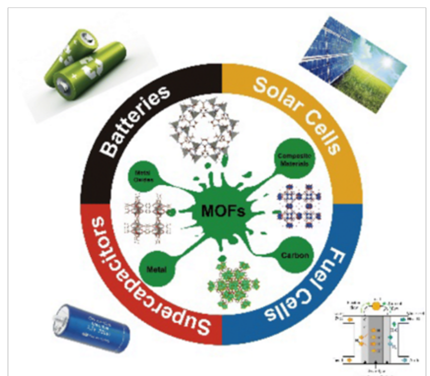
Bayesian Optimization over Hybrid Spaces

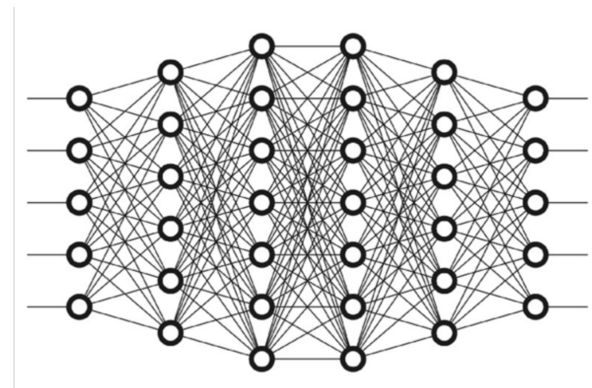
BO Over Hybrid Spaces: The Problem

- **Goal:** find optimized hybrid structures via expensive experiments
  - x = mixture of x_d (discrete) and x_c (continuous) variables

Microbiome design



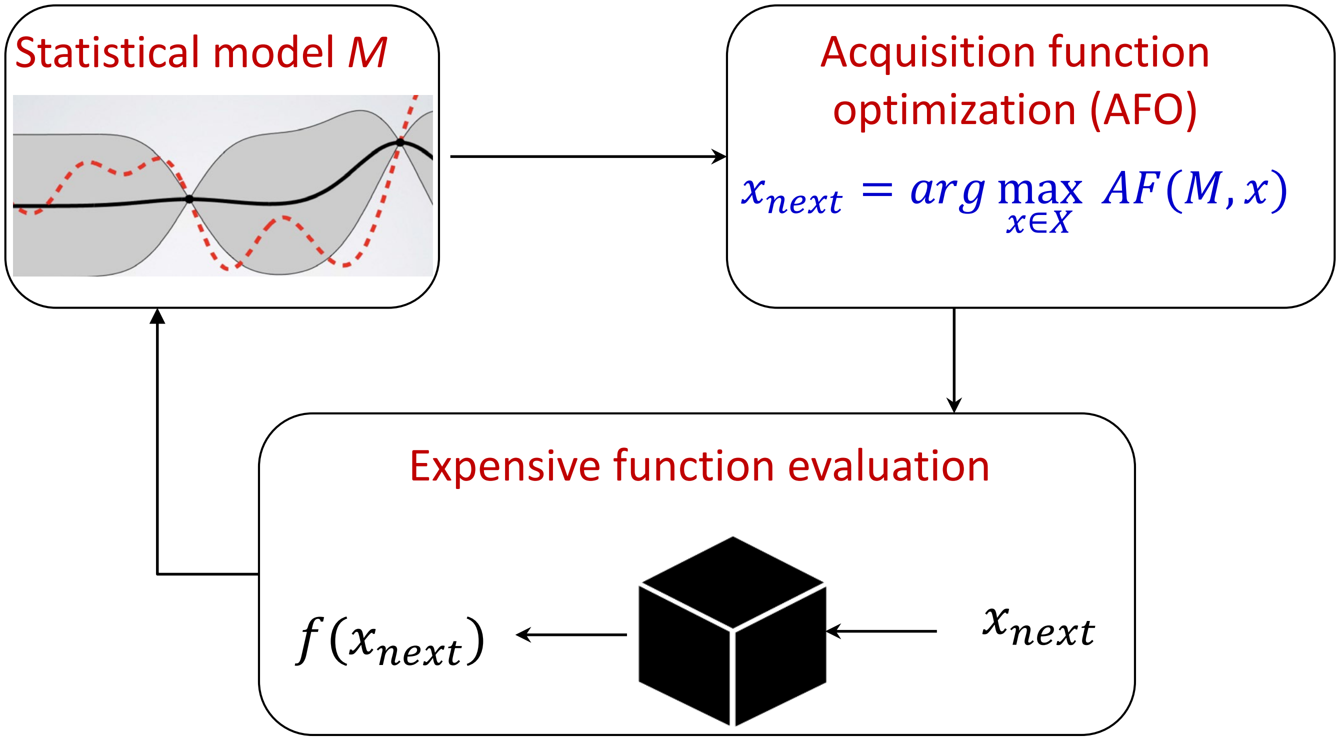
Material design



Hyper-parameter tuning / Auto ML

- Many other science, engineering, industrial applications

Hybrid BO: Technical Challenges



- Effective modeling over hybrid structures (capture complex interactions among discrete and continuous variables)
- Solving hard optimization problem over hybrid spaces to select next structure

Hybrid BO: Summary of Approaches

- Trade-off complexity of model and tractability of AFO

- Simple statistical models and tractable search for AFO
  - MiVaBO [Daxberger et al., 2019]

- Complex statistical models and heuristic search for AFO
  - SMAC [Hutter et al., 2011], HyBO [Deshwal et al., 2021] , BO-FM [Oh et al., 2021]

- Complex statistical models and tractable/accurate AFO
  - Reduction to continuous BO: GEBO [Ahn et al.,, 2022]

Hybrid BO: Summary of Approaches

- Trade-off complexity of model and tractability of AFO

- Simple statistical models and tractable search for AFO
  - MiVaBO [Daxberger et al., 2019]

- Complex statistical models and heuristic search for AFO
  - SMAC [Hutter et al., 2011], HyBO [Deshwal et al., 2021] , BO-FM [Oh et al., 2021]

- Complex statistical models and tractable/accurate AFO
  - Reduction to continuous BO: GEBO [Ahn et al.,, 2022]

MiVaBO [Daxberger et al., 2019]

- Linear surrogate model over binary structures
  - $f(x \in X) = \theta^T . \phi(x)$
  - $\phi(x)$ consists of continuous (random Fourier features), discrete (BOCS representation for binary variables), and mixed (products of all pairwise combinations) features

- Thompson sampling as acquisition function

- Alternating search for acquisition function optimization
  - Step 1: Search over continuous sub-space
  - Step 2: Search over discrete sub-space using output of Step #1
  - Repeat (if needed)

MiVaBO [Daxberger et al., 2019]

- Linear surrogate model over binary structures
  - $f(x \in X) = \theta^T . \phi(x)$
  - $\phi(x)$ consists of continuous (random Fourier features), discrete (BOCS representation for binary variables), and mixed (products of all pairwise combinations) features

May not be sufficient to capture desired dependencies

- Thompson

- Alternating search for acquisition function optimization
  - Step 1: Search over continuous sub-space
  - Step 2: Search over discrete sub-space using output of Step #1
  - Repeat (if needed)

MiVaBO [Daxberger et al., 2019]

- Linear surrogate model over binary structures
  - $f(x \in X) = \theta^T . \phi(x)$
  - $\phi(x)$ consists of continuous (random Fourier features), discrete (BOCS representation for binary variables), and mixed (products of all pairwise combinations) features

- Thompson sampling as acquisition function

- Alternating search for acquisition function optimization
  - Step 1: Searc r continuous sub-space
  - Step 2: Se pace using output of Step #1
  - Repeat (i

Can potentially get stuck in local optima

Hybrid BO: Summary of Approaches

- Trade-off complexity of model and tractability of AFO

- Simple statistical models and tractable search for AFO
  - MiVaBO [Daxberger et al., 2019]

- Complex statistical models and heuristic search for AFO
  - SMAC [Hutter et al., 2011], HyBO [Deshwal et al., 2021] , BO-FM [Oh et al., 2021]

- Complex statistical models and tractable/accurate AFO
  - Reduction to continuous BO: GEBO [Ahn et al.,, 2022]

SMAC Algorithm [Hutter et al., 2010, 2011]

- Random forest as surrogate model
  - works naturally for categorical/continuous variables
  - Prediction/Uncertainty (= empirical mean/variance over trees)

- Expected improvement as acquisition function

- Hand-designed local search with restarts for AFO

SMAC Algorithm [Hutter et al., 2010, 2011]

- Random forest as surrogate model
  - works naturally for categorical variables
  - Prediction/Unce inty (= empirical mean/variance over trees)

Uncertainty estimates
can be poor

- Expected impr n function

- Hand-designed local search with restarts for AFO

SMAC Algorithm [Hutter et al., 2010, 2011]

- Random forest as surrogate model
  - works naturally for categorical variables
  - Prediction/Uncertainty (= empirical mean/variance over trees)

- Expected improvement as acquisition function

- Hand-designed local search with restarts for AFO

Can potentially get
stuck in local optima

HyBO Algorithm [Deshwal et al., 2021]

- GP surrogate model with additive diffusion kernels

- Expected improvement as acquisition function

- Alternating search for acquisition function optimization
  - Step 1: Search over continuous sub-space
  - Step 2: Search over discrete sub-space using output of Step #1
  - Repeat (if needed)

HyBO Algorithm [Deshwal et al., 2021]

- GP surrogate model with additive diffusion kernels
  - Exploits the general recipe of additive kernels [Duvenaud et al., 2011]
  - Instantiation w/ discrete & continuous diffusion kernels
  - Bayesian treatment of the hyper-parameters

$$\mathcal{K}_{HYB} = \boxed{\sum_{p=1}^{m+n} \left(\boxed{\theta_p^2} \sum_{i_1, \cdots, i_p} \prod_{d=1}^{p} \boxed{k_{i_d}(x_{i_d}, x'_{i_d})} \right)}$$

HyBO Algorithm [Deshwal et al., 2021]

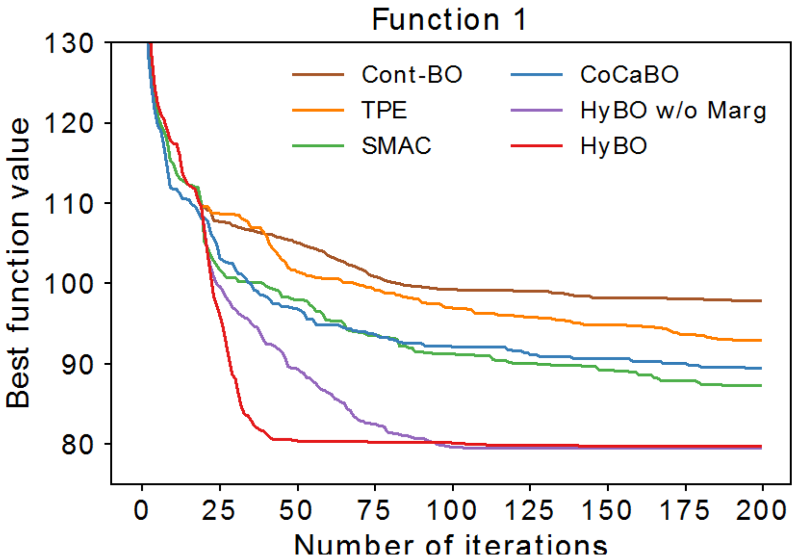
- GP surrogate model with additive diffusion kernels

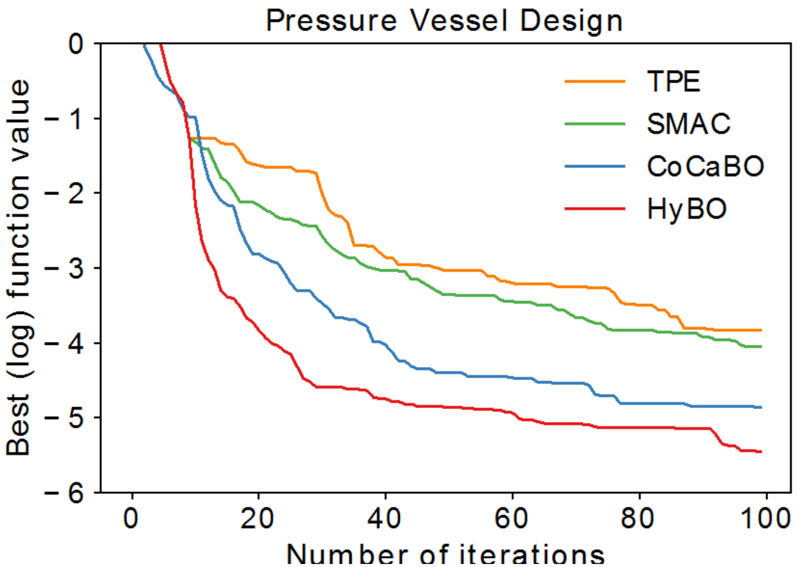
- Expected improvement as acquisition function

- Alternating search for acquisition function optimization
  - Step 1: Search over continuous sub-space
  - Step 2: Search over discrete sub-space using output of Step #1
  - Repeat (if needed)

Can potentially get stuck in local optima

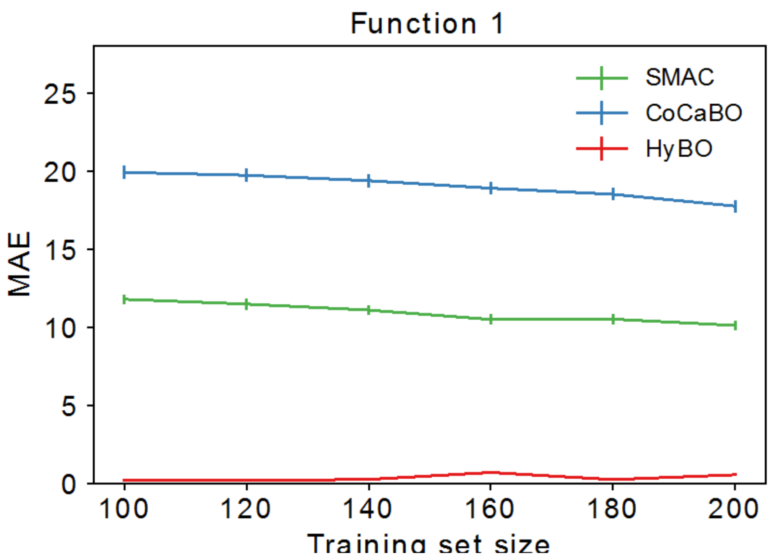
Hybrid BO: Experimental Results #1

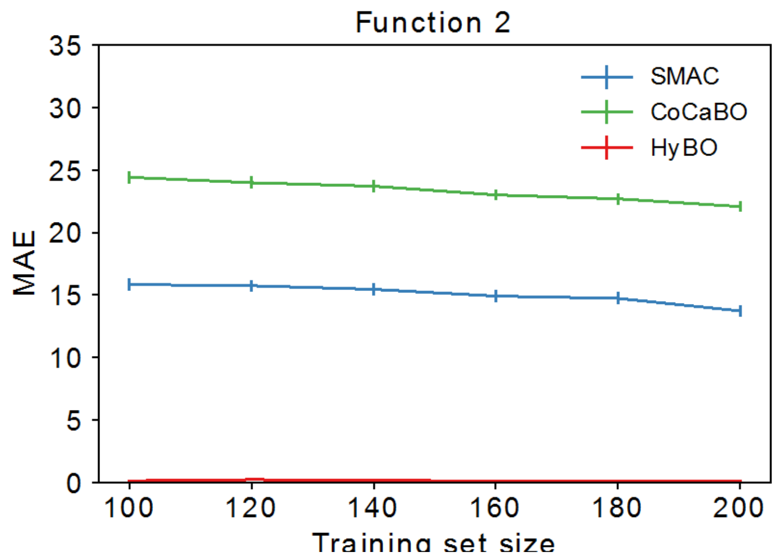




- HyBO performs significantly better than prior methods

Hybrid BO: Experimental Results #2





- HyBO's better BO performance is due to better surrogate model

BO-FM Algorithm [Oh et al., 2021]

- GP surrogate model with frequency modulation kernels

- Expected improvement as acquisition function

- Alternating search for acquisition function optimization
  - Step 1: Search over continuous sub-space
  - Step 2: Search over discrete sub-space using output of Step #1
  - Repeat (if needed)

BO-FM Algorithm [Oh et al., 2021]

- GP surrogate model with frequency modulation kernels

- Key idea: Generalize the COMBO kernel [Oh et al., 2019] by parametrizing via a function of continuous variables

$$\mathrm{K} = exp(-\beta L(G))$$

Remember the COMBO kernel

$$\mathrm{K} = U^T exp(-\beta \Sigma) U$$

$$\mathrm{K} = U^T f(\Sigma, X_c, X_{c'}) U$$

- Requirement on f for K to be a positive definite kernel
  - f should be positive definite w.r.t $X_c, X_{c'}$

Code and Software

- HyBO: https://github.com/aryandeshwal/HyBO
- SMAC: https://github.com/automl/SMAC3

Questions ?