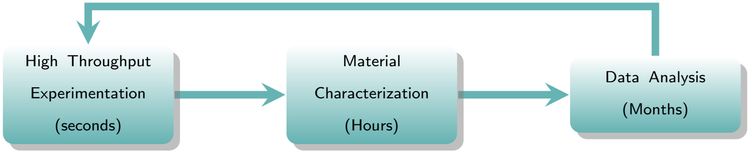
Accelerating Catalyst Discovery through Gaussian Processes and Active Learning

Kiran Vaddi [1] Olga Wodo [1] Krishna Rajan [1]

[1]Department of Material Design and Innovation
University at Buffalo, NY

December 2, 2019

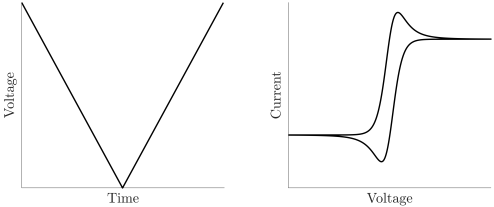
Introduction



- Cyclic Voltammetry is a type of catalyst characterization

Voltammetry– A primer

- Voltammetry is an electrochemical experiment with cyclic voltage load

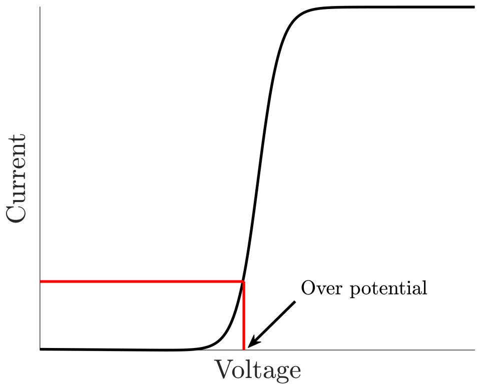


- It finds applications in electrochemical energy conversion, biomedicine, batteries, fuel cells, and glucose biosensors
- Has attracted much attention in analytical chemistry and evaluation of molecular catalysts

University at Buffalo
Department of Materials Design and Innovation
School of Engineering and Applied Sciences
College of Arts and Sciences

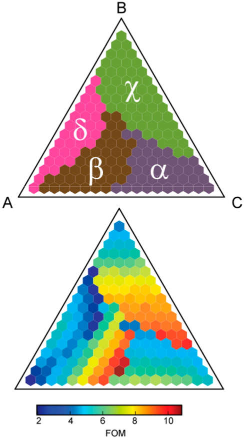
Machine Learning for Catalyst Discovery

- Two types of problems can be tackled
  - Features based search : down selection (Suram et.al [1])



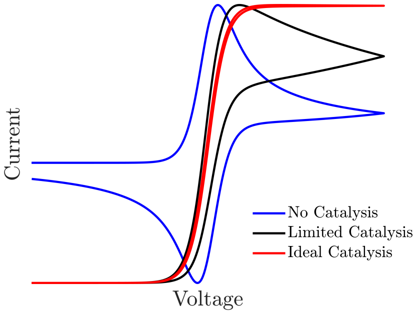
Machine Learning for Catalyst Discovery

- Two types of problems can be tackled
  - Features based search : down selection (Suram et.al [1])



Machine Learning for Catalyst Discovery

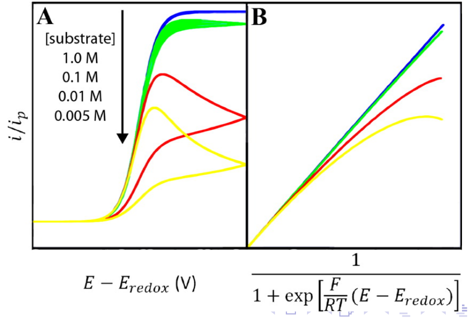
- Two types of problems can be tackled
  - Features based search : down selection (Suram et.al [1])
  - Shape based search : Knowledge extraction [2], Virtual screening



University at Buffalo
Department of Materials Design and Innovation
School of Engineering and Applied Sciences
College of Arts and Sciences

Machine Learning for Catalyst Discovery

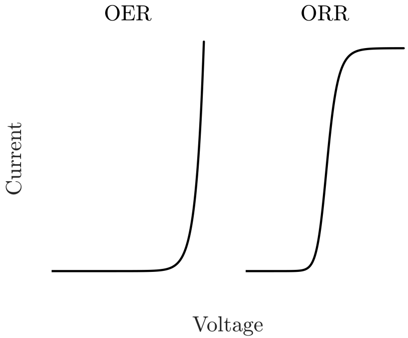
- Two types of problems can be tackled
  - Features based search : down selection (Suram et.al [1])
  - Shape based search : Knowledge extraction [2], Virtual screening



University at Buffalo
Department of Materials Design and Innovation
School of Engineering and Applied Sciences
College of Arts and Sciences

Machine Learning for Catalyst Discovery

- Two types of problems can be tackled
  - Features based search : down selection (Suram et.al [1])
  - Shape based search : Knowledge extraction [2], Virtual screening



Active Learning

- Suppose you start with a small dataset

$$\mathcal{D} = (\mathbf{K} \in \mathbb{R}^d, \mathbf{S} \in \{+1, -1\})$$

- Assume we can build belief about the value or label at unknown points $K^* \in \mathbb{R}^d$ for eg: k-NN probability
- Using $Pr(S^*|K^*, \mathcal{D})$, we can select next input $k^* \in K^*$ to get actual value of y^* eg: using an oracle
- For shape based search, we label the target shape using an "oracle" we built

Gaussian Processes $\mathcal{GP}$

- Start with a prior over functions $f, p(f) \sim \mathcal{GP}(\mu(.), k(., .))$
- We observe data $\mathcal{D} = (\mathbf{X}, \mathbf{y})$ corresponding to a CV response
- We judge our model $\mathcal{M}$, with parameter index θ using

$$\text{Model Evidence: } p(\mathbf{y}|\mathbf{X}, \mathcal{M}) = \int p(\mathbf{y}|\mathbf{X}, \theta, \mathcal{M}) p(\theta|\mathcal{M})\, \mathrm{d}\theta$$

- Represents probability of generating $\mathcal{D}$ under model $\mathcal{M}$

University at Buffalo
Department of Materials Design and Innovation
School of Engineering and Applied Sciences
College of Arts and Sciences

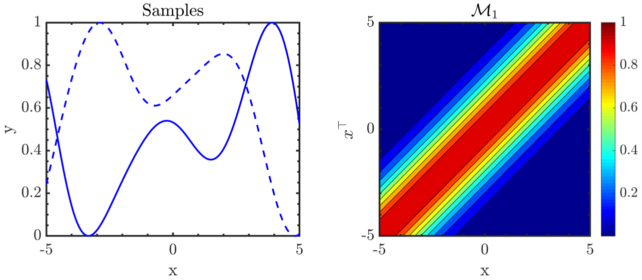
Bayesian Model Selection

- Suppose you have set of models $\{\mathcal{M}_i\}_{i=1}^{n}$ and computed model evidence for any given model $\mathcal{M}$ as $p(\mathbf{y}|\mathbf{X}, \mathcal{M})$
- We compare two models using model posterior–probability of model given the data– using

$$p(\mathcal{M}|\mathcal{D}) = \frac{p(\mathbf{y}|\mathbf{X}, \mathcal{M})p(\mathcal{M})}{p(\mathbf{y}, \mathbf{X})} = \frac{p(\mathbf{y}|\mathbf{X}, \mathcal{M})p(\mathcal{M})}{\sum_i p(\mathbf{y}|\mathbf{X}, \mathcal{M}_i)p(\mathcal{M}_i)}$$

Null model

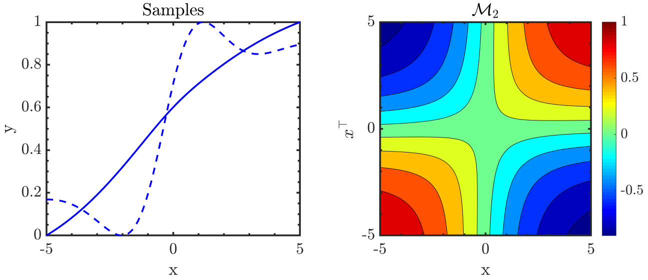
the continuous, smooth nature of any CV response using a model $\mathcal{M}_1$.



Catalytic Model

non-stationary nature of catalytic response using model $\mathcal{M}_2$

.



Data Generation

- Simulated data by solving EC mechanism kinetic equations [3]

$$\mathrm{Cat} + \mathrm{e} \longleftrightarrow \mathrm{Cat}^{+} \tag{R1}$$

$$\mathrm{Cat}^{+} + \mathrm{A} \longrightarrow \mathrm{Cat} + \mathrm{B} \tag{R2}$$

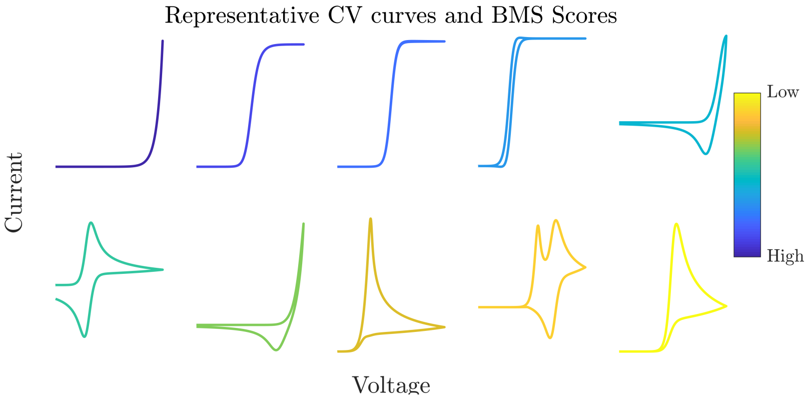
- A total of $\approx 17 \times 10^3$ curves in a 6-dimensional search space **K**

S.No.	Parameter	Range
1	$\log C^0_{cat}$	[-2,3]
2	$\log C^0_A$	[-2,3]
3	$\log \nu$	[-2,4]
4	E^0	[-0.4,0.4]
5	$\log k_f$	[-1,6]
6	$\log k_s$	[-1,6]

Table: Blue are material properties, red are experimental settings

Bayesian Model Selection Scores

posterior log-likelihood



University at Buffalo
Department of Materials Design and Innovation
School of Engineering and Applied Sciences
College of Arts and Sciences

Active Search for S-shape CV Curve

Find parameters of EC mechanism that have S-shape CV response

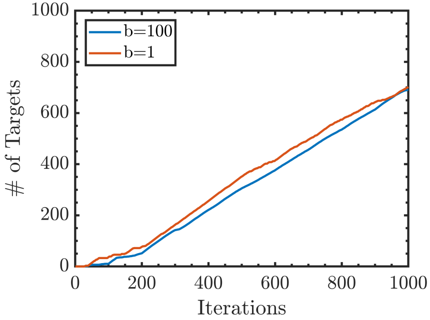


Figure: Efficient Non-myopic Search for S-shapes with a budget of 1000 experiments to find S-shapes out of 17×10^3 CV shapes

Active Area Search for S-shape CV Curve

Find locations in 2D grid with Target Shapes

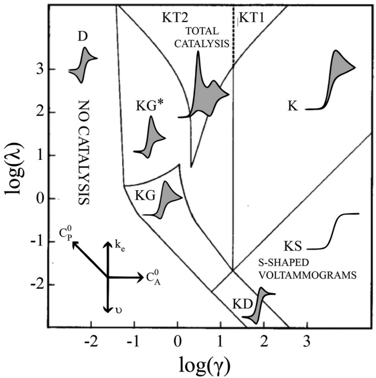


Figure: Find target location with only 10% of total grid experiments

Active Area Search for S-shape CV Curve

Find locations in 2D grid with Target Shapes

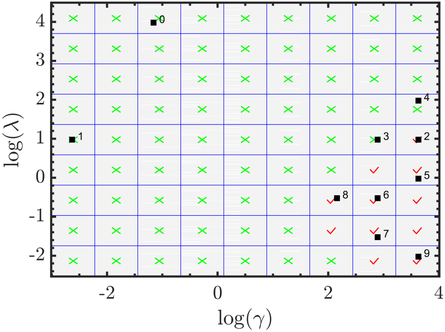


Figure: Find target location with only 10% of total grid experiments

Summary and Future steps

- Introduced an shaped based search for CV experiments of catalyst discovery

Summary and Future steps

- Introduced an shaped based search for CV experiments of catalyst discovery
- Our method has potential to accelerate the knowledge extraction, down sampling and virtual screening catalysts based on CV curves

Summary and Future steps

- Introduced an shaped based search for CV experiments of catalyst discovery
- Our method has potential to accelerate the knowledge extraction, down sampling and virtual screening catalysts based on CV curves
- A probabilistic oracle that has potential to actively search for bi-functional catalysts for fuel cells

References I

- S. K. Suram, J. A. Haber, J. Jin, and J. M. Gregoire, "Generating information-rich high-throughput experimental materials genomes using functional clustering via multitree genetic programming and information theory," *ACS combinatorial science*, vol. 17, no. 4, pp. 224–233, 2015.
- E. S. Rountree, B. D. McCarthy, T. T. Eisenhart, and J. L. Dempsey, *Evaluation of homogeneous electrocatalysts by cyclic voltammetry*, 2014.
- G. Kennedy, *Monash electrochemistry simulator (mecsim)*, 2015.

Thank You!

Questions?