A2I2

APPLIED ARTIFICIAL INTELLIGENCE INSTITUTE

DEAKIN UNIVERSITY

Human-AI Collaborative Bayesian Optimisation

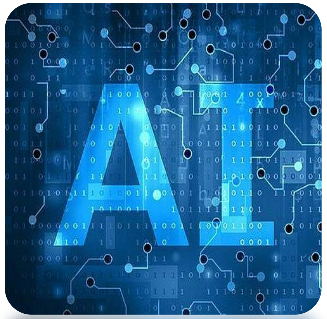
In 36th International Conference on
Neural **I**nformation **P**rocessing **S**ystems **(NeurIPS)** 2022

Authors: **Arun Kumar A V,** Santu Rana, Alistair Shilton, Svetha Venkatesh

A²I² APPLIED ARTIFICIAL INTELLIGENCE INSTITUTE

DEAKIN UNIVERSITY

Introduction

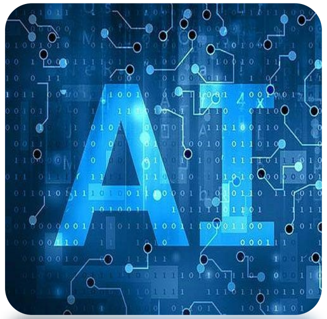


- Applications of AI systems – omnipresent but Sub-optimal

How do we improve the performance in such systems??

Image courtesy: https://www.google.com/imghp?hl=en

Introduction



- Applications of AI systems – omnipresent but Sub-optimal

How do we improve the performance in such systems??

- Human-AI collaborative systems
- Harness the complementary strengths of AI and humans
  - Quantitative abilities of AI systems
  - Cognitive capabilities of humans

Introduction

Human-AI Collaborative Systems

- Bayesian Optimisation (BO) based methods?
  - Assumes an expert performing the experimentations
  - Incorporation of the expert knowledge

Introduction

Human-AI Collaborative Systems

- Bayesian Optimisation (BO) based methods?
  - Assumes an expert performing the experimentations
  - Incorporation of the expert knowledge

- Expert knowledge in Bayesian optimisation
  - Input space characteristics – smoothness
  - Trends – monotonicity, unimodality

Introduction

Human-AI Collaborative Systems

- Bayesian Optimisation (BO) based methods?
  - Assumes an expert performing the experimentations
  - Incorporation of the expert knowledge

- Expert knowledge in Bayesian optimisation
  - Input space characteristics – smoothness
  - Trends – monotonicity, unimodality

knowledge evolves over time

Approach

How to incorporate such fluid knowledge in Bayesian optimisation??

Human-AI Collaborative Bayesian Optimisation

- **H**uman-**A**I **T**eaming (**HAT**) Bayesian optimisation
  - Correction of the current recommendation
  - Specification of good regions and bad regions
- These tasks do not add a significant overhead on the expert
- Extra expert information incorporated into the model selection

Framework

Human-AI Teaming – Rectifying Recommendation (HAT-RR)

- Correct previously suggested observation $\mathcal{X}^A$
- Expert observations $\mathcal{X}^E$
  - Reflects full or partial knowledge of the kernel
- Model selection process

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \log \mathcal{L}$$

Framework

Human-AI Teaming – Rectifying Recommendation (HAT-RR)

- Correct previously suggested observation $\mathcal{X}^A$
- Expert observations $\mathcal{X}^E$
  - Reflects full or partial knowledge of the kernel

- Model selection process

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \log \mathcal{L}$$

such that $u_{\text{GP-UCB}}(\mathbf{x}_i^E | \mathcal{D}_{i-1}) > u_{\text{GP-UCB}}(\mathbf{x}_i^A | \mathcal{D}_{i-1})$

$$\forall\, \mathbf{x}_i^A \in \mathcal{X}^A,\ \mathbf{x}_i^E \in \mathcal{X}^E,\ i \in H$$

$$\mathcal{D} = \{(\mathbf{x}, y = f(\mathbf{x}))\}$$

Framework

Human-AI Teaming – Distance Maximisation (HAT-DM)

- Expert inputs good and bad regions – reflection of expert's acquisition function

- A bi-objective optimisation problem
  1. Maximise log-likelihood $\Theta^* = \underset{\Theta}{\operatorname{argmax}} \log \mathcal{L}$
  2. Define a new co-objective with constraints

Framework

Human-AI Teaming – Distance Maximisation (HAT-DM)

- Expert inputs good and bad regions – reflection of expert's acquisition function

- A bi-objective optimisation problem
  1. Maximise log-likelihood $\Theta^* = \underset{\Theta}{\operatorname{argmax}} \log \mathcal{L}$
  2. Define a new co-objective with constraints

$$\Theta^+ = \underset{start \leftarrow \Theta^*}{\operatorname{argmax}} \sum_{i \in H} \left(u_{\text{GP-UCB}}(\mathbf{x}_i^g | \mathcal{D}_{i-1}) - u_{\text{GP-UCB}}(\mathbf{x}_i^b | \mathcal{D}_{i-1}) \right)$$

$$\text{such that} \quad (\log \mathcal{L})_{\Theta^+} \geq \Delta (\log \mathcal{L})_{\Theta^*}$$

Compromising Threshold

Experimental Results

Variants for Comparisons

HAT-RR

HAT-RR (KFO)

HAT-DM (KFO)

STD-BO

STD-MOD

Squared Exponential Kernel

KFO Tuned Kernel

Squared Exponential Kernel

Experimental Results

Variants for Comparisons

HAT-RR

HAT-RR (KFO)

HAT-DM (KFO)

STD-BO

STD-MOD

Squared Exponential Kernel

KFO Tuned Kernel

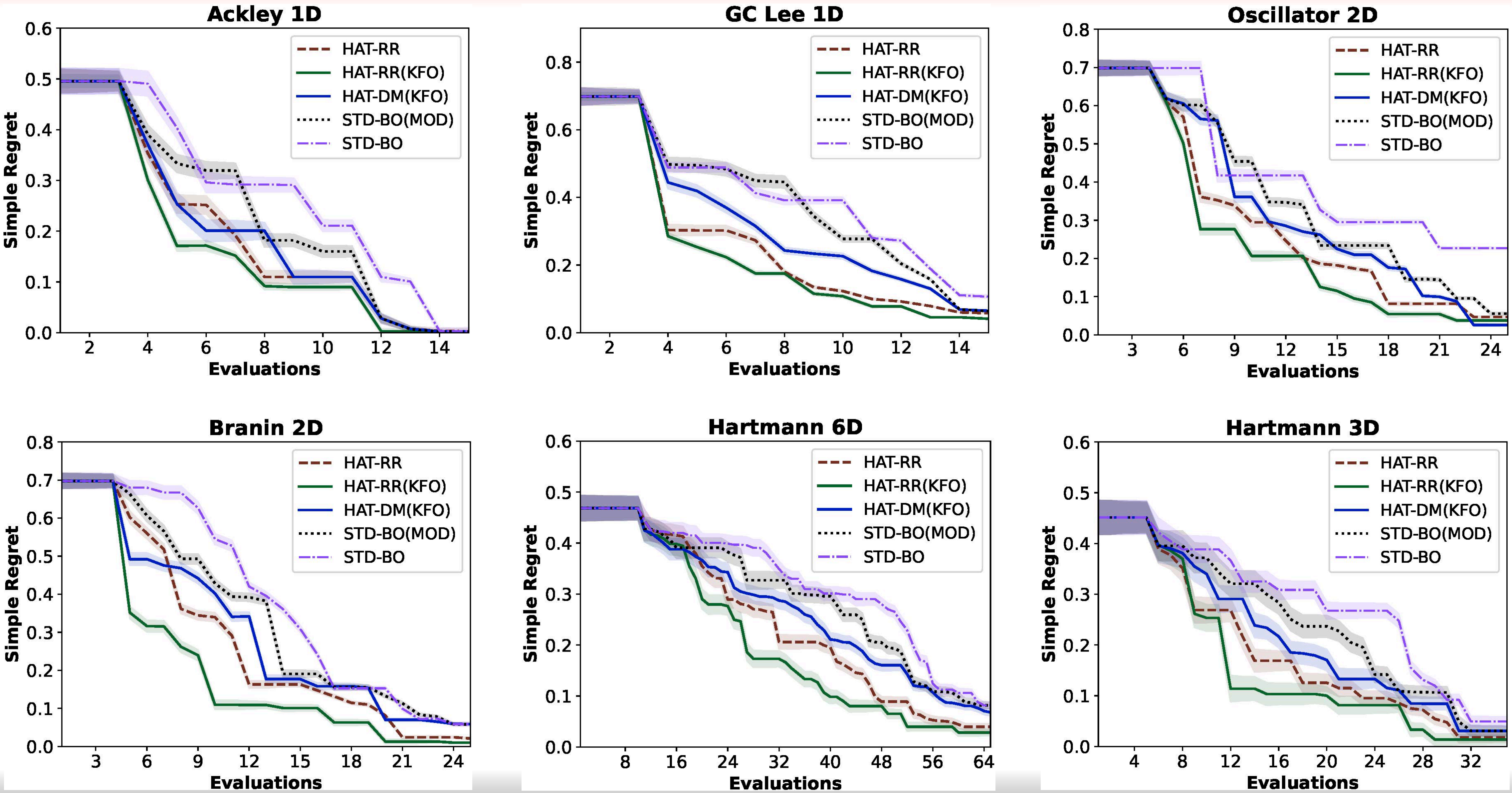
Squared Exponential Kernel

Synthetic Experiments

- Multi-dimensional synthetic benchmark functions
  - Ackley 1D, Gramacy & Lee 1D, Branin 2D, Oscillator 2D, Hartmann 3D and 5D
- Plot simple regret $\hat{r}_t^+$ after $10 \times d + 5$ iterations

$$\hat{r}_t^+ = f(\mathbf{x}^*) - \max_{\mathbf{x}_t \in \mathcal{D}_{1:t}} f(\mathbf{x}_t)$$

Experimental Results



Experimental Results

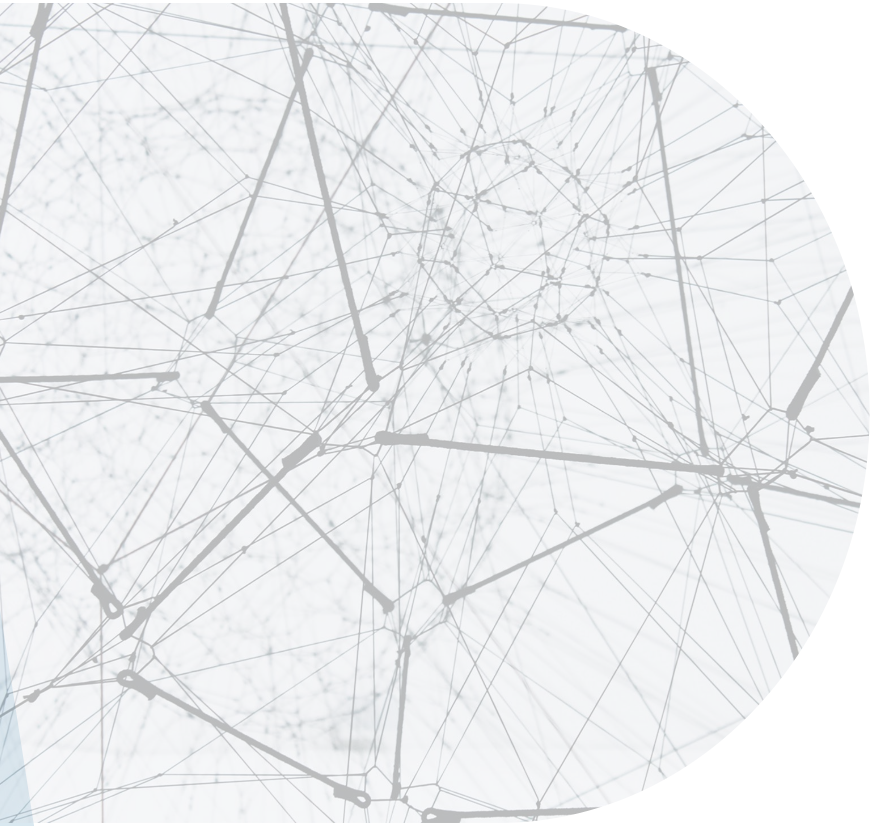
Real-world Experiments

- Multi-dimensional datasets from UCI repository: $80/20$ Train/Test splits
- C-SVM with RBF kernel to minimise classification error.

Instances	STD-BO	STD-BO (MOD)	HAT-RR	HAT-RR (KFO)	HAT-DM (KFO)
WDBC	1.5 ± 0.5	1.24 ± 0.2	*0.98±0.2*	$\mathbf{0.60 \pm 0.75}$	0.95 ± 0.45
Ionosphere	8.9 ± 0.2	6.02 ± 0.6	***5.15±0.4***	5.25 ± 0.10	6.21 ± 0.81
Sonar	8.21 ± 0.9	8.44 ± 0.2	*6.46±0.3*	$\mathbf{6.11 \pm 0.37}$	6.84 ± 0.92
Heart	11.7 ± 0.9	11.10 ± 1.4	*10.2±0.3*	$\mathbf{10.25 \pm 0.41}$	10.97 ± 0.75
Seeds	3.3 ± 0.4	2.58 ± 0.8	*2.8±0.14*	$\mathbf{2.51 \pm 0.65}$	2.63 ± 0.47
Wine	**0**	**0**	**0**	**0**	**0**
Credit	18.1 ± 1.3	14.52 ± 0.6	*12.7±1.3*	$\mathbf{12.42 \pm 0.60}$	12.95 ± 0.78
Biodeg	16.8 ± 1.9	15.05 ± 0.6	***13.4±0.2***	13.88 ± 0.42	14.09 ± 1.83
Car	1.9 ± 0.5	0.39 ± 0.7	*0.21±0.1*	0.35 ± 0.17	$\mathbf{0.30 \pm 0.53}$
Ecoli	2.1 ± 0.6	2.01 ± 0.3	*1.67±0.2*	$\mathbf{1.21 \pm 0.38}$	1.94 ± 0.64

Questions?

Email: aanjanapuravenk@deakin.edu.au



Thank You!!