Accelerating Model Driven Discovery with ML
—
Presented by Sekou L. Remy

Work done in solidarity with:
Irene Wanyana, Makerere University
Joyce Nakatumba-Nabende, Makerere University
Susan Kizito, Makerere University
Balla Diop Ngom, Université Cheikh Anta Diop
Ernest Mwebaze, Sunbird AI
Barbara Han, Cary Institute of Ecosystem Studies
Fadel Thior, IBM
Rose Yvette Essomba, IBM
Stephanie Muller, IBM
Cliff Kerr, Institute for Disease Modeling
...

and many, many friends!

Developed as a workshop proposal for ICLR 2023
3000 Researchers
14 Locations
6 Continents

- 6 Nobel Laureates
- 10 Medals of Technology
- 5 National Medals of Science
- 6 Turing Awards
In 2020, there were \(~247\) million cases of malaria, \(95\%\) of these cases were in Africa, \(~625,000\) malaria deaths, and malaria also increases the risk of death from other diseases.

Approximately \(3.5B\) USD was estimated for elimination and control efforts.


Malaria can be treated and cured with anti-malarial drugs (i.e. interventions exist)
Malaria can be prevented with effective malaria control.
In 2018, there were ~219 million cases of malaria, 95% of these cases were in Africa, ~450,000 malaria deaths, and malaria also increases the risk of death from other diseases.

Approximately $4.5B USD was estimated for elimination and control efforts.


Malaria can be treated and cured with anti-malarial drugs (i.e. interventions exist)
Malaria can be prevented with effective malaria control.
Starting with a Significant Global Health Challenge

**Pre-September Questions 2018**

**Q1** Are there intervention packages that are not being explored today because of lack of tooling?

**Q2** Are there overlooked approaches to generate insight from multiple models, and accommodate the inherent uncertainty therein in a principled manner?

**Q3** How should model output from multiple models be presented to a policy influencer to enable the best policies to be selected?

**Post September Targets**

**T1** Develop and deploy a dashboard which can harness existing resources (malaria models and data) in a manner to support the selection of intervention policies for Uganda.

**T2** Recommend useful intervention packages which would not have been considered by informed policy makers today.

**T3** Identify the context under which additional stratification would be useful for policy decisions in Uganda. Focus is the best overall outcomes from a prevalence, cost, and execution complexity perspective.

---

**New Goal**

By 2025, drive malaria parasite prevalence in all 134 districts of Uganda to 5% or less (from 9%).

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0187680
What you do, and when you do it matters

Crude prevalence of malaria infection stratified by survey round and larviciding status.

Supporting Decisions with ML

O. Bent, et. al. *Novel exploration techniques (NETs) for malaria policy interventions*. In Thirty-Second AAAI Conference on Artificial Intelligence (IAAI 2018)

S.L. Remy, et. al. *Reshaping the use of digital tools to fight malaria*. Multilateral Initiative on Malaria - Workshop on Using Digital Tools to Strengthen the Malaria Supply Chain (MIM 2018)

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0187680
Key Design Change

Environments define:
• Available actions
• Available observations
• Action goodness

Meta-Models define:
• Which parameters can be changed
• What data can be observed

Finding Intervention Programs Which Work Well

Tuning/Calibrating Models for Specific Use

Reward:

\[ C_{DA} = \frac{C_{int} + HSC_{int} - HSC_{no int}}{DALY_{int} - DALY_{no int}} \]
Building Trust With Sample Data

Comparison of 3 Algorithms

COVID MODEL

RL with “COVID MDPs”
Pre-Pandemic Engagement with the ML Community

Humanities Track - KDD Cup 2019

Feedback Phase
Begins: April 15th 2019
Ends: June 29th 2019

Check Phase
Begins: June 30th 2019
Ends: July 6th 2019

Verification Phase
Begins: July 7th 2019
Ends: July 19th 2019

IBM Research

ZINDI

IBM Malaria Challenge
Knowledge

Use reinforcement learning to determine the optimal combination of interventions to control the transmission of malaria

47 data scientists enrolled, 21 on the leaderboard

25 August–30 August 2019 12:59
“I feel like my research topics are shaped by the current benchmarks which are limited in scope and it’s hard to evaluate ideas that are different...

Currently most of the benchmarks are focusing on robotics and games, and going outside of that to actual sequential decision-making problems that would have an impact would be really good.”
Peri-Pandemic Engagement with the ML Community

### Setup

```bash
conda create -n iclr23env python=3.10
conda activate iclr23env
pip install git+https://github.com/IBM/ushiriki-policy-engine-library
```

### Usage

```python
import gym
import ushiriki_policy_engine_library
from gym.wrappers import normalize

e = gym.make('ICLR23Workshop-env1-v0',
             model_name="http://.../ibmracovid19modelv1/",
             driver_data="https://gist.github.com/..",
             numdays=28, duration=280,
             apply_api_compatibility=False)
env = normalize.NormalizeObservation(e)

model = TD3(env, replay_buffer_size = 5e5, hidden_dim = 512)
```

Your algo!
How can ML help a public health professional infer the transmission rate of a disease in a particular location?

Machine learning (ML) can help a public health professional infer the transmission rate of a disease in a particular location by analyzing data related to the disease and the location. Here are some ways ML can be useful:

1. Predictive modeling: ML models...

You are correct that the concerns I mentioned earlier may not apply if ML models are indirectly used to optimize the parameters of known epidemiological models, rather than directly inferring the transmission rates. Indirect use of ML in this context can refer to the application of ML algorithms to optimize model parameters and improve the accuracy of existing epidemiological models, such as SIR (Susceptible-Infectious-Recovered) or SEIR (Susceptible-Exposed-Infectious-Recovered) models...
How can I infer transmission rate from case data for a disease?  

In order to infer the transmission rate from case data for a disease, you can use mathematical models that describe the transmission dynamics of the disease. One commonly used model is the SIR (Susceptible-Infectious-Recovered) model, which divides the population into three groups: those who are susceptible to the disease, those who are infected and can transmit the disease, and those who have recovered and are no longer susceptible.

The transmission rate (also called the reproduction number or R0) is the average number of secondary infections that result from each infected person during the early stages of an outbreak, when the majority of the population is still susceptible. This value can be estimated by analyzing the case data and fitting the SIR model to the observed data.

Is the transmission rate and Reproduction number truly the same thing? I thought one was a probability... 

The transmission rate and the reproduction number are related concepts, but they are not exactly the same thing. ...
Accelerating Model Driven Discovery with ML

Presented by Sekou L. Remy

Work done in solidarity with:
Irene Wanyana, Makerere University
Joyce Nakatumba-Nabende, Makerere University
Susan Kizito, Makerere University
Balla Diop Ngom, Université Cheikh Anta Diop
Ernest Mwebaze, Sunbird AI
Barbara Han, Cary Institute of Ecosystem Studies
Fadel Thior, IBM
Rose Yvette Essomba, IBM
Stephanie Muller, IBM
Cliff Kerr, Institute for Disease Modeling
...

and friends!
Accelerated discovery for neglected tropical diseases (NTDs)
IBM is developing partnerships with non-profit and government organizations on new discovery accelerators for NTDs.
Two phases of exploration

- Accelerated design of new drug molecules
- Accelerated generation of intervention program
Our Goal for 2025
IBM is developing tech with the goal of overcoming bottlenecks, accelerating discovery

10x

increase the speed of discovering more effective drugs for NTDs with partners
Questions

Do you think this can work?

What else do you think we need?

What could go wrong/right?

Who do you thing might want to join us in the journey?
• Or whose journey can we join?
Data Sources

https://ibm.github.io/wntrac/
https://www.worldometers.info/coronavirus/
https://coronavirus.jhu.edu/map.html
Motivation
Artificial Intelligence
Mimic Humans

Machine Learning
Learn with Experience

Reinforcement Learning
Interactively Learn from Data

Supervised Learning
Learn from Labeled Data

Unsupervised Learning
Learn from Unlabeled Data

... Deep Learning?
... Data Science?

Usage

<table>
<thead>
<tr>
<th>Year</th>
<th>DeepMind Control Suite</th>
<th>Atari 100k</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>2020</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>2021</td>
<td>100</td>
<td>75</td>
</tr>
<tr>
<td>2022</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>2023</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>