Public Policy and Deep Reinforcement Learning on AWS

Emily Webber | Machine Learning Specialist at Amazon Web Services | To-be-open-sourced research project
Public Policy Has Unique Challenges

- Structural Inefficiency
- Lack of single goal
- Synthesize Information
- Leadership Turnover
What if we used machine learning to optimize public policy?

- Personalized
- Collaborative
- Transparent
- Decades of Economic Data
- Normalized Policy Data
- Reinforcement Learning
Data-Driven Public Policy Analysis Is Not New

- Causal Inference
- Counterfactual analysis
- Intuitively, what would have happened if the policy (or, treatment) had not been applied?
- Can we convince ourselves that the two groups were nearly identical otherwise?

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment: Illinois</td>
<td>100</td>
<td>250</td>
</tr>
<tr>
<td>Control: New York</td>
<td>100</td>
<td>150</td>
</tr>
</tbody>
</table>

Did the treatment *cause* this difference?

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_T X_T + \ldots + \epsilon \]
Learning Theory Fundamentals

- Machine Learning
  - Use Case
  - Model
  - Data

- Reinforcement Learning
  - Use Case
  - Actions
  - Rewards
Mathematically speaking

Bellman Equation for Reinforcement Learning

\[
U(s) = R(s) + \gamma \max_a \sum_{s'} T(s, a, s') U(s')
\]

- **Utility per state, or value**
- **Discount factor**
- **Transition value**
- **Available action**
- **Current state**
- **Adjacent state, iterable**
- **Recursive call on utility function**

- **Reward per state, a real number**
- **For each possible adjacent state**
Our reward function

- Ask, are they similar? T-test
- Use logical reasoning
- Eventually, scale with another ML model using data labelled by experts

A deep learning model maps the economic variables to a policy suggestion

The simulator picks treatment and control states and runs a regression on historical data

We use the estimated effect of the policy as our reward signal, scaled by validity of the experiment

Causal Inference

Reinforcement Learning Policy Estimation

"Pareto"
But how do we pick the right way to optimize?
## Philosophical Foundations

<table>
<thead>
<tr>
<th>Utilitarianism</th>
<th>Egalitarianism</th>
<th>Kantian Rights</th>
<th>Libertarianism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Value</td>
<td>Equality</td>
<td>Universal Rights</td>
<td>Freedom</td>
</tr>
</tbody>
</table>

### Pareto Improvements

Improve at least one person, without making anyone worse off
There is no single best optimization strategy

What we can do is use data to automatically suggest policies based on user-defined preferences.
What do you want to see in public policy?

- Personal Freedom
- Equality of outcomes
- Less crime
- Access to education
- Access to social services
- Less waste
- Equality of opportunity
- Less traffic
- Better health care

What do you think impacts crime the most?

In my neighborhood, people commit crimes because there are no jobs here.

Given your views, we recommend evaluating:

- Outcomes
  - Crime
  - Income

- Indicators
  - Employment
  - Savings

Confirm?
These policies are impacting you today.

Here’s how to engage your elected officials

Please correct bill 789, it is lowering my income

Email

Your policy recommendations

Reduce taxation

Continue investment

Build more highways

Reducing income 13.45

Creating jobs 42.66

Increasing traffic .05
What if we could step into someone else’s shoes?
Your policy recommendations

- Bill 789
- Reduce taxation
- Bill 238
- Continue investment
- Bill 121
- Build more highways
- Personal Freedom Increase

Another point of view

- Bill 789
- Increase taxation
- Bill 238
- Continue investment
- Bill 121
- More Public Transit
- Increase Equality Overall
Technically speaking:

```python
for ism in philosophical_frameworks:
    utility = define_utility(ism)
    data = update_data(utility)
    model = get_pareto(data)
```
How should we handle air traffic delays?
<table>
<thead>
<tr>
<th><strong>Utilitarianism</strong></th>
<th><strong>Egalitarianism</strong></th>
<th><strong>Kantian Rights</strong></th>
<th><strong>Libertarianism</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Do whatever increases overall utility</td>
<td>Do what increases overall equality</td>
<td>Uphold human rights</td>
<td>Preserve Freedom</td>
</tr>
<tr>
<td>• Different people value timeliness differently</td>
<td>• Don’t prioritize airline status travelers</td>
<td>• Uphold the human sanctity of travelers</td>
<td>• Let people pick for themselves</td>
</tr>
<tr>
<td>• Need multiple ways of defining utility for diverse stakeholders</td>
<td>• Don’t let people pay more for perks</td>
<td>• Provide food, lodging, respectful notice</td>
<td>• Don’t automatically make decisions for travelers</td>
</tr>
<tr>
<td>• Use testing and surveys to get a numerical estimate for how different people value certain outcomes</td>
<td>• Don’t do special favors</td>
<td>• Make reasonable attempts to avoid delays</td>
<td>• Let travelers switch across airliners</td>
</tr>
<tr>
<td></td>
<td>• Treat each traveler, airliner, and airport the same</td>
<td></td>
<td>• Ensure freedom of airliners and airports</td>
</tr>
</tbody>
</table>
There is fundamental overlap between the philosophical frameworks.

This overlap can be scaled by reward functions.
There is no single right answer

We need a computational system that can:

- Synthesize different points of view
- Weight these based on criteria, like population size
- Be transparent, collaborative, timely
- Change with the times

To efficiently support existing governing bodies
Thank you!

effective-policies@amazon.com ⇐ email me to collaborate!