Ambulance Routing with Reinforcement Learning

Carrie Rucker

March 2021

Problem Overview

The agent's goal is to position ambulances in the environment space so as to minimize both the distance traveled to those locations and the distance traveled by the nearest ambulance to an incoming call

Two environments have been implemented:

- **Metric:** ambulances can be positioned and calls can arrive anywhere along a line from 0 to 1
  Distances between locations are calculated with the ℓ_1 distance
- **Graph:** ambulances can be positioned and calls can arrive at any node in a given graph
  Distances between nodes are calculated using the shortest path

Observation and Action Spaces

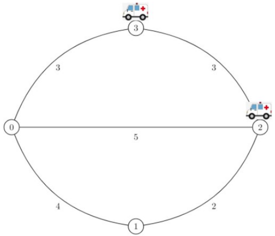
Metric Environment:

- Observation Space: $S = X^k$ where $X = [0, 1]$ and there are k ambulances
- Action Space: $A = X^k$ where $X = [0, 1]$ and there are k ambulances

If $k = 2$ the ambulances could be positioned at 0.4 and 0.6

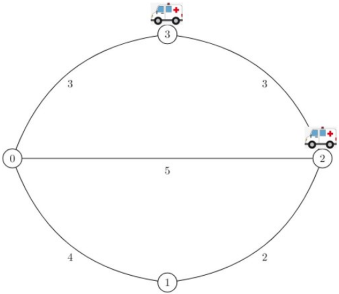
Graph Environment:

- Observation Space: $S = V^k$, where V is all the nodes in the graph and there are k ambulances
- Action Space: $A = V^k$, where V is all the nodes in the graph and there are k ambulances



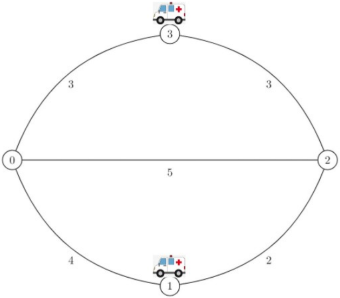
If $k = 2$ the ambulances could be positioned at nodes 2 and 3

Graph Environment Example



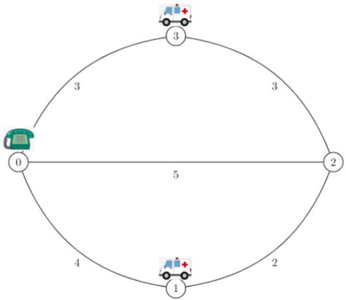
Suppose there are two ambulances in the environment and they start at nodes 3 and 2

Graph Environment Example cont.



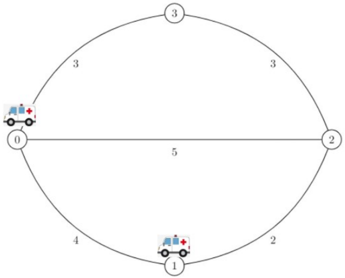
The agent can then choose new nodes for the ambulances. In this case, the agent chooses to leave one ambulance at node 3 and move the other ambulance to node 1. The total distance travelled from the original state S_O to the action A is $D(S_O, A) = 2$

Graph Environment Example cont.



A call arrives at a node randomly sampled from the arrival distribution. The nodes in this example is 0

Graph Environment Example cont.



The closest ambulance travels to the call. The distance travelled from the action A to the new state S_N in this example is $D(A, S_N) = 3$

The reward returned by this iteration is

$$-1 * [\alpha * (D(S_O, A)) + (1 - \alpha) * D(A, S_N)]$$

where α is a parameter controlling the tradeoff between the cost of travel between calls and the cost of travel to respond to a call

Stable Agent (Graph and Metric)

The stable agent never moves the ambulances between calls.

The stable agent is optimal when $\alpha = 1$ and there is no cost to travel to respond to a call.

Median Agent (Metric)

The median agent takes the past call arrival data sorted by location and splits it into k quantiles where k is the number of ambulances. It will use the center data point in each quantile as the location for an ambulance.

The median agent performs best when $\alpha = 0$ and there is no cost to travel between calls.

Median Agent (Graph)

The median agent finds the k nodes from which an ambulance would have to travel the least distance to respond to all of the past calls, where k is the number of ambulances.

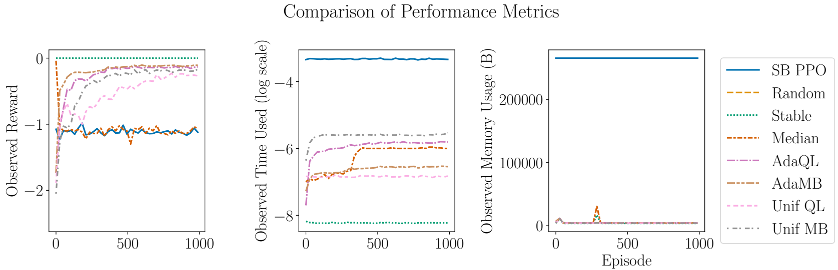
The median agent performs best when $\alpha = 0$ and there is no cost to travel between calls.

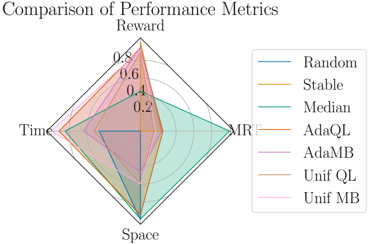
Mode Agent (Graph)

The mode agent finds the k nodes where the most calls have appeared in the past call data where k is the number of ambulances, and stations the ambulances at those locations.

The mode agent performs better when $\alpha = 0$ and there is no cost to travel between calls.

Metric Agent Performance w/ Beta distribution ($\alpha = 1$)

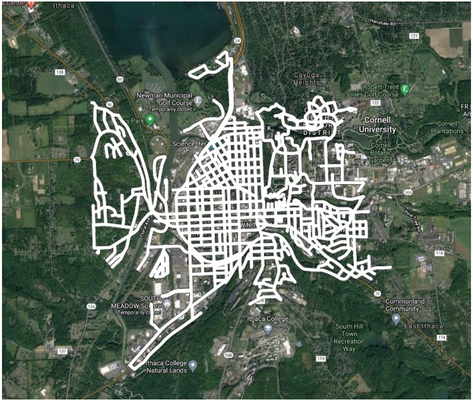




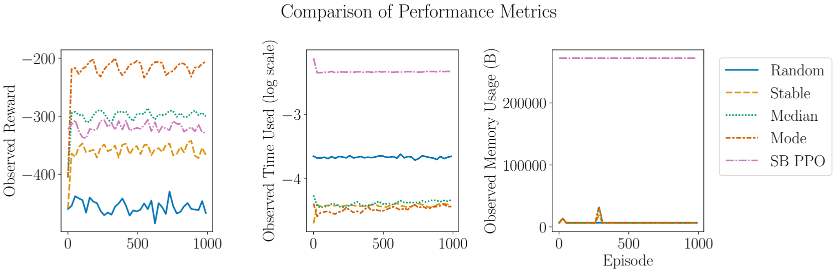
Using Real Data with the Graph Environment

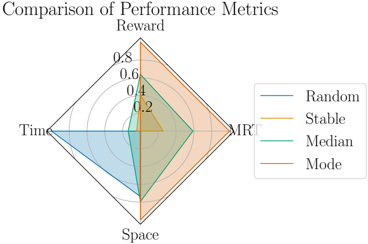
We used the osmnx package to generate a graph of Ithaca's streets, which is used as the state space for the environment.

Generated arrivals using police data on call locations; data is from the Ithaca police department.



Graph Agent Performance w/ Data ($\alpha = 0.25$)





Metric Environment Example

Suppose there are two ambulances in the environment and they start at locations 0.4 and 0.6

Metric Environment Example cont.

The agent can then choose new locations for the ambulances. In this case, the agent chose 0.3 and 0.65. The total distance travelled from the original state S_O to the action A is $D(S_O, A) = 0.15$

Metric Environment Example cont.

A call arrives at a location randomly sampled from the arrival distribution. The location in this example is 0.95

Metric Environment Example cont.

The closest ambulance travels to the call. The distance travelled from the action A to the new state S_N in this example is
$D(A, S_N) = 0.3$

The reward returned by this iteration is

$$-1 * [\alpha * (D(S_O, A)) + (1 - \alpha) * D(A, S_N)]$$

where α is a parameter controlling the tradeoff between the cost of travel between calls and the cost of travel to respond to a call