

# Human Reward Learning with BIRL

April 18, 2018

## 1 Domain Representations

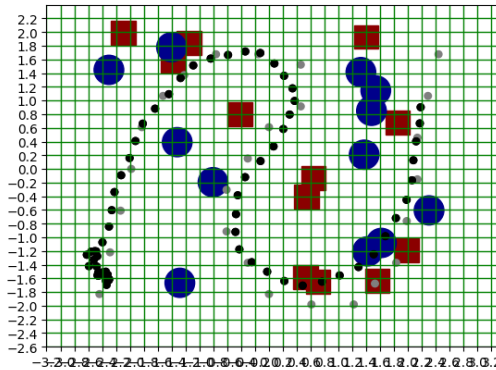


Figure 1: Example of discretized gridworld

In order to make the problem tractable by BIRL, the test area is discretized into a 2D gridworld of size  $32 \times 24$  with  $0.2 \times 0.2 \text{ m}^2$  cells. Each cell is a state in the MDP. The actions are discretized into 8 directions so that an agent can move to any adjacent state in the gridworld. The (center) location of targets, obstacles and pathpoints are treated as different feature points, which contribute to each state’s feature by distance. The problem is formulated as learning the weights for the three different features: targets, obstacles and pathpoints. The three features are represented using three different continuous values at each state. More specifically, the closer a state is to an target/obstacle/pathpoint, the higher the feature value for the particular object at that state. The reward at any given state is computed as the linear combination of these features using their corresponding weights. The observations are a set of state-action pairs extracted from the human’s trajectory, which are fitted to the discretization of the space.

## 2 Experimental Design Choices

The parameters for BIRL are set empirically. The confidence factor  $\alpha$  is set at 80 and the chain length is set to be 3000 (since there are only three values, i.e. feature weights to be

tweaked, which is relatively small). 0.5 is used as the discount factor for MDPs with the assumption that the decision making process of humans tends to prefer immediate rewards.

### 3 Results

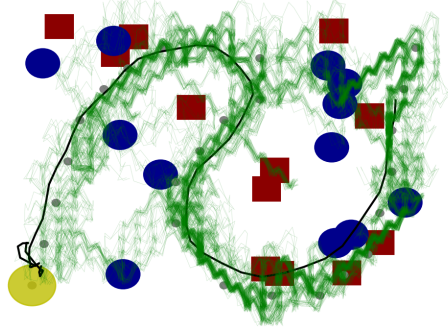


Figure 2: Example BIRL traces for Task 1

Performance is measured by angular differences in actions taken by the BIRL agent and by the human subject along all the states the human subject was in.

#### 3.1 Average performance across subjects

Total:

Task	Average angular difference
1	$53.87 \pm 2.537$
2	$53.369 \pm 2.71$
3	$59.864 \pm 1.997$
4	$51.088 \pm 2.596$

#### 3.2 Performance per subject

Subject ID 26 :

Task	Average angular difference
1	$27.901 \pm 6.638$
2	$56.231 \pm 8.117$
3	$72.996 \pm 2.01$
4	$66.936 \pm 2.782$

Subject ID 27 :

Task	Average angular difference
1	$71.298 \pm 2.143$
2	$63.729 \pm 8.428$
3	$73.013 \pm 1.335$
4	$63.328 \pm 10.718$

Subject ID 28 :

Task	Average angular difference
1	$52.038 \pm 11.266$
2	$63.386 \pm 2.171$
3	$72.775 \pm 2.296$
4	$21.301 \pm 1.613$

Subject ID 31 :

Task	Average angular difference
1	$68.127 \pm 11.727$
2	$62.79 \pm 4.864$
3	$51.341 \pm 12.189$
4	$75.07 \pm 3.244$

Subject ID 32 :

Task	Average angular difference
1	$62.917 \pm 11.987$
2	$39.125 \pm 11.89$
3	$30.295 \pm 2.014$
4	$39.843 \pm 11.66$

Subject ID 33 :

Task	Average angular difference
1	$61.881 \pm 12.049$
2	$58.009 \pm 13.688$
3	$71.132 \pm 3.861$
4	$57.484 \pm 13.724$

Subject ID 34 :

Task	Average angular difference
1	$76.871 \pm 8.415$
2	$74.661 \pm 4.55$
3	$73.39 \pm 2.909$
4	$77.317 \pm 3.127$

Subject ID 35 :

Task	Average angular difference
1	$47.282 \pm 15.993$
2	$69.235 \pm 0.837$
3	$65.007 \pm 11.036$
4	$31.16 \pm 12.727$

Subject ID 36 :

Task	Average angular difference
1	$47.036 \pm 10.412$
2	$43.565 \pm 17.449$
3	$49.129 \pm 13.396$
4	$79.417 \pm 5.735$

Subject ID 37 :

Task	Average angular difference
1	$79.528 \pm 5.452$
2	$65.886 \pm 16.982$
3	$71.9 \pm 4.103$
4	$77.418 \pm 3.452$

Subject ID 38 :

Task	Average angular difference
1	$67.316 \pm 2.553$
2	$78.237 \pm 5.031$
3	$33.727 \pm 1.826$
4	$15.505 \pm 2.04$

Subject ID 39 :

Task	Average angular difference
1	$48.528 \pm 15.749$
2	$82.112 \pm 7.056$
3	$40.801 \pm 7.712$
4	$43.805 \pm 15.795$

Subject ID 42 :

Task	Average angular difference
1	$45.479 \pm 11.671$
2	$10.618 \pm 1.683$
3	$54.398 \pm 12.366$
4	$56.364 \pm 9.343$

Subject ID 43 :

Task	Average angular difference
1	$68.002 \pm 2.478$
2	$11.637 \pm 1.286$
3	$81.341 \pm 2.508$
4	$36.878 \pm 11.635$

Subject ID 44 :

Task	Average angular difference
1	$77.104 \pm 4.322$
2	$46.058 \pm 15.931$
3	$69.221 \pm 3.912$
4	$70.405 \pm 2.103$

Subject ID 45 :

Task	Average angular difference
1	$71.065 \pm 7.463$
2	$12.925 \pm 0.665$
3	$57.408 \pm 9.027$
4	$56.713 \pm 12.502$

Subject ID 46 :

Task	Average angular difference
1	$32.997 \pm 12.932$
2	$67.45 \pm 3.604$
3	$78.446 \pm 4.298$
4	$18.815 \pm 0.644$

Subject ID 47 :

Task	Average angular difference
1	$54.126 \pm 14.343$
2	$19.558 \pm 1.839$
3	$29.846 \pm 1.748$
4	$44.143 \pm 11.713$

Subject ID 48 :

Task	Average angular difference
1	$44.224 \pm 10.455$
2	$74.592 \pm 5.037$
3	$75.599 \pm 3.458$
4	$39.064 \pm 8.041$

Subject ID 54 :

Task	Average angular difference
1	$34.989 \pm 13.017$
2	$79.911 \pm 1.405$
3	$68.111 \pm 4.007$
4	$44.253 \pm 11.792$

Subject ID 56 :

Task	Average angular difference
1	$45.078 \pm 7.215$
2	$70.324 \pm 5.918$
3	$28.161 \pm 2.13$
4	$29.036 \pm 11.581$

Subject ID 59 :

Task	Average angular difference
1	$53.279 \pm 7.645$
2	$53.219 \pm 10.744$
3	$49.667 \pm 8.677$
4	$72.353 \pm 4.009$

Subject ID 61 :

Task	Average angular difference
1	$46.511 \pm 15.732$
2	$11.256 \pm 2.559$
3	$36.445 \pm 6.13$
4	$77.837 \pm 1.816$

Subject ID 63 :

Task	Average angular difference
1	$57.092 \pm 11.014$
2	$72.277 \pm 4.968$
3	$71.945 \pm 3.254$
4	$58.373 \pm 10.365$

Subject ID 64 :

Task	Average angular difference
1	$16.468 \pm 4.118$
2	$24.179 \pm 10.995$
3	$51.322 \pm 13.496$
4	$26.093 \pm 11.173$