

Human Reward Learning with BIRL

October 28, 2017

1 Domain Representations

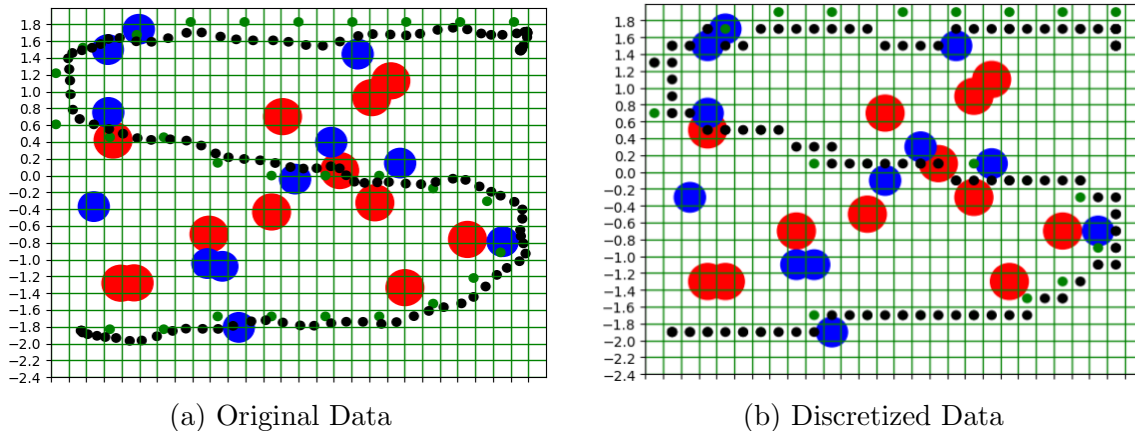


Figure 1: Example of data discretization

In order to make the problem tractable by BIRL, the test area is discretized into a 2D gridworld of size 28×22 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, pathpoints and the subject's waypoints are approximated with their closest discrete state as shown in figure 1.

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 center cell of targets and obstacles have a feature value of 1.0 for the corresponding feature and the eight adjacent cells will take the value of 0.5, while the centers of pathpoints take an increasing value as their order increases. 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.

2 Experimental Design Choices

Since the discretization as well as test-time error can introduce noise into the data, only a (randomly sampled) subset of available human state-action pairs is used per iteration of BIRL. The model with lowest training error across 50 iterations is used as its final prediction.

The parameters for BIRL are set empirically. The confidence factor α is set at 100 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). 0.25 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

3.1 Average performance across subjects

Task	Average angular difference
1	53.913 ± 2.54
2	53.409 ± 2.709
3	59.856 ± 1.997
4	51.063 ± 2.593

3.2 Performance per subject

Subject ID 26 :

Task	Average angular difference
1	28.102 ± 6.593
2	56.231 ± 8.117
3	72.996 ± 2.01
4	66.936 ± 2.782

Subject ID 27 :

Task	Average angular difference
1	71.298 ± 2.143
2	63.631 ± 8.417
3	73.013 ± 1.335
4	63.328 ± 10.718

Subject ID 28 :

Task	Average angular difference
1	52.038 ± 11.266
2	63.698 ± 2.403
3	72.775 ± 2.296
4	21.3 ± 1.613

Subject ID 31 :

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

Subject ID 32 :

Task	Average angular difference
1	62.917 ± 11.987
2	39.36 ± 11.98
3	30.295 ± 2.014
4	39.843 ± 11.66

Subject ID 33 :

Task	Average angular difference
1	61.881 ± 12.049
2	58.009 ± 13.688
3	71.132 ± 3.861
4	56.822 ± 13.512

Subject ID 34 :

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

Subject ID 35 :

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

Subject ID 36 :

Task	Average angular difference
1	47.036 ± 10.412
2	43.106 ± 17.267
3	49.129 ± 13.396
4	79.417 ± 5.735

Subject ID 37 :

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

Subject ID 38 :

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

Subject ID 39 :

Task	Average angular difference
1	48.528 ± 15.749
2	82.601 ± 6.687
3	40.801 ± 7.712
4	43.805 ± 15.795

Subject ID 42 :

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

Subject ID 43 :

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

Subject ID 44 :

Task	Average angular difference
1	78.791 ± 2.986
2	46.058 ± 15.931
3	69.221 ± 3.912
4	70.405 ± 2.103

Subject ID 45 :

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

Subject ID 46 :

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

Subject ID 47 :

Task	Average angular difference
1	54.126 ± 14.343
2	19.99 ± 1.742
3	29.846 ± 1.748
4	44.143 ± 11.713

Subject ID 48 :

Task	Average angular difference
1	44.499 ± 10.585
2	74.592 ± 5.037
3	75.599 ± 3.458
4	39.064 ± 8.041

Subject ID 54 :

Task	Average angular difference
1	34.989 ± 13.017
2	79.911 ± 1.405
3	68.111 ± 4.007
4	44.712 ± 11.939

Subject ID 56 :

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

Subject ID 59 :

Task	Average angular difference
1	52.767 ± 7.486
2	53.219 ± 10.744
3	49.437 ± 8.545
4	72.352 ± 4.009

Subject ID 61 :

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

Subject ID 63 :

Task	Average angular difference
1	56.442 ± 11.574
2	72.277 ± 4.968
3	71.945 ± 3.254
4	57.861 ± 10.215

Subject ID 64 :

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