

## Assignment 9

a) The environment  $m$  is modeled as a graph  $G = (V, E)$ , where nodes  $v \in V$  are street intersections, and edges  $e \in E$  are streets. The space of all directions in  $m$  from current location  $v_1$  to a destination location  $v_{goal}$  according to given route is defined by a deterministic MDP  $M$ ; actions of MDP - single instruction that guides user from one adjacent node to another. Then each trajectory  $\tau$  (sequence of states and actions that defines the state transition dynamics) in  $M$  corresponds to valid directions that guide the user from its current location to its target location. So, route directions are generated as trajectories in  $M$ .

b) Goal of learning is to recover unknown reward function given a set of demonstrations (trajectories)  $\{\tau_1, \dots, \tau_n\}$ .

c) A set of directions (trajectories) for each route from a set of given routes was recorded from the participants. Then set of feature vector for each trajectory was derived. Using this set, reward function was learned. It was used further to generate a set of test routes.

e) Maximum entropy inverse reinforcement learning. Within MaxEnt IRL it is assumed that behavior of agent is governed by a probability distribution of trajectories that is a function  $f_\tau \in R^n$ . The method considers maximum entropy distribution for trajectories whose expected values match observed feature values  $\tilde{f}_\tau$ :

$$\tilde{f}_\tau = E_{p_\theta}[f_\tau]$$

Under this constraint, algorithm optimizes objective function:

$$p_\theta^* = \operatorname{argmax}_{p_\theta} H_{p_\theta}(\tau)$$

Here  $H(p_\theta)$  is Shannon entropy of distribution  $p_\theta$ . The resulting distribution:

$$p_\theta^* = \exp^{-\theta^T f(\tau)}$$

and the gradient with respect to its parameters  $\theta$  is:

$$\tilde{f}_\tau - E_{p_\theta}[f_\tau]$$

d) Parameters  $\theta$  (vector) are adjusted by learning. The size of one vector  $\theta$  is the same as the size of feature vector of corresponding trajectory.