Differentiable inter-agent learning (DIAL)

Drawback

Since DIAL is designed for independent learners, it inevitably faces the challenge of not being able to tackle the non-stationary environments.

CommNet

Drawback

As the communication network is fully symmetric and embedded in the original network, it lacks the ability of handle heterogeneous agent types. Also it is a single network for all agents, and therefore its scalability is unclear.

3.1

Global reward.

Minimax Q-function.

Deal with b:

When the enemies are AI agents, we consider fictitious play, where both the controlled agents and the enemies apply empirical strategies learned from the other players so far, and then iteratively optimise the above Q-function.

When dealing with the Game’s own control of the enemies, we modelled the deterministic policy b through DNN in a supervised learning setting. By learning the policy of enemies and fixing them, SG defined in Eq. (2) effectively turns into a simper MDP problem.

3.2

Local reward, top-K list.

Actor-critic method.

Grouping (based on geometry):

Intra-group behaviour

Bi-directional RNN:

Inter-group behaviour.

The dependency of our agents are built upon the internal layers, rather than directly from the actions.

Homogenuous agents: share parameters.

Heterogenuous agents: don’t share parameters.

State space:

The state space of each individual agent is represented as a tensor of size 72 x 72 x 16 extracted from the map of size 72 x 72. Each channel in the state space describes either the hit points, the damage or the safety attacking range for all the agents on the map.

Questions:

两个agent位置很近，但是biRNN网络中离得远，互相影响会变小，怎么办？