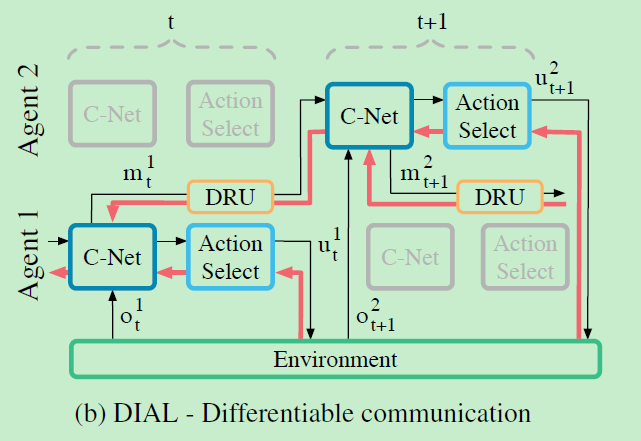
QValue-Based method.

Deep Recurrent Q-Networks.

Differentiable Inter-Agent Learning



Message mt send from previous C-Net to the next C-Net. Hence, gradients flow across agents, from the recipient to the sender

1. During centralised learning, communication actions are replaced with direct connections between the output of one agent’s network and the input of another’s.
2. C-Net, outputs two distinct types of values, as shown in Figure 1(b), a) Q(.), the Q-values for the environment actions, which are fed to the action selector, and b) mta , the real-valued message to other agents, which bypasses the action selector and is instead processed by the discretise/regularise unit (DRU(mta)).

Review from ali starcraft paper

Since DIAL is designed for independent learners, it inevitably faces the challenge of not being able to tackle the non-stationary environments; in other words, the environment will keep changing for each agent.