Actor-critic.

Centralized Q function. (infer policy)

Policy Ensembles.

Use counterfactual baseline: Actor not only follow global reward, but also take into account agent’s own contribution.

Local reward, top-K list.

Grouping.

Network

Bi-directional RNN.

Share parameters.