Problem

Centralised approach fail:

one agent learns a useful policy, but a second agent is discouraged from learning because its exploration would hinder the first agent and lead to worse team reward.

Design individual reward functions:

Reward shaping is difficult and only a small class of shaped reward functions are guaranteed to preserve optimality w.r.t. the true objective.

Related works

Discuss approaches how to deal with global objective vs local objective.

4.1

Since observations are from a local perspective, we do not benefit from convolutional networks, but use a fully connected linear layer to process the observations.