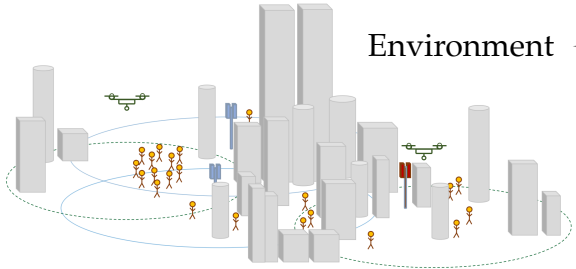


Environment

Airborne-BSs (Agents)

Reinforcement Learning Algorithm

Observation

Control

SARSA .....

Algorithm 1: SARSA implementation of two drones scenario

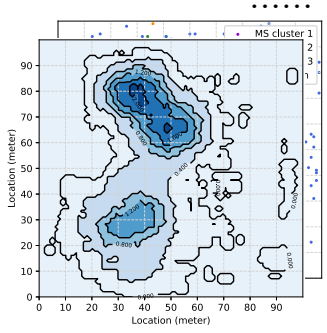
Initialization
for every episode j **do**
  $s_1 \leftarrow$ random
  for Every iteration t **do**
    for Every drone δ **do**
      $a_t \leftarrow$ Choose action based on current Q Table $(Q_{s_t}, \varepsilon_t, \delta)$
      $s_{t+1} \leftarrow$ Take the previous action (s_t, a_t, δ)
      $a_{t+1} \leftarrow$ Choose action based on real scenario $(Q_{s_{t+1}}, \varepsilon_t, \delta)$
      $r_t \leftarrow$ Compute the reward (s_{t+1})
      $Q(s_t, a_t, d) \leftarrow$ Update Q Table
      $s_t \leftarrow s_{t+1}$
    end for
  end for
end for

Deep Q Network

Algorithm 2: DQN implementation of two drones scenario

Initialization
for every episode j **do**
  $s_1 \leftarrow$ random
  for Every iteration t **do**
    for Every drone δ **do**
      $a_t \leftarrow \max_a Q(\phi(s_t), a; \theta)$ with probability ϵ select a random actio
      $r_t, x_{t+1} \leftarrow$ based on the action a_t
      $\phi_{t+1} = \phi(s_{t+1}) \leftarrow s_{t+1} = (s_t, a_t, x_{t+1})$
      $D \leftarrow$ Add data to the dataset $D + (\phi_t, a_t, r_t, \phi_{t+1})$
      $NN_{t+1} \leftarrow$ NN learn from data selected in D
      $s_t \leftarrow s_{t+1}$
    end for
  end for
end for

Mobile Station
Airborne-BSs
Buildings
Connection Condition



Environment information send to