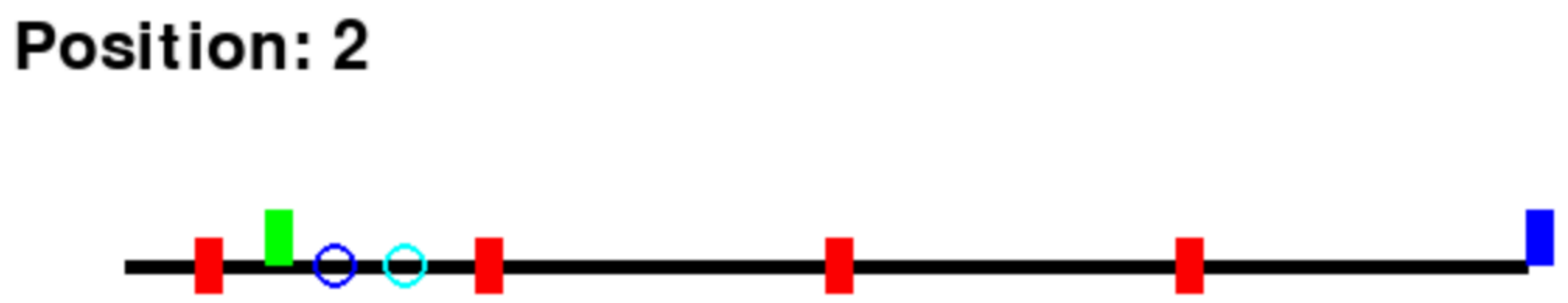
Markov Decision Process



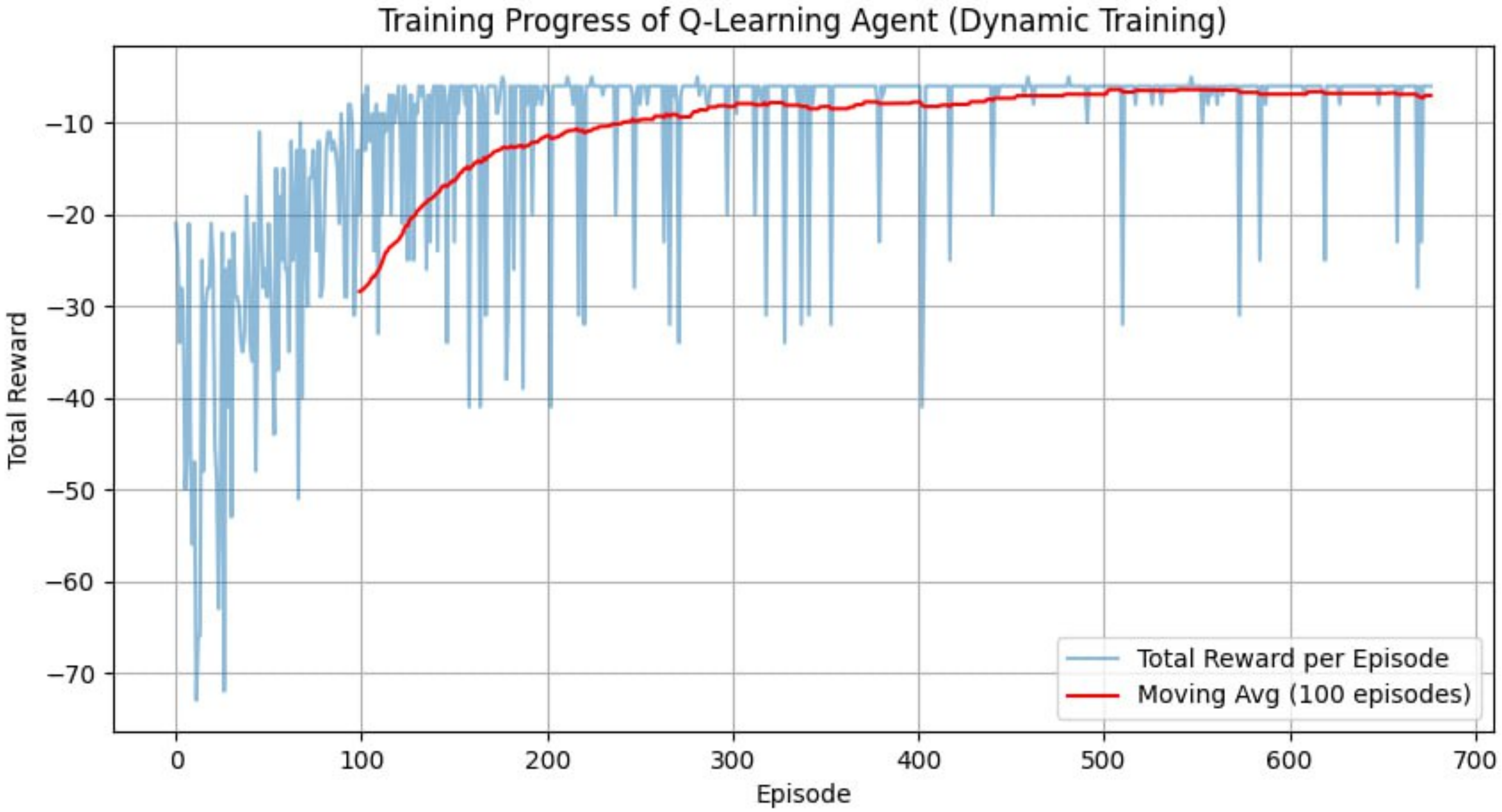
- State space: {0, 1, ..., 20}
  - goal state: 20
  - obstacles: {1, 5, 10, 15}
- Action space:
  - move left (x -= 1)
  - move right (x += 1)
  - jump (x += 2)

- Rewards:
  - move left/right: -1
  - jump: -3
  - obstacle bumped: -20
  - goal state reached: +20

Q-learning

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right)$$

Results


Training Progress of Q-Learning Agent (Dynamic Training)
Total Reward
−10
−20
−30
−40
−50
−60
−70
0
100
200
300
400
500
600
700
Episode
Total Reward per Episode
Moving Avg (100 episodes)