Exploration of Optimal Path in Poi Traversal by RL

Reporter：徐博文

content

ONE

Introduction

Part. 01

Introduction

Application

Intelligence Navigation for optimal path

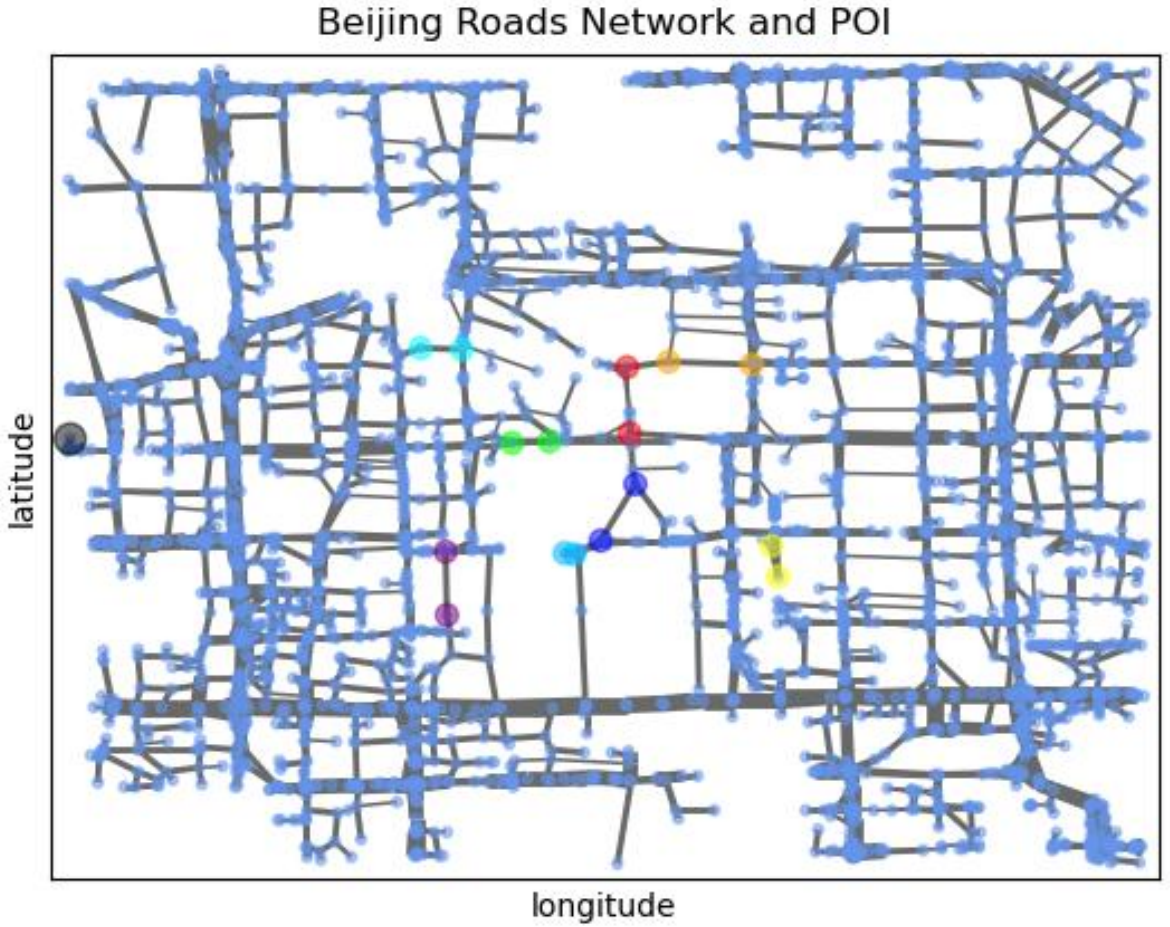
Method

Dijkstra algorithm

RL algorithm: Q-Learning, Sarsa, DQN

Ant colony Optimization

Background: Beijing Road Network



TWO

Algorithm

Part. 02

2.1 Dijkstra

——solve the shortest path with single source problem

Core Idea (Greedy):

- Start from a vertex s.
- Choose the unvisited nearest v to s and set v as visited.
- Tranverse every unvisited vertex v_i adjacent to v:
  - Compare dis(v_i,s) and dis(v_i,v) + dis(v, s)
  - Update dis(v_i,s)
- Repeat step2 and step3 until every vertex is visited

Easy to fall in local optimum and have large time cost

2.2 Q-Learning

Type: value-based, off-policy

Basis: Bellman Optimality Function:

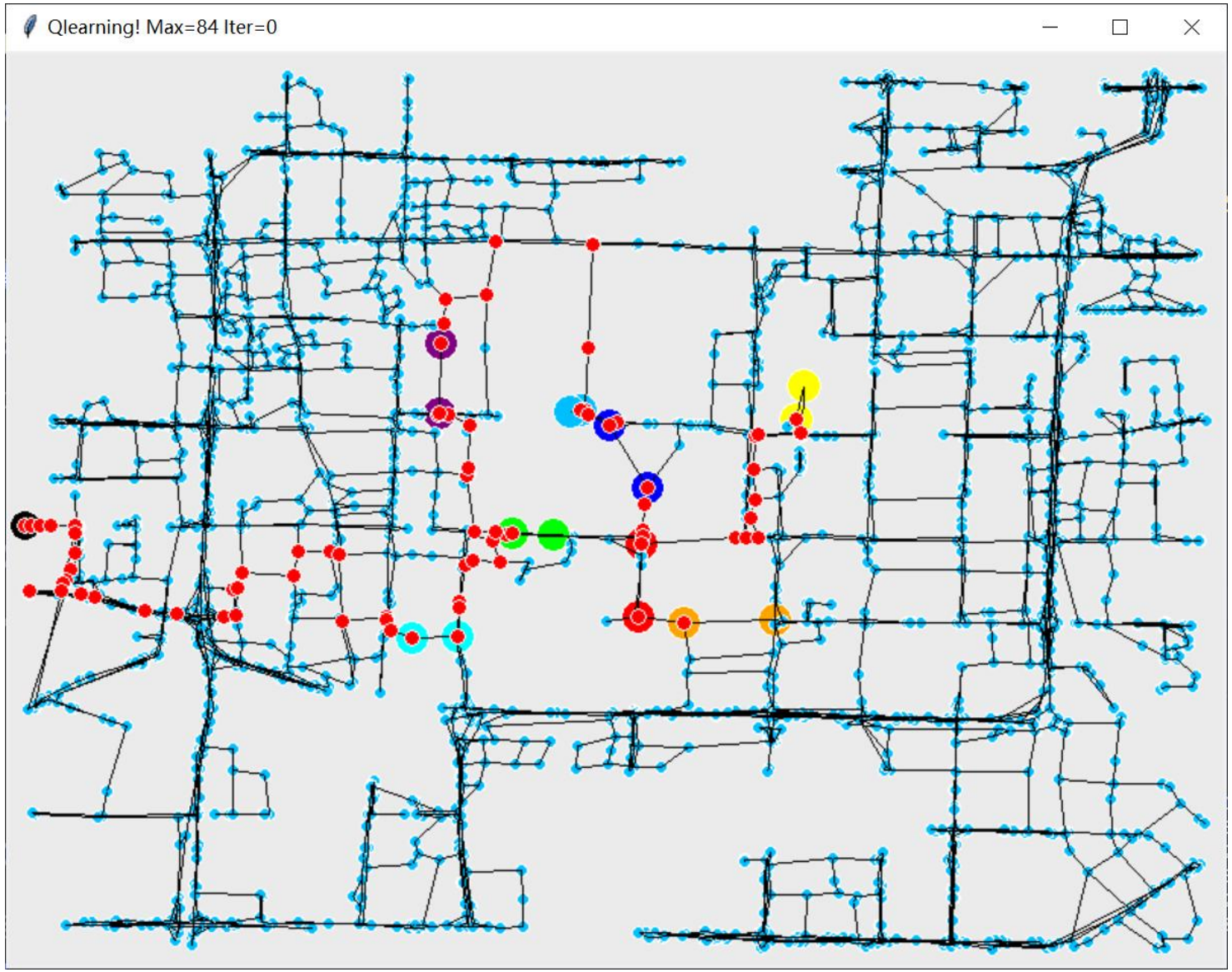
$$Q(s,a) = R + \gamma \sum_{s' \in S} P^{a}_{ss'} \max_{a' \in A} Q(s',a'),$$

Control Algorithm:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(R + \gamma \max_{a' \in A} Q(s',a') - Q(s,a))$$

Parameter Set: State S, Action A, Reward R

2.2 Q-Learning Performance



2.3 Sarsa

Type: value-based, on-policy

Basis: Bellman Expectation Function:

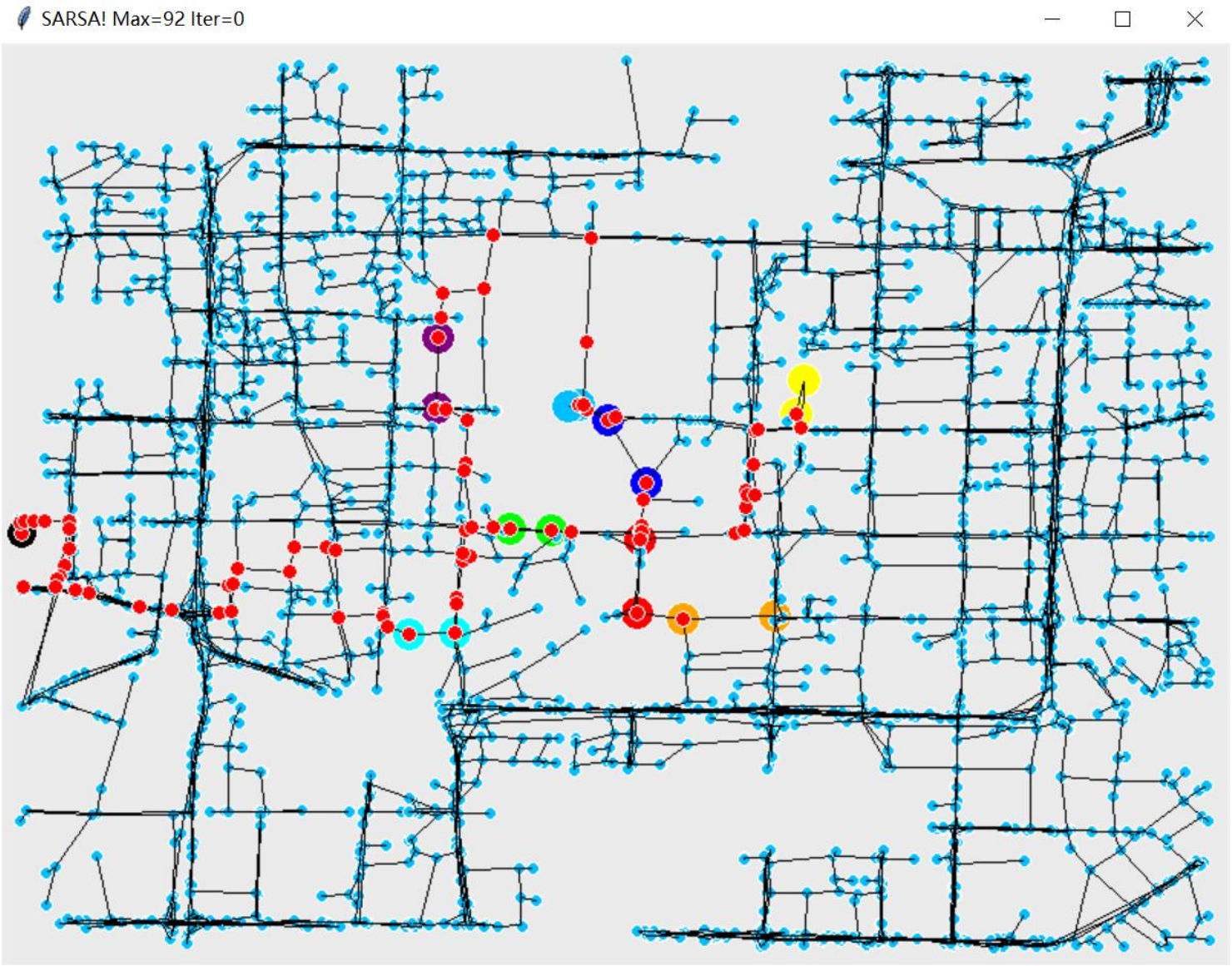
$$Q(s,a) = R + \gamma \sum_{s' \in S} P^{a}_{ss'} \dot{Q}(s',a')$$

Control Algorithm:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(R + \gamma Q(s',a') - Q(s,a))$$

Parameter Set: State S, Action A, Reward R

2.3 Sarsa Performance



2.4 DQN

Basis: Bellman Optimality Function
(Similar to Q-learning)

Core Idea: Approximate the action value
by neural network

2.5 Ant Colony

Core Idea:

Assuming that there are many ants at the initial position, the ants will use ε-greedy method according to the pheromone and choose the next step to be taken, while they will leave the pheromone behind.

Compare with the current optimal ant and update the optimal solution until all cities are traversed.

THREE

Different Scenarios

Part. 02

3.1 Shortest Route

Situation: The optimal path is the route with the **shortest total length**

Solution: Change the reward to the difference of the **total road length** between the two states and all unvisited poi

3.1 Shortest Time

Situation: The optimal path is the route with the **shortest total time**

Solution: Decrease the reward by the difference of the **total time** between the two states and all unvisited poi

Time of each road = road length / road speed limit

Thanks!