Xây dựng hệ thống điều khiển đèn giao thông bằng Deep Reinforcement Learning

Nguyễn Đỗ Đức Anh, Đỗ Lê Duy

Ngày 18 tháng 6 năm 2019

Mục lục

Mục lục

Giới thiệu

Phạm vi đề tài

- Xây dựng hệ thống điều khiển đèn giao thông trên phần mềm mô phỏng **SUMO**.
- Mô hình đạt được áp dụng cho **ngã tư**, **ngã ba**.

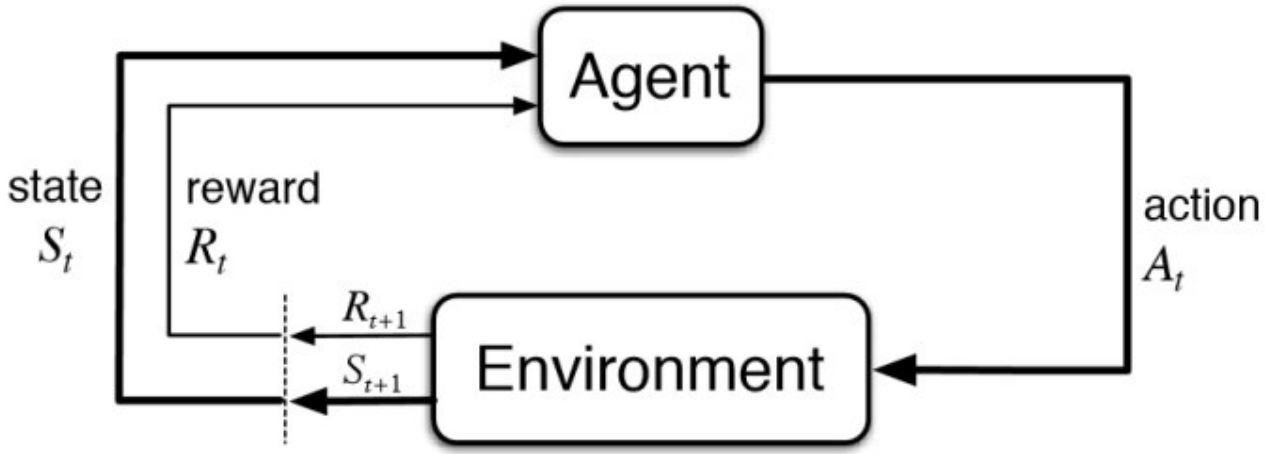
Mục tiêu

- Giảm **thời gian chờ** của mỗi người khi đi qua giao lộ.
- Mô hình thử nghiệm hiệu quả với mọi loại kịch bản giao thông **ngẫu nhiên**.

Phương pháp tiếp cận

Deep Reinforcement Learning

Học tăng cường (reinforcement learning) là một lĩnh vực con của học máy.



Kiến thức nền tảng

Giải thuật Deep-Q-Learning, Markov decision process, Convolution neuron network, giải thuật Greedy-ϵ.

Mục lục

Mô hình đề xuất

Bài báo tham khảo

- Traffic Light Control Using Deep Policy-Gradient and Value-Function Based Reinforce-ment Learning (DePGVF) in 2017
- Deep Reinforcement Learning for Traffic Light Control in Vehicular Networks (DeTLC) in 2018

Lựa chọn mô hình

- Trạng thái
- Hành động
- Phần thưởng
- Mạng nơ-ron

Mô hình đề xuất

Trạng thái

Trạng thái (state) là dữ liệu đầu vào của mô hình, giúp Agent nhận biết môi trường mà nó tương tác.

Mô hình đề xuất

Trạng thái

Trạng thái (state) là dữ liệu đầu vào của mô hình, giúp Agent nhận biết môi trường mà nó tương tác.

Thông tin môi trường

- Vị trí.
- Tốc độ.
- Chiều dài hàng đợi.
- Thời gian chờ.

Mô hình đề xuất

Trạng thái

Trạng thái (state) là dữ liệu đầu vào của mô hình, giúp Agent nhận biết môi trường mà nó tương tác.

Thông tin môi trường

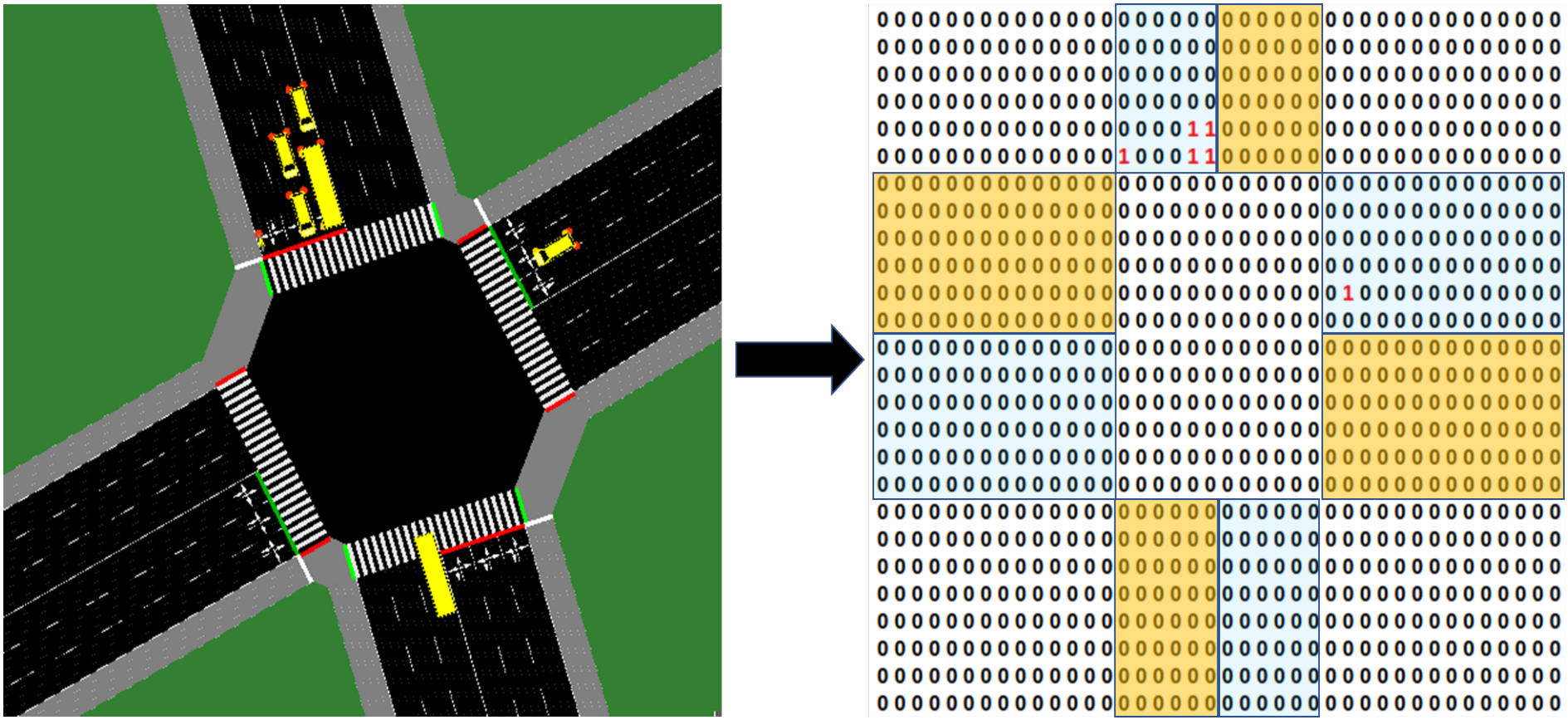
- Vị trí.
- Tốc độ.
- Chiều dài hàng đợi.
- Thời gian chờ.

Hiện thực trạng thái

Ma trận: $60x60x2$

Mô hình đề xuất

Ánh xạ 1-1:

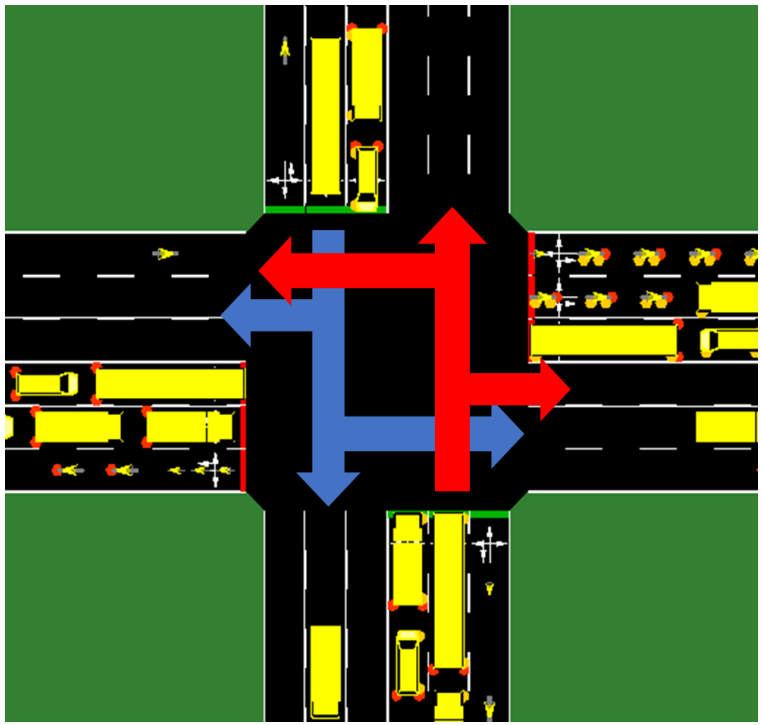


Hình: Ma trận 1-1

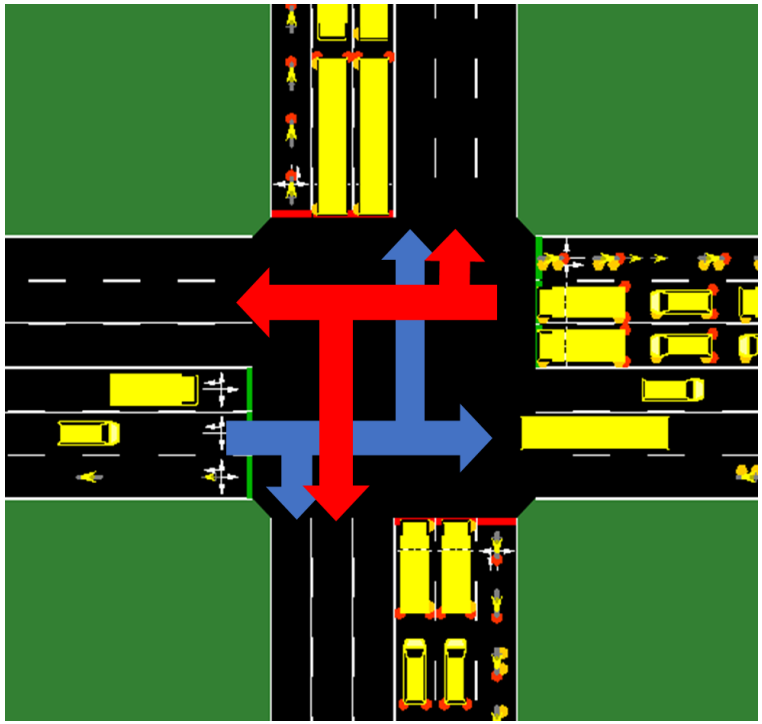
Mô hình đề xuất

Hành động

Hành động được Agent lựa chọn và thực hiện nhằm thay đổi môi trường mà nó tương tác.

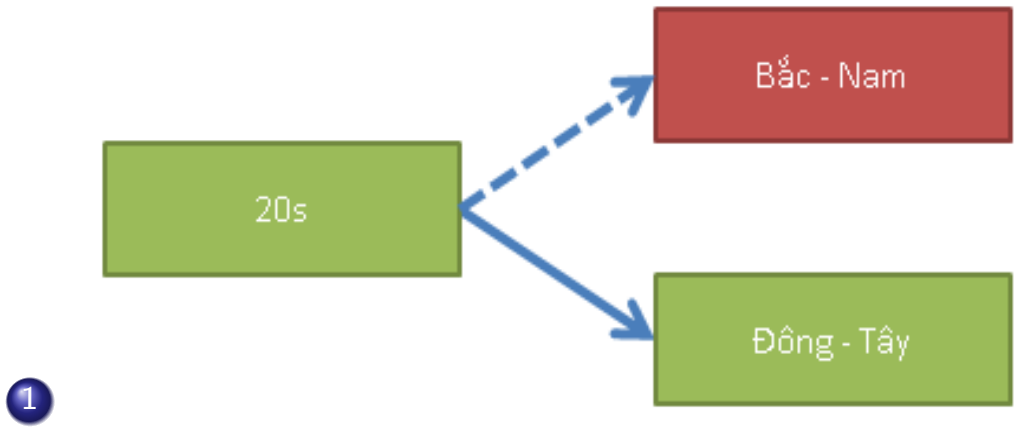


Hình: Bắc - Nam

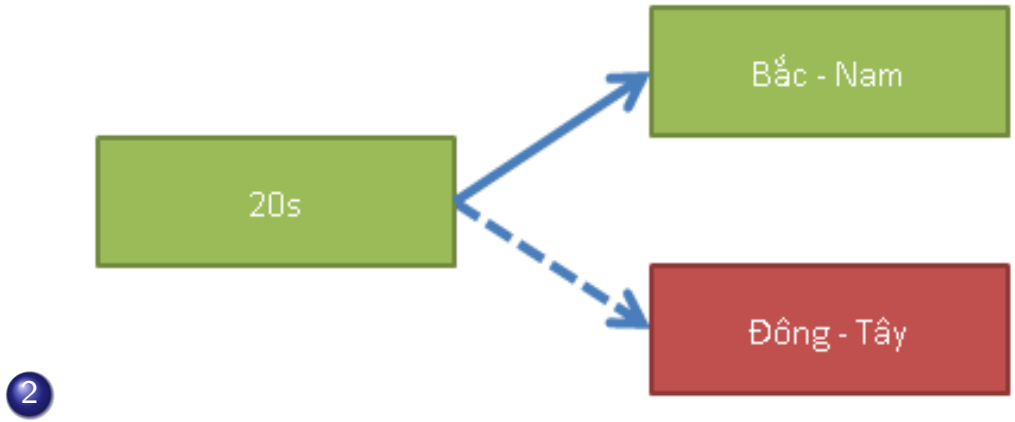


Hình: Đông - Tây

Mô hình đề xuất



Hình: Chọn chiều Đông - Tây



Hình: Chọn chiều Bắc - Nam

Mô hình đề xuất

Phần thưởng

Vai trò của phần thưởng là thể hiện kết quả phản hồi từ môi trường về hành động mà Agent vừa thực hiện.

Mục tiêu quan tâm

Giảm thời gian chờ.

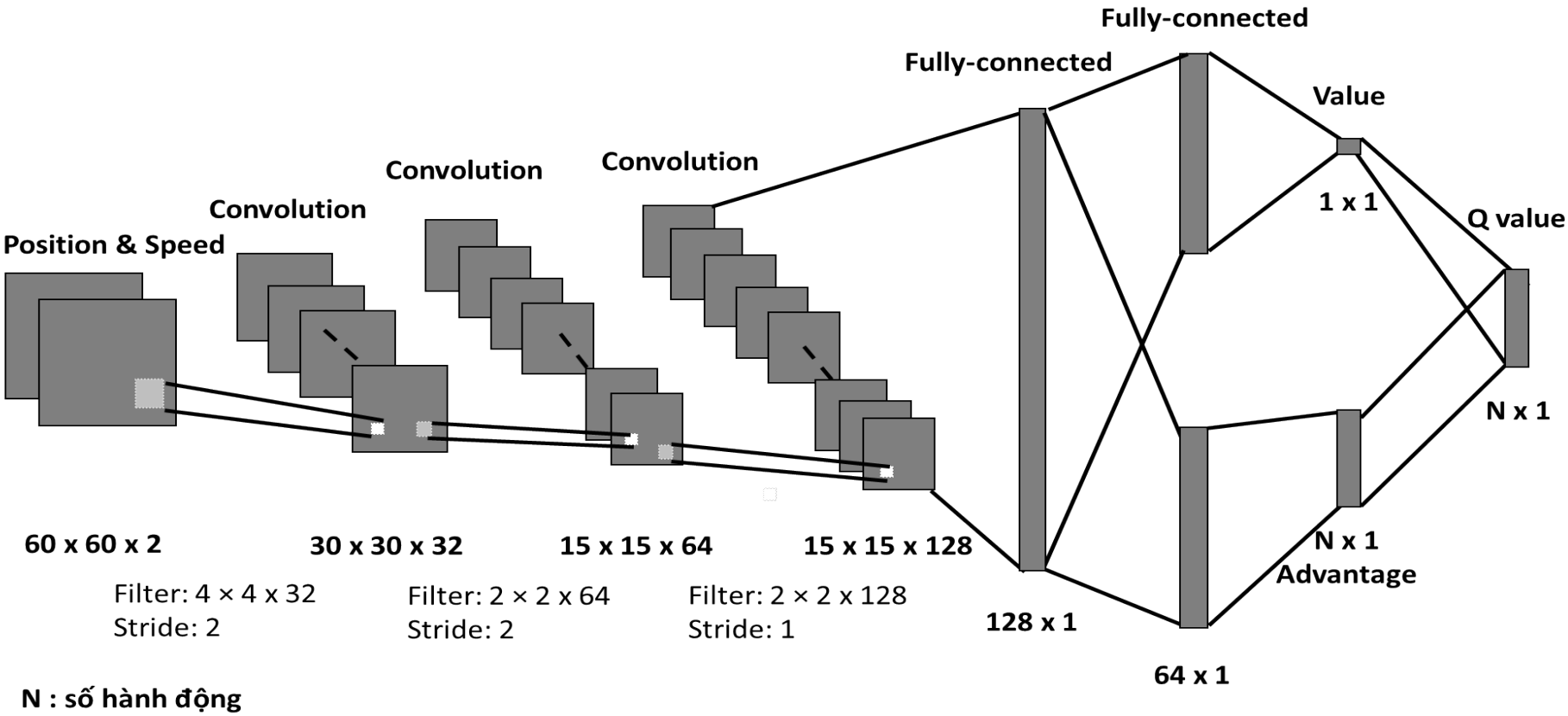
Mô hình đề xuất

t: step t
D_t: tổng thời gian chờ tức thời tại step t.

$$r_t = D_{t-1} - D_t \tag{1}$$

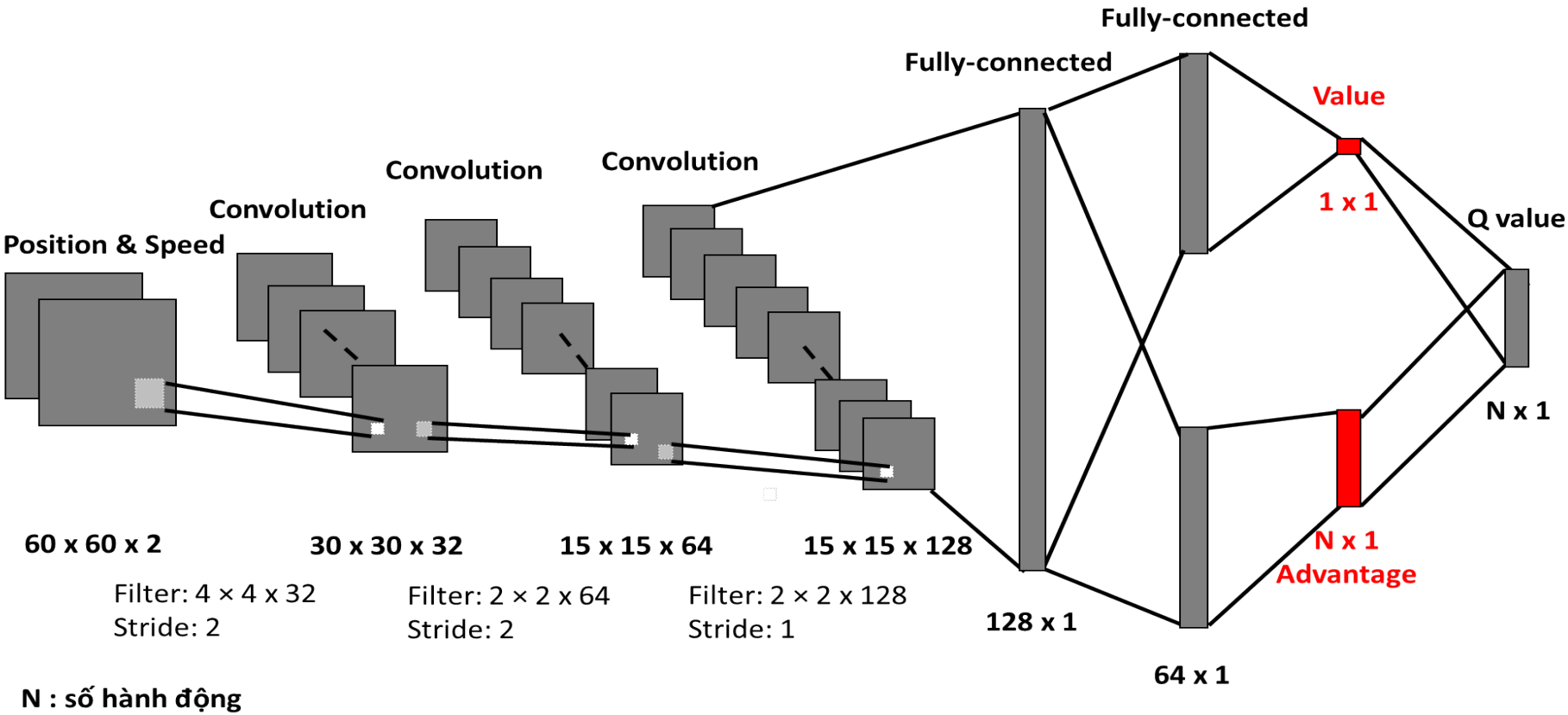
- $r_t > 0$: giảm tổng thời gian chờ giữa hai steps.
- $r_t < 0$: tăng tổng thời gian chờ giữa hai steps.
- $r_t = 0$: tổng thời gian chờ không thay đổi giữa hai steps.

Mạng nơ-ron



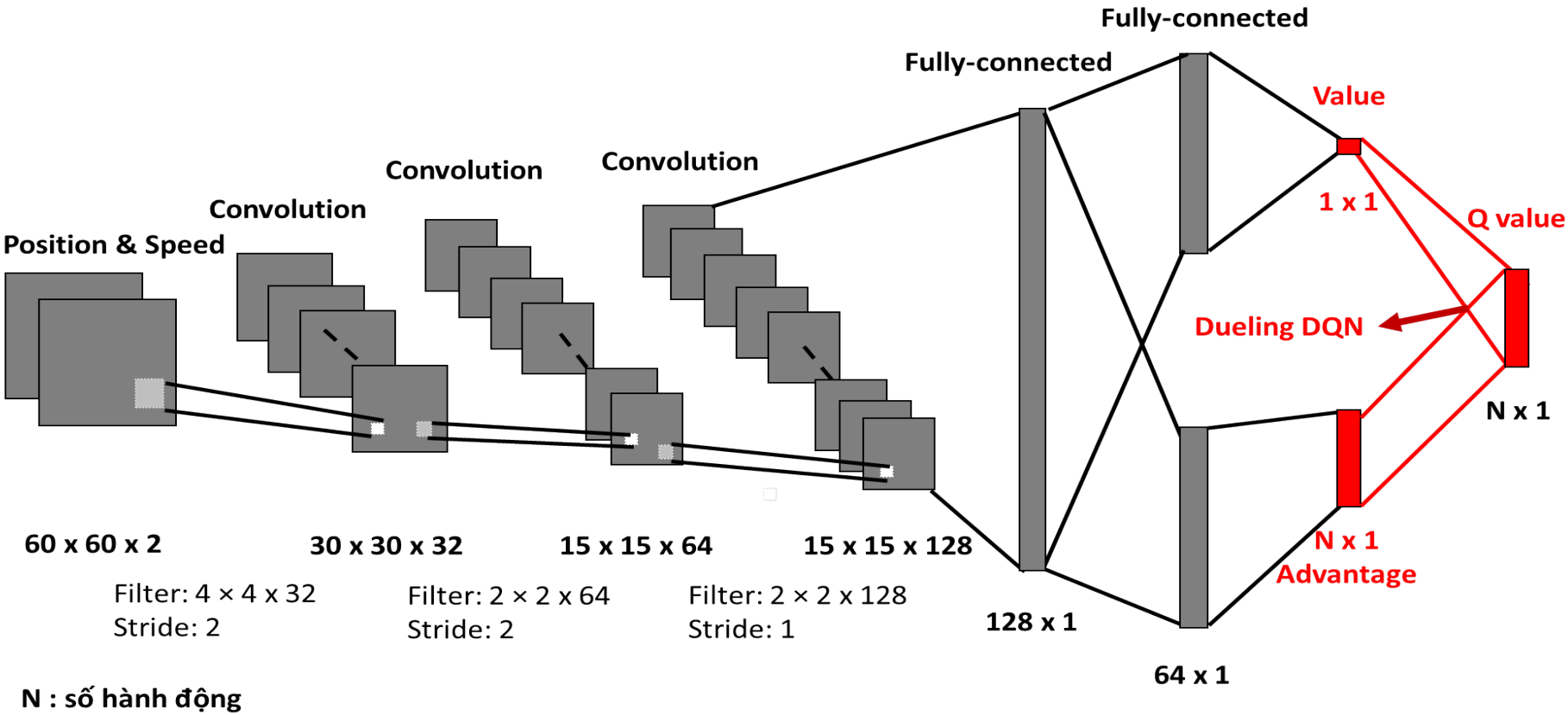
Hình: Mạng nơ-ron

Mạng nơ-ron



Hình: Mạng nơ-ron

Mạng nơ-ron



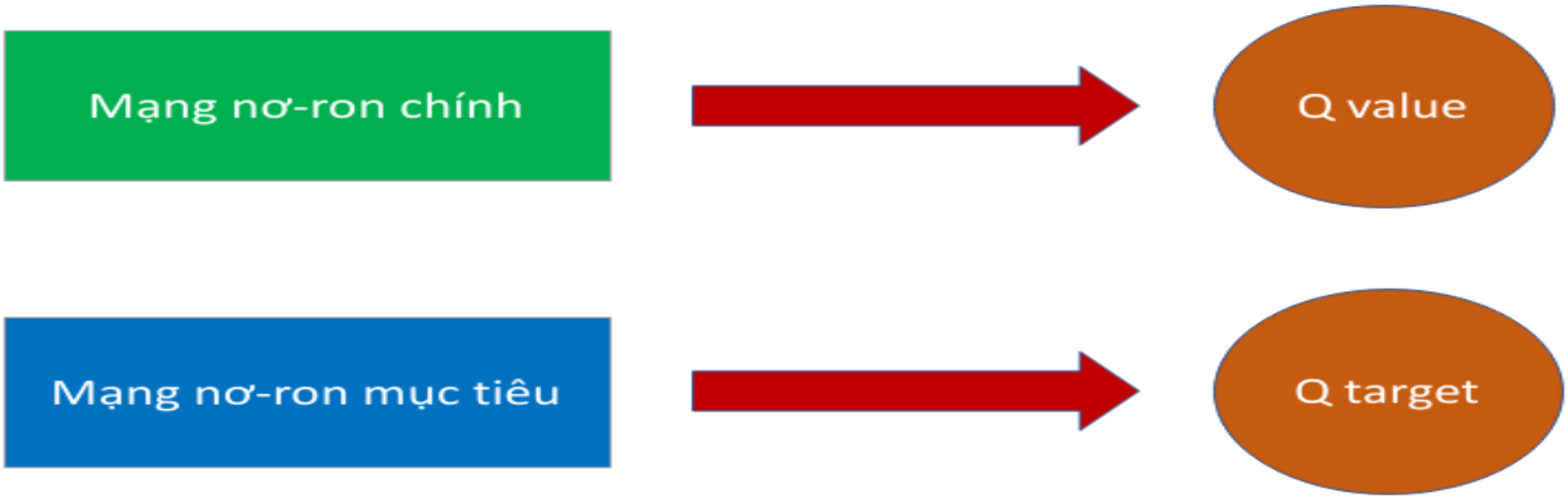
Hình: Mạng nơ-ron

Mạng nơ-ron

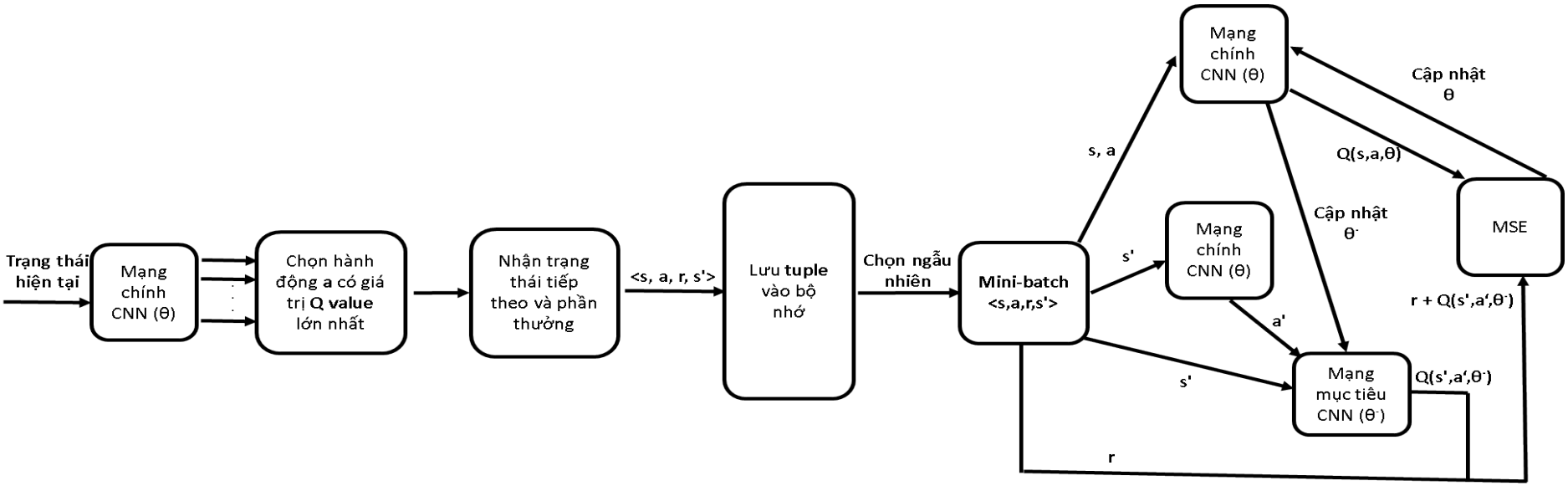
.

Mạng mục tiêu

Kiến trúc tương tự mạng nơ-ron chính, đóng vai trò như nhãn trong các giải thuật huấn luyện đánh nhãn.

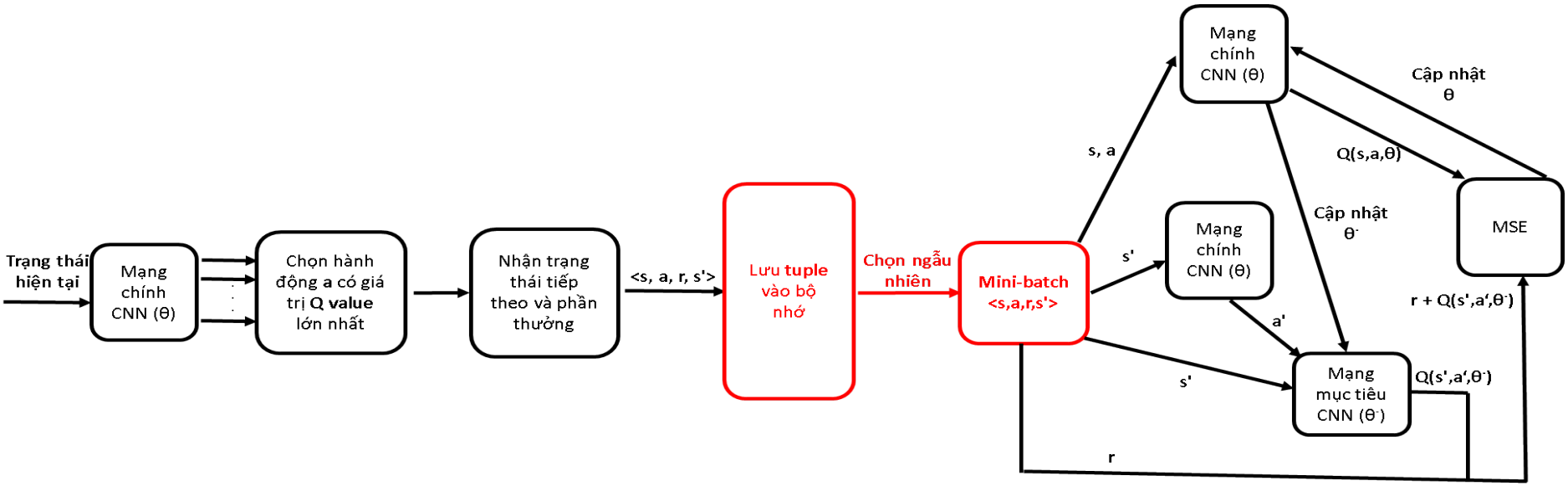


Kiến trúc tổng quan



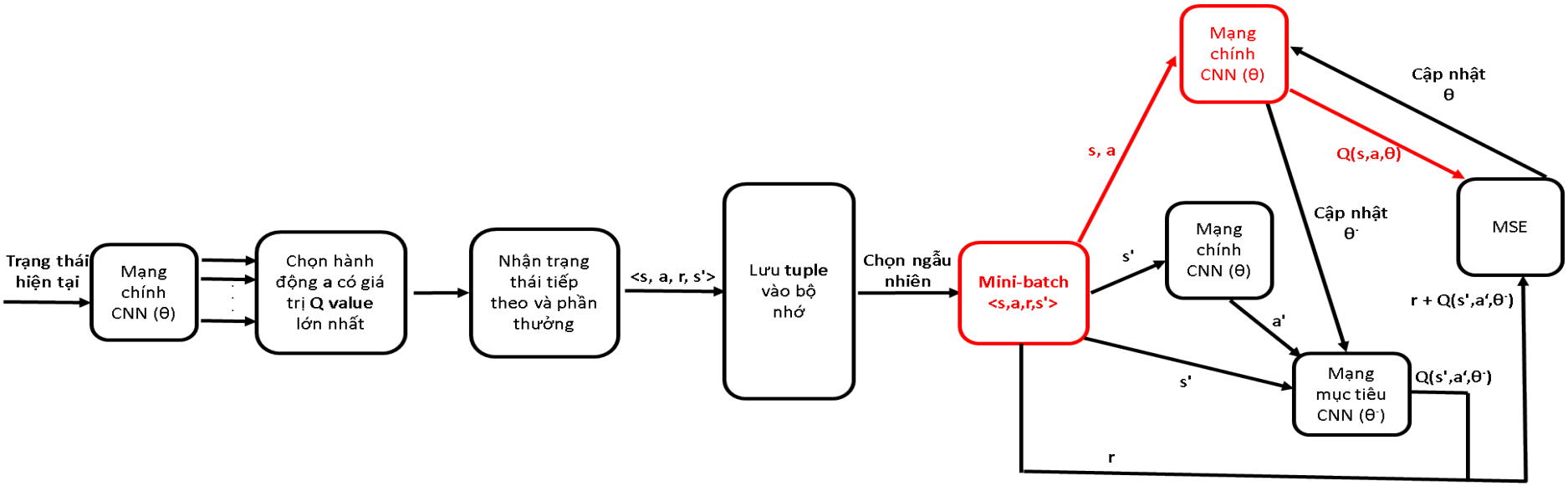
Hình: Kiến trúc tổng quan

Kiến trúc tổng quan



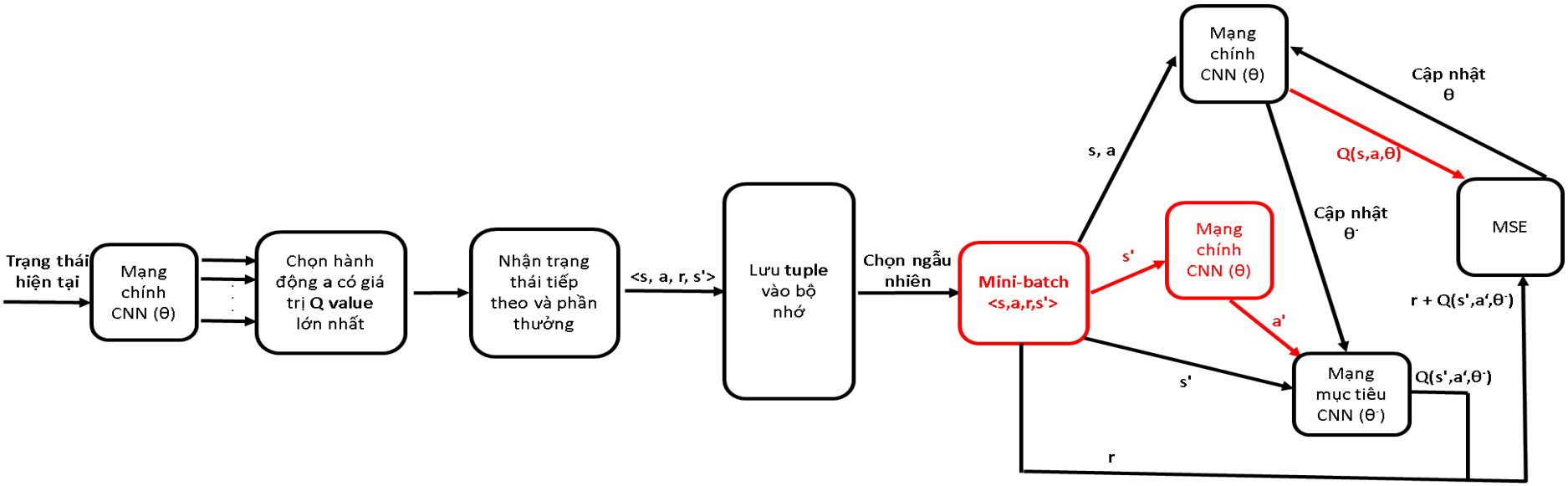
Hình: Kiến trúc tổng quan

Kiến trúc tổng quan



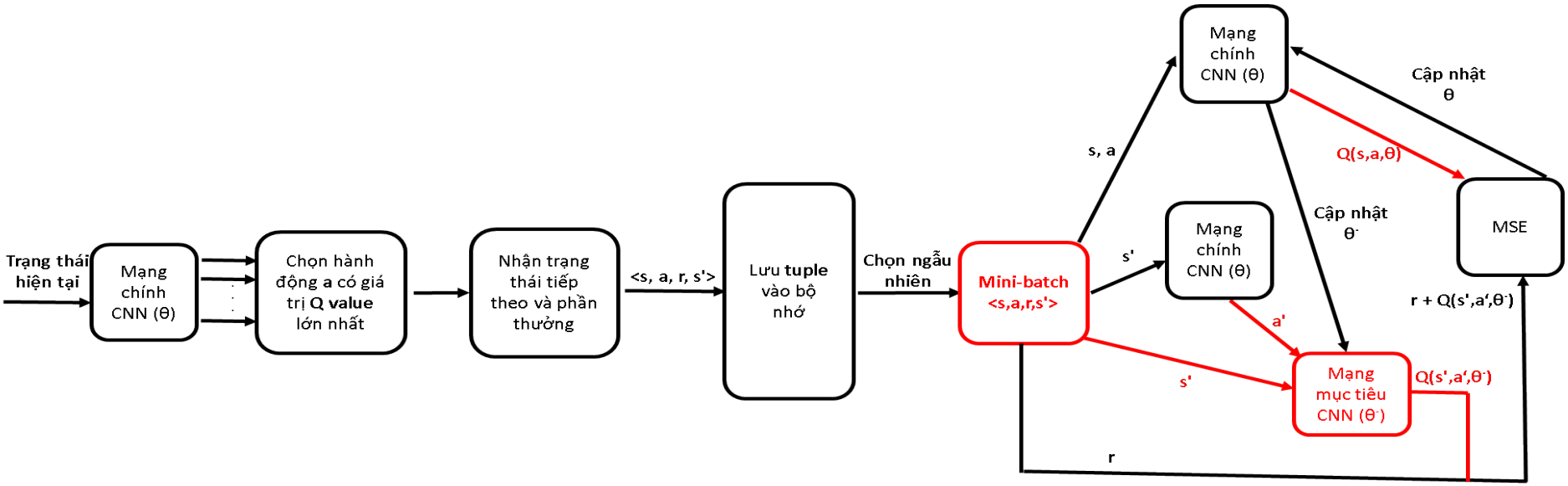
Hình: Kiến trúc tổng quan

Kiến trúc tổng quan



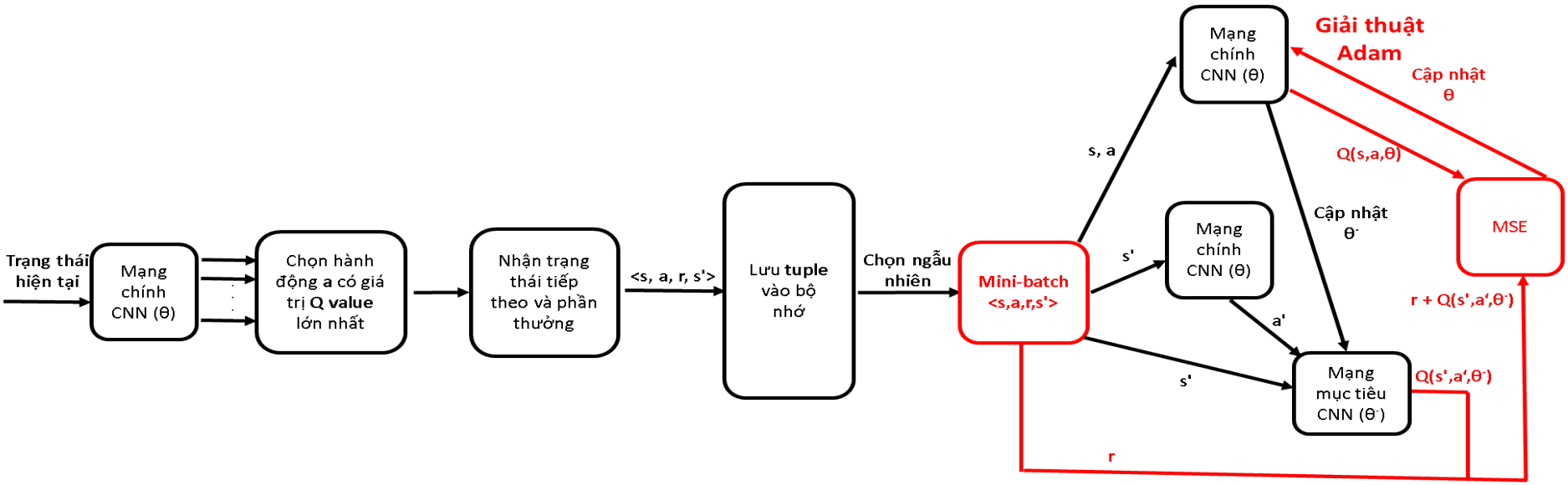
Hình: Kiến trúc tổng quan

Kiến trúc tổng quan



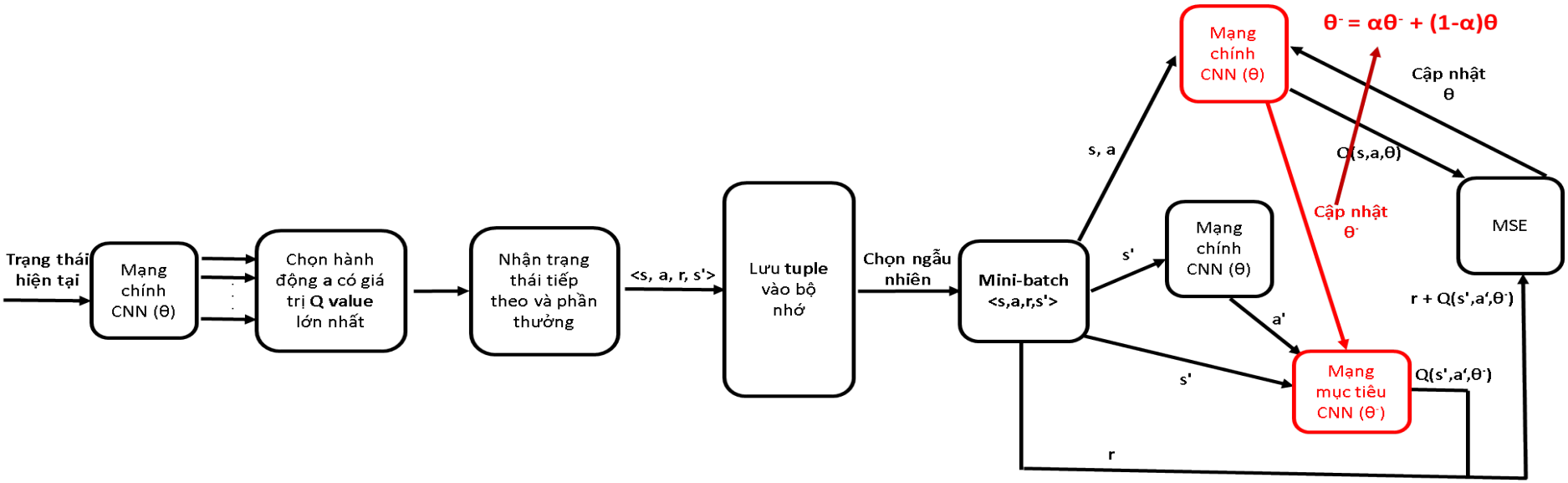
Hình: Kiến trúc tổng quan

Kiến trúc tổng quan



Hình: Kiến trúc tổng quan

Kiến trúc tổng quan

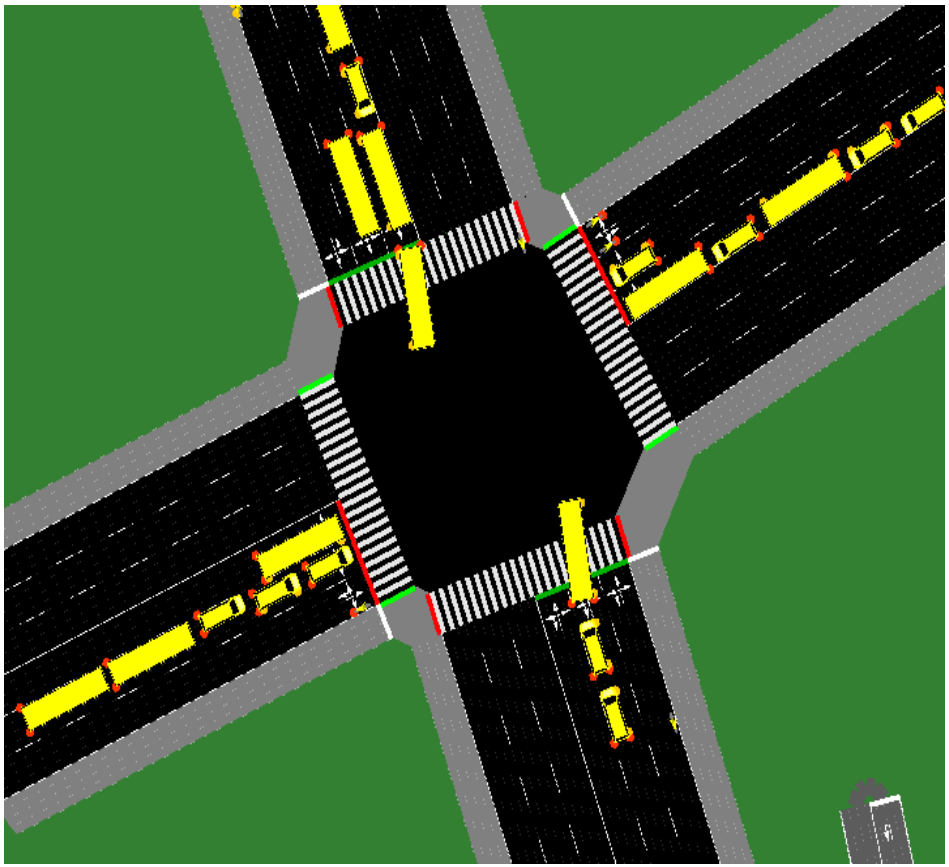


Hình: Kiến trúc tổng quan

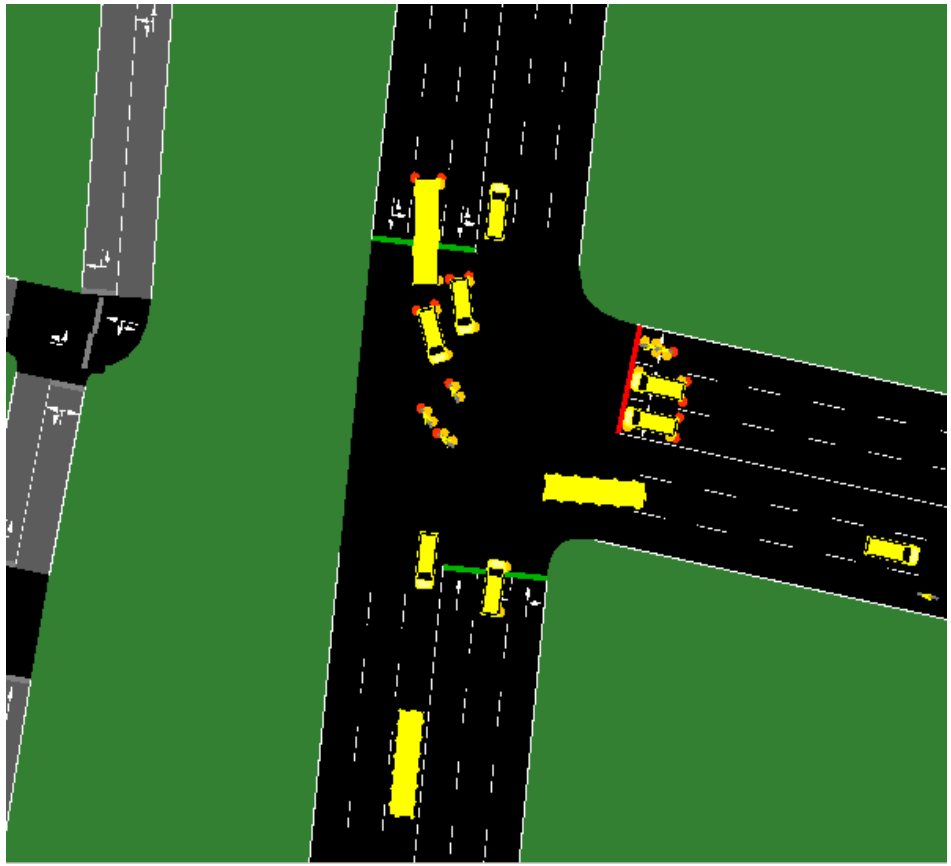
Mục lục

Chiến lược huấn luyện

Xây dựng bản đồ



Hình: Ngã tư



Hình: Ngã ba

Chiến lược huấn luyện

Xây dựng 4 loại kịch bản

Kịch bản	**Mật độ**
LOW	Mật độ xe ít.
HIGH	Mật độ xe nhiều.
EW	Mật độ xe nhiều ở chiều Đông - Tây.
NS	Mật độ xe nhiều ở chiều Bắc - Nam.

Chiến lược huấn luyện

Xây dựng 4 loại kịch bản

Kịch bản	**Mật độ**
LOW	Mật độ xe ít.
HIGH	Mật độ xe nhiều.
EW	Mật độ xe nhiều ở chiều Đông - Tây.
NS	Mật độ xe nhiều ở chiều Bắc - Nam.

2 chiến lược huấn luyện

- Chiến lược 1: mật độ xe **cố định**.
- Chiến lược 2: mật độ xe **ngẫu nhiên trong khoảng nhất định**.

Mục lục

Đánh giá kết quả

Phương pháp đánh giá

- Thời gian chờ trung bình của mỗi xe **trong 1 kịch bản**.

$$AWT = \frac{TWT}{TV}\ (s) \tag{2}$$

Đánh giá kết quả

Bản đồ

- Ngã tư (50m)
- Ngã ba (200m)

Đánh giá kết quả

Bản đồ

- Ngã tư (50m)
- Ngã ba (200m)

Kịch bản

- 4 kịch bản cố định (được dùng để huấn luyện chiến lược 1).
- 5 kịch bản ngẫu nhiên.

Đánh giá kết quả

Bản đồ

- Ngã tư (50m)
- Ngã ba (200m)

Kịch bản

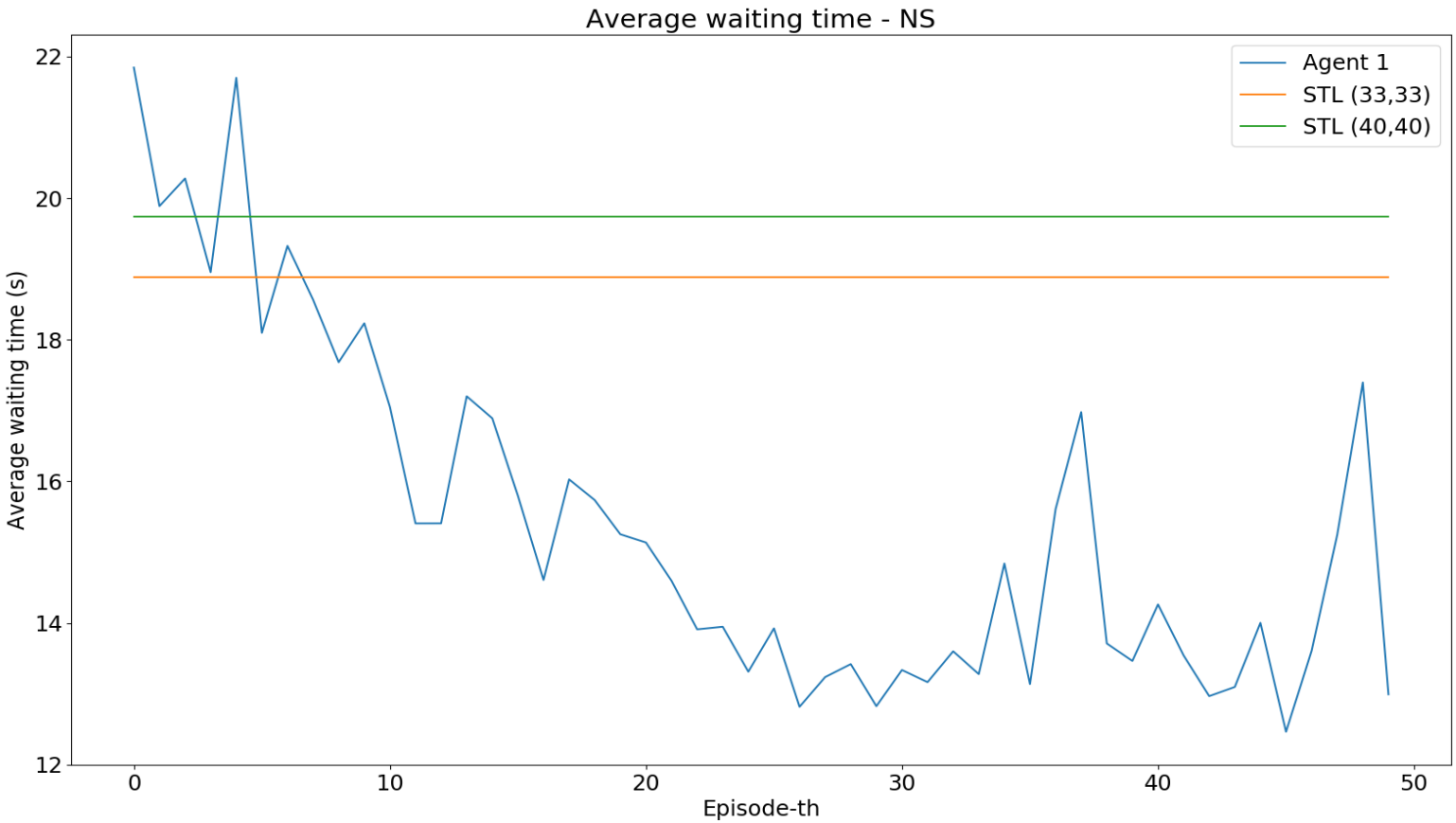
- 4 kịch bản cố định (được dùng để huấn luyện chiến lược 1).
- 5 kịch bản ngẫu nhiên.

Hệ thống đèn giao thông cố định

- **STL (33,33)**: 33s đèn xanh, 4s đèn vàng.
- **STL (40,40)**: 40s đèn xanh, 4s đèn vàng.

Chiến lược 1

Kết quả quá trình huấn luyện ngã tư



Hình: Thời gian chờ trung bình của mỗi xe trong quá trình huấn luyện (NS)

Chiến lược 1

Kết quả quá trình huấn luyện ngã tư

Hình: Thời gian chờ trung bình của mỗi xe trong quá trình huấn luyện (EW)

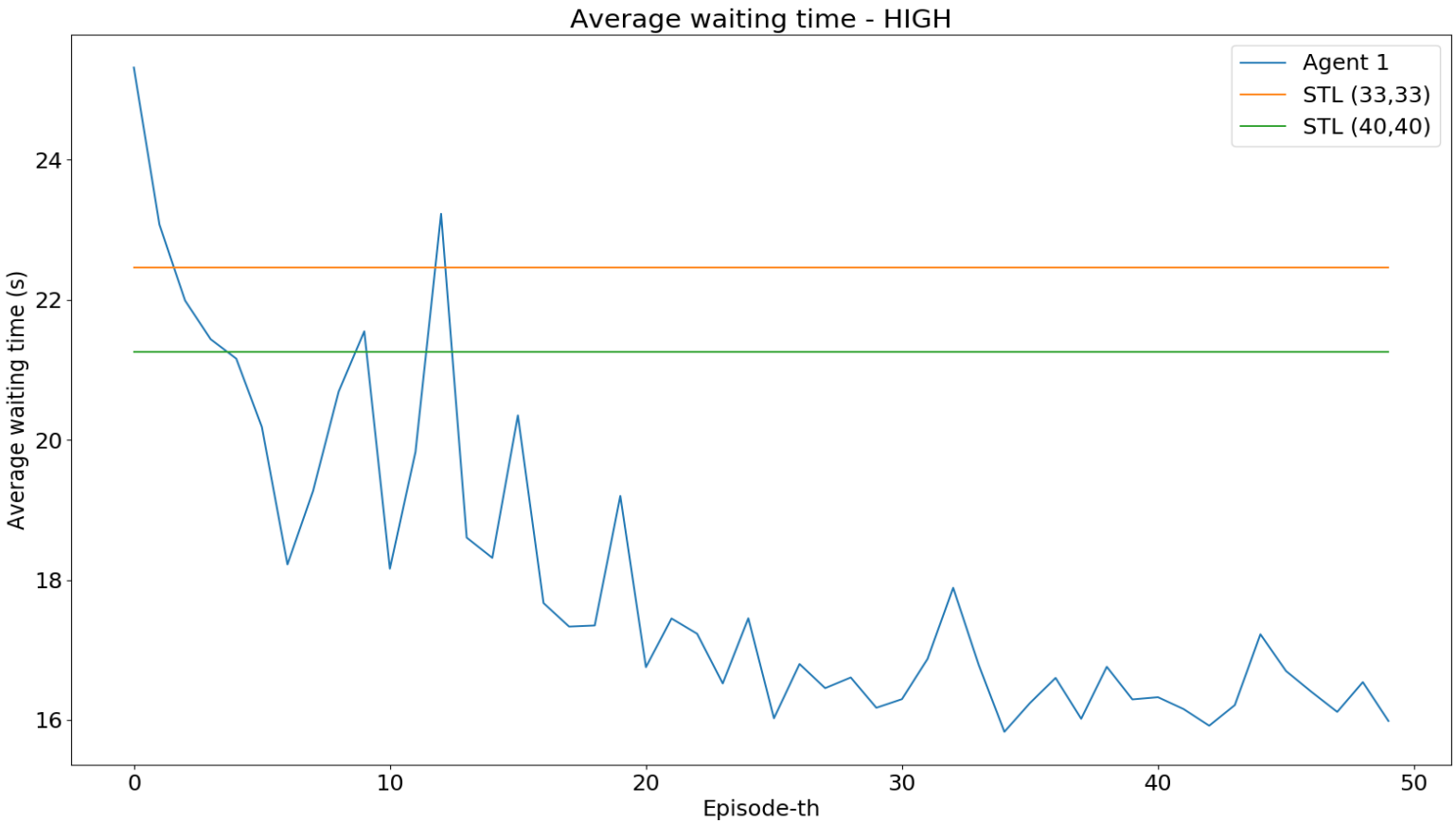
Chiến lược 1

Kết quả quá trình huấn luyện ngã tư

Hình: Thời gian chờ trung bình của mỗi xe trong quá trình huấn luyện (LOW)

Chiến lược 1

Kết quả quá trình huấn luyện ngã tư



Hình: Thời gian chờ trung bình của mỗi xe trong quá trình huấn luyện (HIGH)

Chiến lược 1

	STL (33,33)	**STL (40,40)**	**Agent 1**	**Độ cải thiện**
LOW	11.41	13.41	8.81	23% - 34%
HIGH	20.37	21.48	18.54	10% - 14%
NS	16.92	18.24	14.84	12% - 19%
EW	16.79	18.91	13.14	22% - 31%

Bảng: So sánh AWT của 2 hệ thống STL kịch bản cố định Agent 1

Chiến lược 1

	STL (33,33)	STL (40,40)	Agent 1	Độ cải thiện
LOW	11.41	13.41	8.81	23% - 34%
HIGH	20.37	21.48	18.54	10% - 14%
NS	16.92	18.24	14.84	12% - 19%
EW	16.79	18.91	13.14	22% - 31%

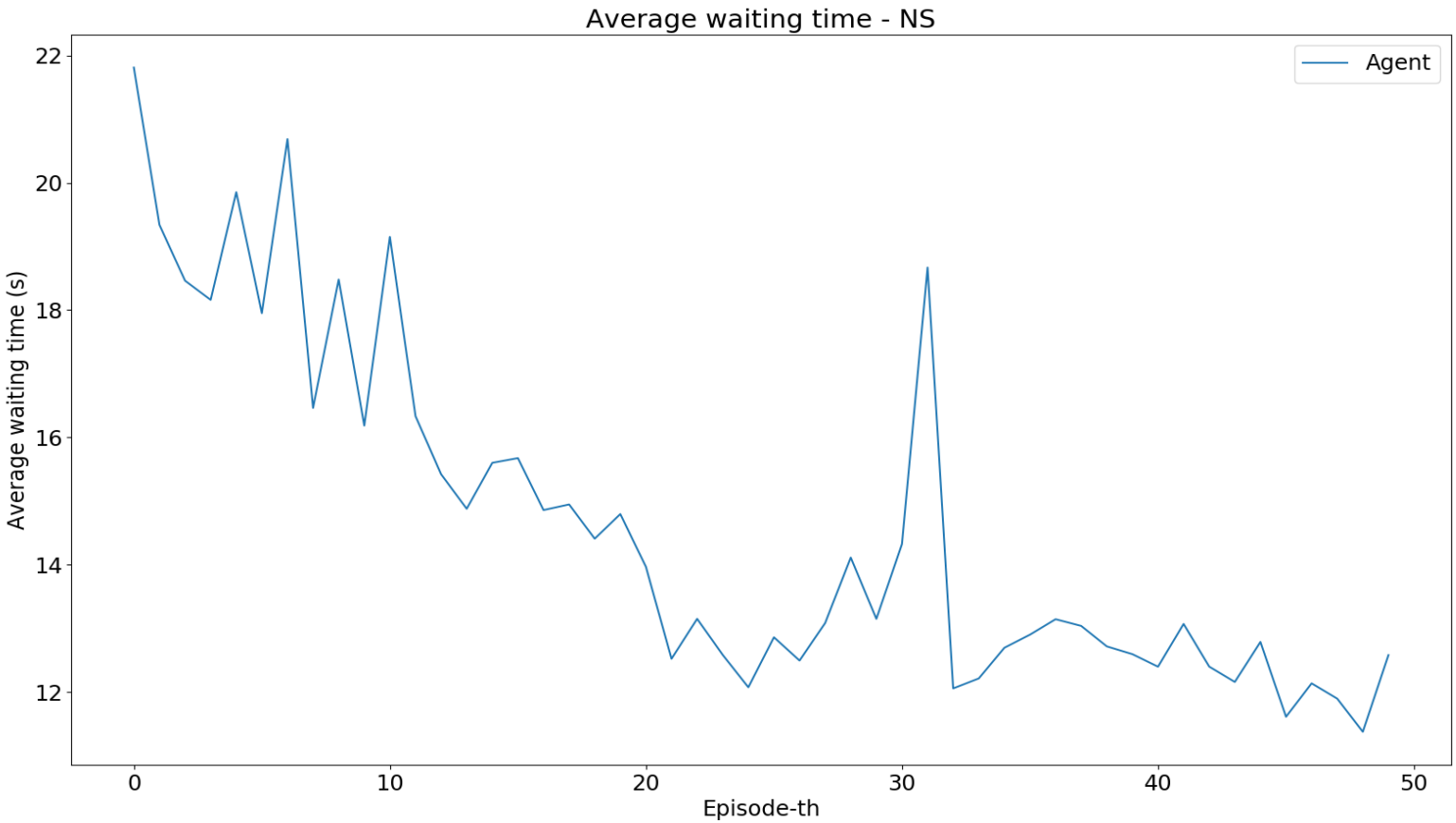
Bảng: So sánh AWT của 2 hệ thống STL kịch bản cố định Agent 1

	STL (33,33)	STL (40,40)	Agent 1	Độ cải thiện
Ngẫu nhiên 1	16.03	18.59	13.47	16% - 28%
Ngẫu nhiên 2	20.12	21.74	17.19	15% - 21%
Ngẫu nhiên 3	16.71	17.87	13.04	22% - 27%
Ngẫu nhiên 4	14.09	15.76	11.95	15% - 24%
Ngẫu nhiên 5	18.17	19.44	16.32	10% - 16%

Bảng: So sánh AWT của 2 hệ thống STL kịch bản ngẫu nhiên Agent 1

Chiến lược 2

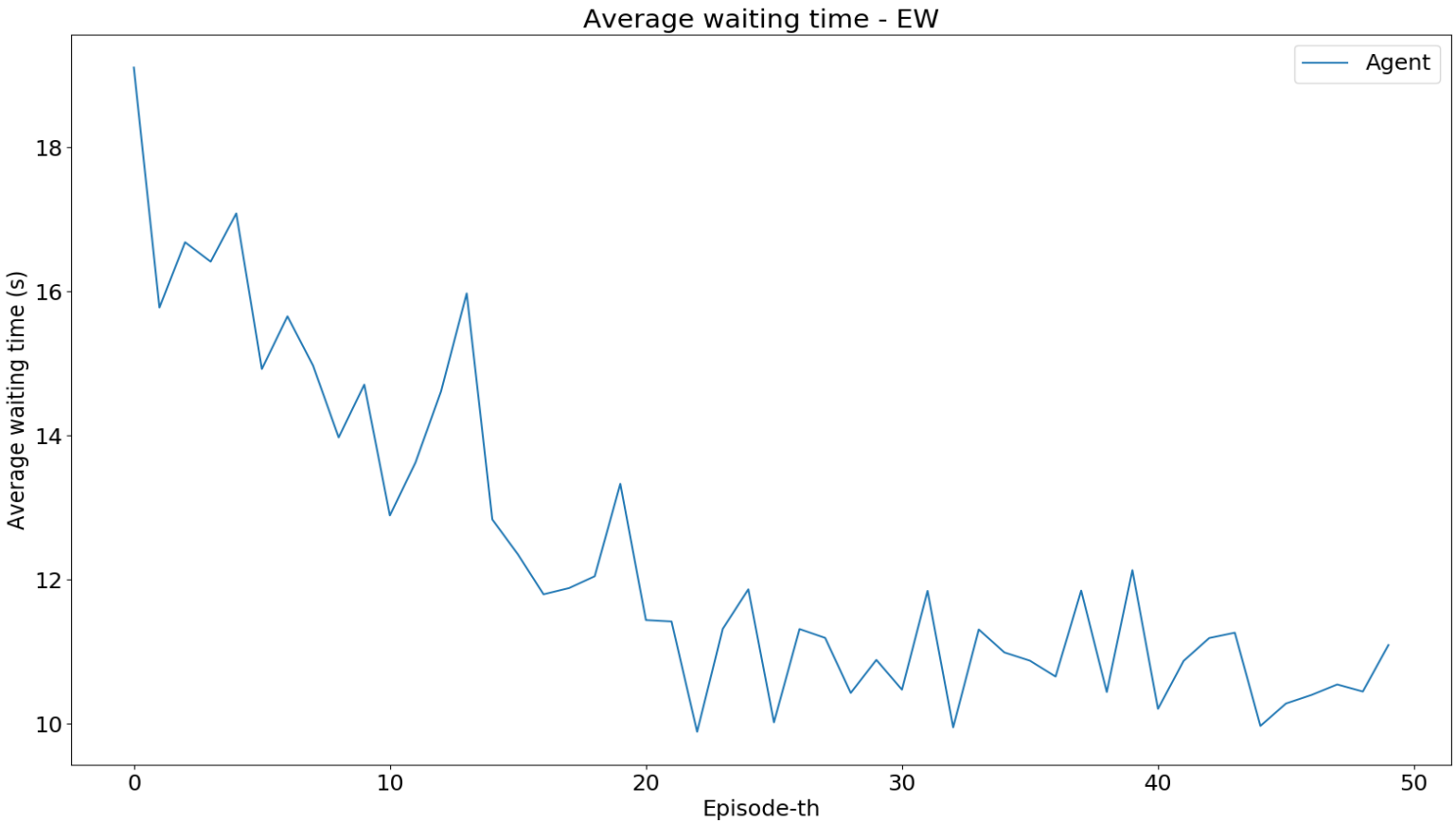
Kết quả quá trình huấn luyện ngã tư



Hình: Thời gian chờ trung bình của mỗi xe trong quá trình huấn luyện (NS)

Chiến lược 2

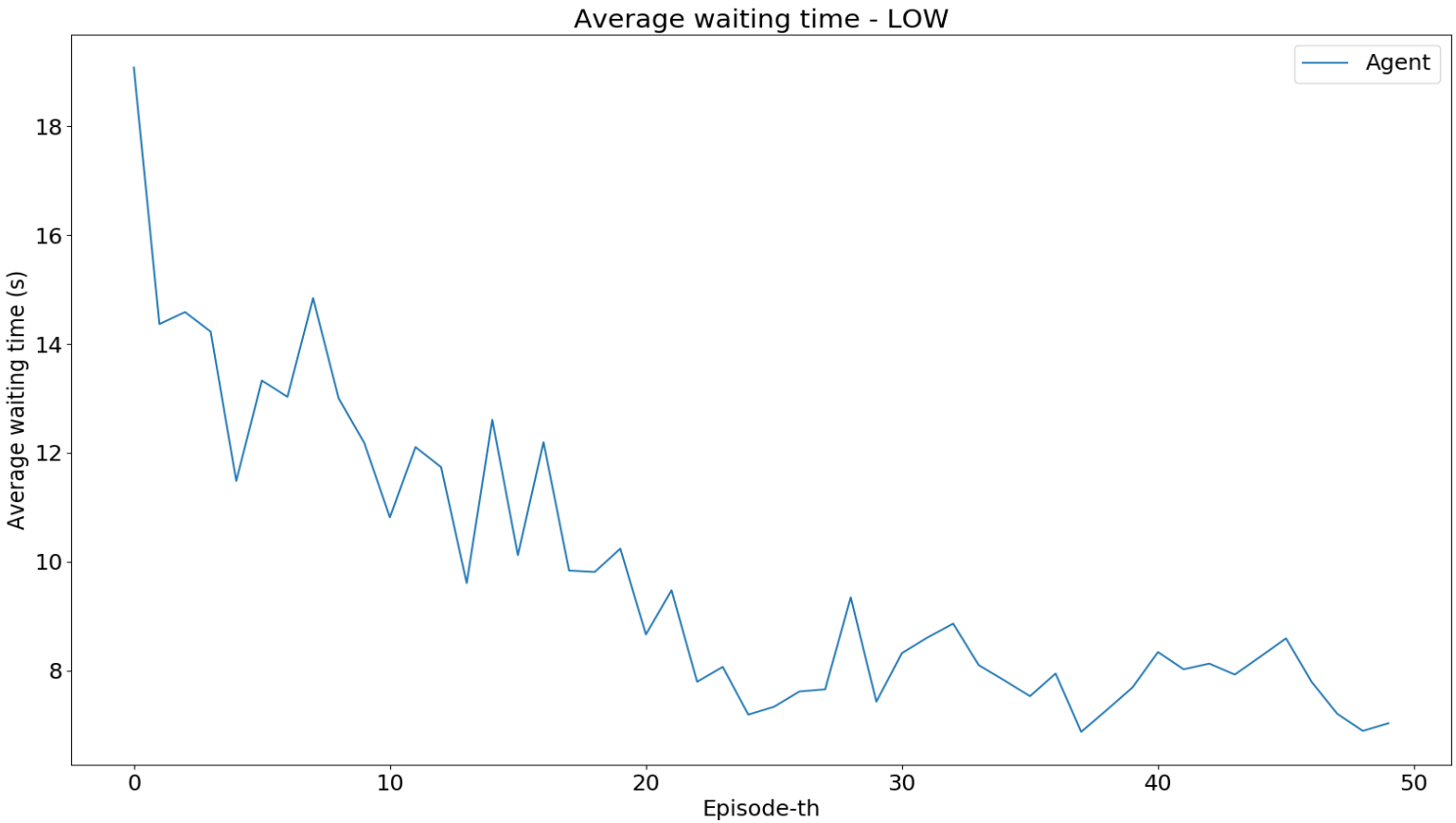
Kết quả quá trình huấn luyện ngã tư



Hình: Thời gian chờ trung bình của mỗi xe trong quá trình huấn luyện (EW)

Chiến lược 2

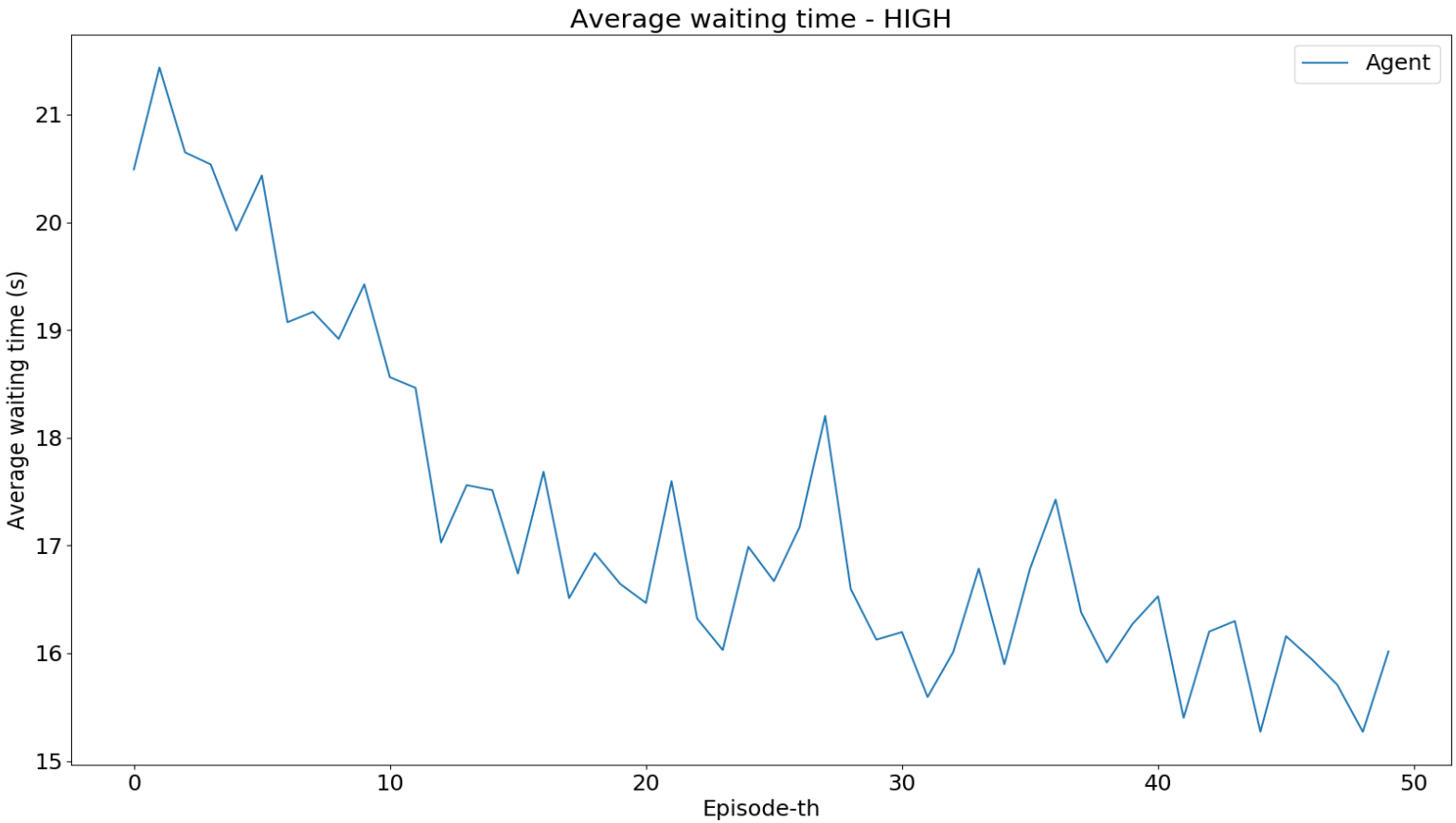
Kết quả quá trình huấn luyện ngã tư



Hình: Thời gian chờ trung bình của mỗi xe trong quá trình huấn luyện (LOW)

Chiến lược 2

Kết quả quá trình huấn luyện ngã tư



Hình: Thời gian chờ trung bình của mỗi xe trong quá trình huấn luyện (HIGH)

Chiến lược 2

	STL (33,33)	**STL (40,40)**	**Agent 2**	**Độ cải thiện**
LOW	11.41	13.41	8.38	26% - 37%
HIGH	20.37	21.48	19.09	6% - 11%
NS	16.92	18.24	13.19	22% - 28%
EW	16.79	18.91	14.46	14% - 24%

Bảng: So sánh AWT của 2 hệ thống STL kịch bản cố định Agent 2

Chiến lược 2

	STL (33,33)	STL (40,40)	Agent 2	Độ cải thiện
LOW	11.41	13.41	8.38	26% - 37%
HIGH	20.37	21.48	19.09	6% - 11%
NS	16.92	18.24	13.19	22% - 28%
EW	16.79	18.91	14.46	14% - 24%

Bảng: So sánh AWT của 2 hệ thống STL kịch bản cố định Agent 2

Kịch bản	STL (33,33)	STL (40,40)	Agent 2	Độ cải thiện
Ngẫu nhiên 1	16.03	18.59	15.24	5% - 18%
Ngẫu nhiên 2	20.12	21.74	16.23	19% - 25%
Ngẫu nhiên 3	16.71	17.87	12.00	28% - 33%
Ngẫu nhiên 4	14.09	15.76	11.34	20% - 28%
Ngẫu nhiên 5	18.17	19.44	16.13	11% - 17%

Bảng: So sánh AWT của 2 hệ thống STL kịch bản ngẫu nhiên Agent 2

Nhận xét

Kịch bản	Agent 1	Agent 2
LOW	23% - 34%	26% - 37%
HIGH	10% - 14%	6% - 11%
NS	12% - 19%	22% - 28%
EW	22% - 31%	14% - 24%
Ngẫu nhiên 1	16% - 28%	5% - 18%
Ngẫu nhiên 2	15% - 21%	19% - 25%
Ngẫu nhiên 3	22% - 27%	28% - 33%
Ngẫu nhiên 4	15% - 24%	20% - 28%
Ngẫu nhiên 5	10% - 16%	11% - 17%

Bảng: So sánh độ cải thiện của 2 chiến lược trên ngã tư

Nhận xét

Kịch bản	Agent 1	Agent 2
LOW	28% - 33%	39% - 43%
HIGH	14% - 16%	18% - 20%
NS	31% - 38%	30% - 37%
EW	58% - 60%	72% - 73%
Ngẫu nhiên 1	38% - 59%	31% - 55%
Ngẫu nhiên 2	3% - 10%	26% - 31%
Ngẫu nhiên 3	49% - 61%	51% - 61%
Ngẫu nhiên 4	68% - 71%	54% - 58%
Ngẫu nhiên 5	13% - 14%	15% - 16%

Bảng: So sánh độ cải thiện của 2 chiến lược trên ngã ba

Kết luận

Kết luận

- Huấn luyện bằng chiến lược 2 (kịch bản ngẫu nhiên) mang lại kết quả tốt hơn chiến lược 1 (kịch bản cố định)
- Độ dài quan sát càng lớn độ cải thiện càng lớn.

Mục lục

Tổng kết

Kết quả đạt được

- Kiến thức về lĩnh vực **Học tăng cường**, **Học máy**.
- Mô phỏng **ngã ba**, **ngã tư** gần giống với giao thông thực tế trên mô phỏng **SUMO**.
- Đưa ra được **mô hình đề xuất** sau khi thử nghiệm, phân tích các mô hình đã tham khảo.
- Mô hình cải thiện khoảng **20%-30%** so với 2 hệ thống STL(33,33) và STL(40,40).
- Mô hình mang tính **tổng quát**.

Tổng kết

Hạn chế

1. **Chưa** áp dụng được kĩ thuật **PER** (prioritized experience replay).
2. Hệ thống **chưa** hiển thị thời gian đèn.
3. Mô hình mới được áp dụng vào hai giao lộ **ngã ba** và **ngã tư**.
4. Chưa mô phỏng được các **kịch bản phức tạp**: vượt đèn đỏ, ưu tiên loại xe khẩn cấp, tình huống có người đi bộ, các xe lấn làn...
5. Chưa áp dụng được trong **thực tế**.

Tổng kết

Kế hoạch tương lai

1. Áp dụng được kĩ thuật **PER** (prioritized experience replay).
2. Tìm một **mô hình hành động khác** thay thế.
3. Áp dụng mô hình cho **ngã 5**, **ngã 6**, **vòng xoay**.
4. Tìm hiểu thêm các tính năng của công cụ SUMO.
5. **Xây dựng hệ thống** trích xuất thông tin từ camera, máy đo tốc độ để áp dụng cho thực tế.

Kết thúc

Cảm ơn quý thầy cô và các bạn đã lắng nghe!