

Phương pháp Học tăng cường sử dụng trong việc chuyển làn đường cho xe tự động

Nguyễn Đàm Trường – Email:18021333@vnu.edu.vn

Khoa Điện tử Viễn thông, Đại học Công Nghệ, Đại học quốc gia Hà Nội

Tóm tắt – Những tiến bộ của các phương pháp học có giám sát và học tăng cường đã mang lại sự thay đổi lớn để áp dụng các phương pháp vào trong việc vận hành một chiếc xe tự lái. Tuy nhiên còn nhiều thách thức trong quá trình vận hành một chiếc xe tự lái trong một môi trường nhiều tác nhân và nhiều tác động của ngoại cảnh. Các thuật toán học tập có giám sát có thể được xây dựng trên một lượng lớn các dữ liệu để chuyển đổi sang các môi trường mới, mặc dù vậy thì việc thu thập dữ liệu mới cho mỗi một môi trường mới rất không thực tế và tốn kém. Học tăng cường có thể giảm thiểu vấn đề phụ thuộc vào dữ liệu. Ở bài báo này, tôi sẽ bàn về một phương pháp là Meta Reinforcement Learning (MRL) trong việc áp dụng vào bài toán chuyển làn cho xe tự lái. Và so sánh MRL với DRL áp dụng trong cùng một bài toán.

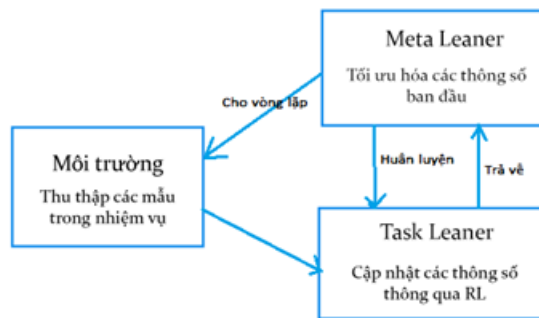
Từ khóa – Học tăng cường, Xe tự lái, Thay đổi làn đường, Điều khiển xe.

I. Giới thiệu

Các phương tiện giao thông tự động hoặc bán tự động đang ngày càng phổ biến trong giao thông vận tải. Ở lĩnh vực xe tự lái đã có những phát triển đáng kể trong suốt thập kỉ qua. Đặc biệt những năm gần đây, những tiến bộ trong học máy có sự ảnh hưởng mạnh mẽ đến lĩnh vực xe tự động. Việc xây dựng các hệ thống tự động áp dụng vào bài toán chuyển làn cho xe tự lái một cách an toàn và hiệu quả là điều rất quan trọng. Điều này đặc biệt ảnh hưởng trong bối cảnh đường cao tốc có nhiều làn xe khi tham gia giao thông và di chuyển với tốc độ cao, vì nếu có bất kỳ sự cố nào xảy ra, thì việc dẫn đến tắc nghẽn, tồi tệ nhất có thể xảy ra tai nạn giao thông nghiêm trọng. Bài toán chuyển làn cho một chiếc xe cần phải cảnh giác quan sát xe trên làn đường của mình và các xe xung quanh trên làn đường mục tiêu, đồng thời thực hiện các hành động phù hợp theo phản ứng bất lợi hoặc hợp tác tiềm ẩn do các xe xung quanh thực hiện. Một nghiên cứu cho thấy rằng khoảng 10% các vụ tai nạn giao thông trên cao tốc là do chuyển làn đường.

Theo nghiên cứu về thao tác chuyển làn tự động có được tiến hành rộng rãi và công việc có thể được chia thành hai loại chức năng: Mô hình ra quyết định và Mô hình thực thi điều khiển. Mô hình ra quyết định có thể được xem như một chiến lược chuyển làn dựa trên một tuyến đường đã lên kế hoạch. Mô hình thực thi khi có hiệu lệnh chuyển làn sẽ thực hiện kiểm soát vận tốc để làm cho xe chuyển làn an toàn, trơn tru và hiệu quả. Các phương pháp Học tăng cường ngày càng áp dụng nhiều hơn trong việc ra quyết định và điều khiển xe tự lái. Ví dụ, Deep Q Network (DQN) được nêu ra để quyết định với việc điều khiển tốc độ tự động khi xe đang chạy trên đường cao tốc. Một mô hình DQN được đề xuất bởi Hoel trong môi trường mô phỏng để thực hiện các lệnh (ví dụ: chuyển làn

sang trái/phải, di chuyển trên làn hiện tại, ...) và nghiên cứu khi thay đổi các yếu tố bên trong mạng nơ-ron khác nhau ảnh hưởng đến hiệu suất của xe tự lái. Trong bài báo này, tôi muốn nói đến một cách tiếp cận Meta Reinforcement Learning tích hợp trong Deep Reinforcement Learning. Giờ đây Meta Reinforcement Learning sẽ được áp dụng để phát triển các chiến lược ra quyết định cho các tình huống chuyển làn trên đường cao tốc. Phương pháp MRL cho phép tạo ra một mô hình chung có thể tổng quát để thực hiện chuyển làn tự động cho xe tự lái trong các môi trường giao thông khác nhau. Để thể hiện sự khái quát hóa thì trong bài báo này sẽ đề cập đến mô hình model-agnostic meta learning (MAML).



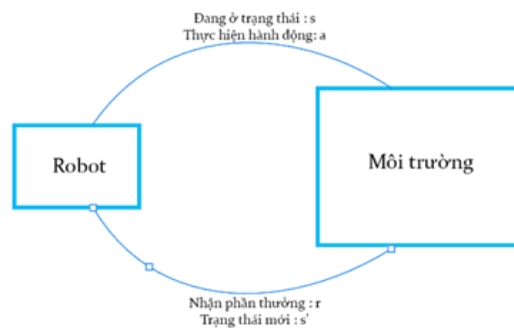
Hình 1.1: Sơ đồ MAML

Các phần tiếp theo của bài báo sẽ được sắp xếp như sau: Phần II Nền tảng lý thuyết, Phần III Phương pháp MRL áp dụng vào chuyển làn xe tự động, Phần IV Mô phỏng , Phần V So sánh và Kết Luận, Phần VI Tài liệu tham khảo.

II. Nền tảng lý thuyết

a. Reinforcement Learning

Giả sử, có một con Robot trong một môi trường không biết trước và con robot này có thể nhận được một số phần thưởng khi tương tác với môi trường. Con robot sẽ phải tối ưu hóa phần thưởng tích lũy nhận được.



Hình 2.1: Robot tương tác với môi trường và tối đa hóa phần thưởng tích lũy

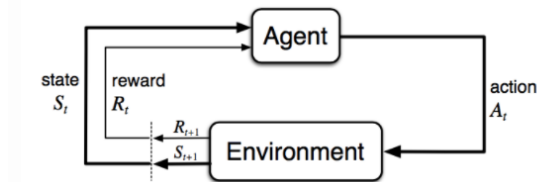
b. Markov Decision Processes

Hầu hết các vấn đề của Reinforcement Learning có thể được gọi là Markov Decision Processes (MDPs). Tất cả các trạng thái trong MDP đều có tính “Markov”, tức là tương lai chỉ phụ thuộc vào trạng thái hiện tại không phụ thuộc vào quá khứ.

$$\mathbb{P}[S_{t+1}|S_t] = \mathbb{P}[S_{t+1}|S_1, \dots, S_t]$$

P là xác suất chuyển đổi giữa hai trạng thái

Nói cách khác là tương lai và quá khứ độc lập có điều kiện với hiện tại và trạng thái hiện tại tập trung tất cả các số liệu để có thể đưa ra quyết định trong tương lai.



Hình 2.2: Sự tương tác của tác nhân với môi trường trong một quá trình Markov
(Nguồn ảnh: Sec. 3.1 Sutton & Barto (2017).)

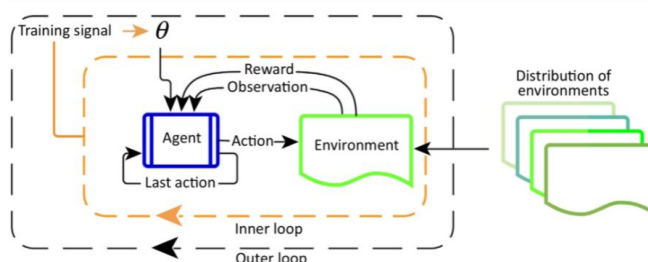
c. Meta-Learning

Meta – Learning hay còn được biết đến với cái tên Learning to Learn được mong đợi để thiết kế các mô hình có thể học được các kỹ năng mới và thích ứng với môi trường nhanh chỉ trong một vài ví dụ đào tạo. Một số ví dụ về nhiệm vụ của Meta – Learning:

- + Một mô hình phân loại đào tạo trên tập dữ liệu không phải con mèo sẽ biết được một bức ảnh có con mèo không sau khi cho biết một vài bức ảnh có mèo.
- + Một bot làm chủ được trò chơi mới một cách nhanh chóng

d. Meta Reinforcement Learning

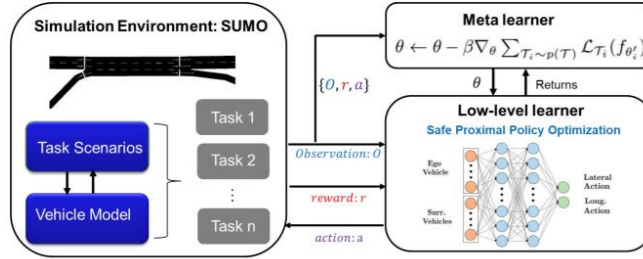
Meta Reinforcement Learning là thực hiện Meta–Learning trong lĩnh vực học tăng cường. Thông qua các nhiệm vụ đào tạo và kiểm tra là khác nhau nhưng được rút ra từ cùng một nhóm các vấn đề.



Hình 2.3: Minh họa về MRL với hai vòng lặp tối ưu hóa. Vòng lặp bên ngoài lấy mẫu môi trường trong mỗi lần lặp và Vòng bên trong để đưa ra thông số với môi trường và tối ưu thông số cho việc nhận thưởng
(Nguồn ảnh : Botvinick, et al. 2019)

III. Phương pháp MRL áp dụng vào chuyển làn xe tự động

A. Tổng quan



Hình 3.1 Cấu trúc chuyển làn xe (Nguồn ảnh: Fig.2 Fei Ye- Pin Wang(2020))

a. Task Learner (Vòng lặp trong)

Nhiệm vụ của vòng lặp trong đưa ra các quyết định chuyển làn tự động có tính năng an toàn và hiệu quả. Sử dụng thuật toán Reinforcement Learning với:

1. Action Space

Không gian hành động được thiết kế theo gia tốc nghiêng của xe (a_{yaw} -yaw acceleration) để đảm bảo xe có thể chuyển làn trơn tru không gặp nhiều dao động do quá trình lấy mẫu khi điều khiển thực tế.

$$a = a_{yaw} \in A$$

2. State Space

Không gian trạng thái của xe khi chuyển làn không chỉ phụ thuộc vào mô hình động học của xe mà còn phụ thuộc vào hình dạng của đường chạy. Tức là xe có thể chuyển làn trên đường thẳng hoặc trên đường cong. Không gian trạng thái của xe gồm: vận tốc của xe v , gia tốc dài a , vị trí (x, y) , góc nghiêng θ , làn mục tiêu id , chiều rộng của làn đường w , độ cong của làn đường c .

$$s = (v, a, x, y, \theta, id, w, c) \in S$$

3. Hàm đánh giá

Trong việc chuyển làn đường chúng ta phải chú ý đến sự an toàn, trơn tru và hiệu quả vì vậy chúng ta đánh giá trên 3 thông số gia tốc nghiêng, đánh giá gia tốc nghiêng và thời gian:

$$R = \sum_{i=1}^N (r_{acce})_i + \sum_{i=1}^N (r_{rate})_i + \sum_{i=1}^N (r_{time})_i$$

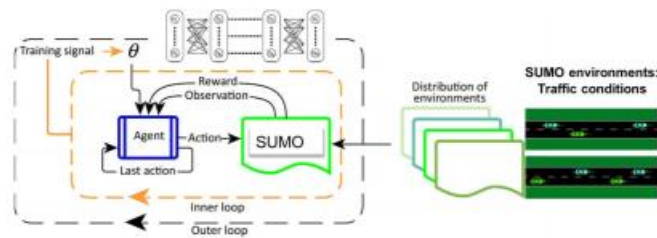
Nếu hàm đánh giá có giá trị âm, thì có thể được coi là một hành động phạt vì tác động bất lợi cho hành động chuyển làn, ví dụ như chuyển làn không trơn tru và kém hiệu quả. Bằng cách này chiếc xe có thể học được cách tránh các hành động dẫn đến hình phạt lớn.

4. Q-Learning

Q-Learning là một kỹ thuật học tăng cường không có mô hình được sử dụng để tìm ra các hành động tối ưu thông qua Q-function được thể hiện bằng cách ước tính giá trị của tổng phần thưởng mà không cần đợi đến khi kết thúc để tích lũy phần thưởng.

b. Meta-Learner (Vòng lặp ngoài)

Meta-learner cho phép điều chỉnh hiệu quả của quá trình thay đổi làn đường theo các tình huống mới như là mật độ tham gia giao thông khác nhau, ảnh hưởng từ môi trường khác nhau. Meta-Learner sẽ tối ưu hóa các tham số nhận được từ Task Learner và sẽ tương tác lại với môi trường.



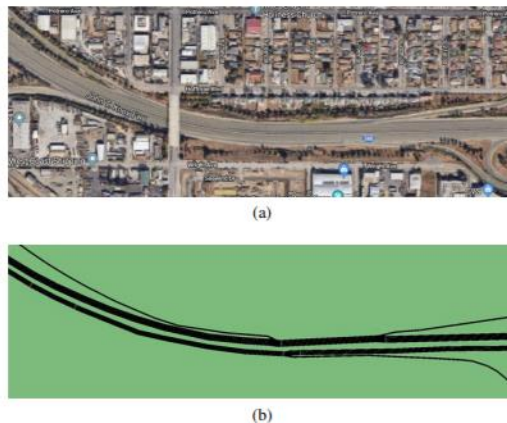
Hình 3.2: Tương tác giữa môi trường mô phỏng và meta learner

Các nhiệm vụ khác nhau được xây dựng với các mô hình mật độ giao thông khác nhau. Khi đào tạo, các nhiệm vụ đào tạo được rút ra từ những lưu lượng thấp và trung bình. Và khi thử nghiệm, sẽ được đào tạo trên những mô hình mật độ giao thông dày đặc.

IV. Mô Phỏng

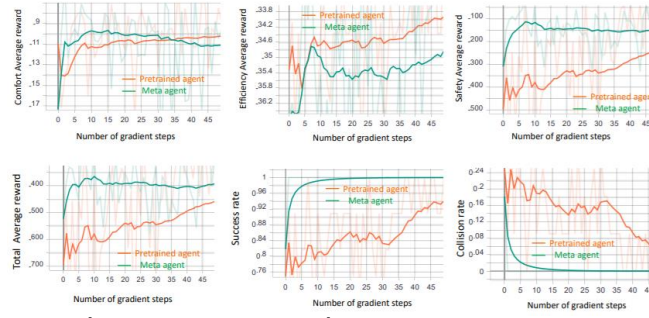
Mô phỏng được thể hiện trên cao tốc 3 làn và chiều dài cao tốc là 800m chiều rộng của mỗi làn là 3.75m (như hình 4)

Meta agent: sử dụng mạng đào tạo meta learner với hệ số khởi tạo ban đầu, khi có nhiệm vụ mới từ môi trường, meta agent có thể điều chỉnh thông số thông qua quá trình xử lý meta-learning.



Hình 4 Mô phỏng (Nguồn ảnh: Fei Ye và Pin Wang (2020))

V. So sánh và kết luận



Hình 5: Kết quả thích nghi (Nguồn ảnh: Fei Ye và Pin Wang (2020))

Bảng 3			
Trung bình tỷ lệ thành công của hai mô hình			
Mô hình	Trung bình tỷ lệ thành công		
	Sau 5 bước	Sau 20 bước	Sau 45 bước
Mô hình DRL	78%	86%	92%
Mô hình MRL	98%	100%	100%

Bảng 4			
Trung bình tỷ lệ va chạm của hai mô hình			
Mô hình	Trung bình tỷ lệ va chạm		
	Sau 5 bước	Sau 20 bước	Sau 45 bước
Mô hình DRL	20%	14%	6%
Mô hình MRL	2%	0%	0%

Từ bảng 3 và bảng 4 ta nhận ra rằng với so với sử dụng MRL với DRL thông thường có lợi thế hơn rất nhiều khi cả về hiệu suất, độ an toàn và trơn tru sau 20 lần gradient thì tỷ lệ chuyển đường thành công của MRL là 100% và của DRL chỉ 86%. Ngoài ra DRL bị 14% tỷ lệ va chạm lớn hơn nhiều so với MRL đã chỉ còn 0% sau 20 lần gradient.

Trong bài báo này, tôi đã chỉ ra rằng khi áp dụng phương pháp MRL đã có kết quả tốt hơn rất nhiều so với việc chỉ sử dụng DRL trong tốc độ thích ứng với môi trường và hiệu suất của nhiệm vụ.

VI. Tài liệu tham khảo

- [1] Hoel, Wolff and Laine, “Automated speed and lane change decision making using deep reinforcement learning,” *Proc. Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov 2018.
- [2]. Tsendsuren Munkhdalai and Hong Yu. “Meta Networks.” *ICML*. 2017.
- [3]. Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-agnostic meta-learning for fast adaptation of deep networks.” *ICML* 2017.
- [4]. Fei Ye , Pin Wang , Ching-Yao Chan and Jiucui Zhang. “Meta Reinforcement Learning-Based Lane Change Strategy for Autonomous Vehicles”. 2020.
- [5]. Pin Wang, Ching-Yao Chan , Arnaud de La Fortelle. “A Reinforcement Learning Based Approach for Automated Lane Change Maneuvers”. *IEEE* 2018.
- [6]. Slide on courses UET by Nguyen The Hoang Anh.