

Bài tập về nhà

Buổi số: 12

Bài 1: Học tăng cường (reinforcement learning) là một trong những lĩnh vực của học máy nhận được nhiều sự quan tâm trong những năm gần đây.

a/ Dựa trên những kiến thức đã học trên lớp và đọc thêm tài liệu ở nhà, em hãy trình bày những ứng dụng của học tăng cường.

b/ Em hãy lựa chọn một ứng dụng cụ thể của học tăng cường mà em quan tâm hoặc thích nhất, sau đó xác định các thành phần giúp em xây dựng bài toán học tăng cường như: environment (môi trường), agent (tác tử), actions (các hành động), rewards (các phần thưởng).

Bài 2: Dựa trên những kiến thức đã học trên lớp và đọc thêm tài liệu ở nhà, em hãy trình bày những ưu điểm và nhược điểm của học tăng cường. Theo em trong những trường hợp nào chúng ta nên sử dụng học tăng cường và trong những trường hợp nào chúng ta không nên sử dụng học tăng cường.

Bài 3: (Giải bài toán bằng giấy và bút chì)

Trong hệ thống IoT nhà thông minh, quản lý năng lượng là một trong những khía cạnh quan trọng. Các thiết bị và đồ dùng trong nhà nên được bật hoặc tắt để tối ưu hóa mức tiêu thụ năng lượng trong khi vẫn duy trì sự thoải mái cho cư dân. Trạng thái của mỗi thiết bị có thể được biểu diễn bằng chuỗi Markov, trong đó thiết bị thay đổi trạng thái dựa trên các chuyển đổi xác suất theo thời gian.

Giả sử chúng ta chỉ xem xét một hệ thống nhà thông minh rất đơn giản (ở miền núi phía Bắc vào mùa đông ☺) với một lò sưởi hoạt động ở ba trạng thái:

- Tắt: Lò sưởi tắt,
- Thấp: Lò sưởi đang chạy ở chế độ cài đặt mức thấp,
- Cao: Lò sưởi đang chạy ở chế độ cài đặt mức cao.

Lò sưởi chuyển đổi giữa các trạng thái này dựa trên các chỉ số nhiệt độ từ môi trường (đo lường bởi sensor) và cài đặt mức độ thoải mái của người dùng. Mỗi trạng thái có khả năng chuyển sang trạng thái khác (hoặc ở duy trì trạng thái đó) trong khoảng thời gian tiếp theo. Biết rằng xác suất chuyển đổi của lò sưởi cho như trong Bảng 1, em hãy:

a/ Vẽ chuỗi Markov biểu diễn ba trạng thái của lò sưởi và xác suất chuyển đổi của chúng. Chú ý: Cần ghi rõ nhãn cho từng trạng thái (Tắt, Thấp, Cao) và hiển thị xác suất chuyển đổi giữa các trạng thái.

b/ Phân tích hoạt động chuyển đổi trạng thái:

- Nếu lò sưởi bắt đầu ở trạng thái "Thấp", thì xác suất lò sưởi ở trạng thái "Cao" sau hai khoảng thời gian là bao nhiêu?
- Tính xác suất lò sưởi sẽ vẫn ở trạng thái "Tắt" sau hai khoảng thời gian liên tiếp nếu lò sưởi bắt đầu ở trạng thái "Tắt".

c/ Dựa trên xác suất chuyển đổi, em hãy thảo luận ngắn gọn về cách hệ thống nhà thông minh có thể quản lý cài đặt lò sưởi để tiết kiệm năng lượng trong khi vẫn duy trì sự thoải mái.

Bảng 1: Xác suất chuyển đổi trạng thái của lò sưởi

Trạng thái hiện tại	Chuyển sang: Tắt	Chuyển sang: Thấp	Chuyển sang: Cao
Tắt	0,6	0,3	0,1
Thấp	0,2	0,5	0,3
Cao	0,1	0,3	0,6

Bài 4¹: (Thực hành với Python)

Một trong những thách thức của học tăng cường là để đào tạo một tác tử (agent), trước tiên ta cần phải có một môi trường làm việc (working environment). Nếu ta muốn lập trình một tác tử sẽ học cách chơi một trò chơi (ví dụ như: Atari, Mario, hay Flappy Bird, v.v.), ta sẽ cần một trình mô phỏng trò chơi đó. OpenAI Gym là bộ công cụ cung cấp nhiều môi trường mô phỏng khác nhau (trò chơi Atari, trò chơi cờ, mô phỏng vật lý 2D và 3D, v.v.) giúp ta có thể đào tạo các tác tử, so sánh các tác tử hoặc phát triển các thuật toán RL mới.

a/ Em hãy tiến hành cài đặt và tìm hiểu cách sử dụng OpenAI Gym.

b/ Tìm hiểu về môi trường Cart Pole (trong mục Environments\Classic Control\Cart Pole)². Môi trường này tương ứng với phiên bản của bài toán xe đẩy-cột/con lắc được Barto, Sutton và Anderson mô tả trong bài báo “Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problem”³. Một cột/con lắc được gắn bằng một khớp không có cơ cấu truyền động vào một xe đẩy,

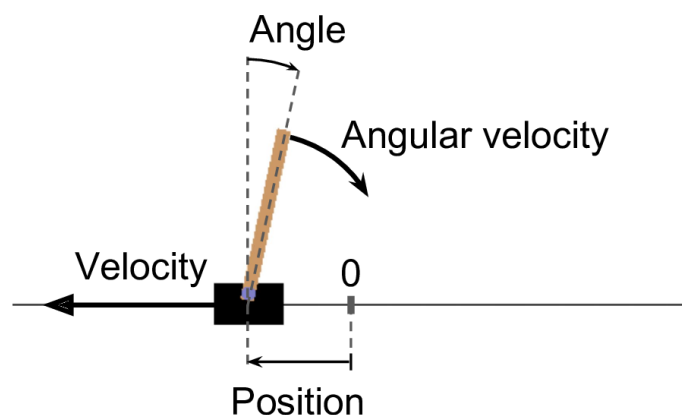
¹ Sinh viên tham khảo thêm về OpenAI Gym tại: <https://gymnasium.farama.org/index.html>

² Sinh viên tìm hiểu về môi trường Cart Pole tại:

https://gymnasium.farama.org/environments/classic_control/cart_pole/

³ Bài báo của Barto, Sutton và Anderson: <https://ieeexplore.ieee.org/document/6313077>

xe đẩy di chuyển dọc theo một đường ray không ma sát (xem minh họa trong Hình vẽ 1). Con lắc/cột được đặt thẳng đứng trên xe đẩy và mục tiêu là cân bằng con lắc/cột bằng cách tác dụng lực theo hướng trái và phải lên xe đẩy.



Hình vẽ 1: Minh họa mô hình bài toán xe đẩy-cột/con lắc. Con lắc/cột được đặt thẳng đứng trên xe đẩy. Mục tiêu của bài toán là cân bằng con lắc/cột bằng cách tác dụng lực theo hướng trái và phải lên xe đẩy.