Nguyễn Khắc Sơn – MSSV: 21085691

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?*

* Trí tuệ nhân tạo trong trò chơi
* Robot học (Robotic)
* Quản lý tài nguyên và tối ưu hệ thống
* Chăm sóc sức khỏe
* Hệ thống đề xuất

*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:* ***enviroment (môi trường), agent (tác tử), actions (các hành động), rewards (các phần thưởng)?***

Lựa chọn ứng dụng cụ thể: **Xe tự lái**

* **Environment (Môi trường):**
  + Mô phỏng môi trường giao thông bao gồm các yếu tố như đường đi, các phương tiện khác, tín hiệu giao thông và các điều kiện thời tiết
  + Trong thực tế, có sử dụng các mô phỏng như CARLA Simulator hoặc môi trường thực tế với các cảm biến trên xe
* **Agent (tác tử):**
  + Tác tử trong trường hợp này là hệ thống điều khiển xe tự lái, chịu trách nhiệm quyết định hành động dựa trên trạng thái hiện tại cửa môi trường.
* **Action (Các hành động):**
  + Các hành động mà xe có thể thực hiện bao gồm: tăng tốc, giảm tốc, rẽ phải, trái,…
* **Rewards (Phần thưởng):**
  + “ + ” : nếu xe di chuyển đúng làn và ko xảy ra va chạm
  + “ - “ : nếu xe va chạm, vượt đèn, lạng lách hay chậm trễ

**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.**

|  |  |
| --- | --- |
| **Ưu điểm** | **Nhược điểm** |
| * **Ưu điểm của học tăng cường:** RL cho phép tác tử học từ quá trình tương tác trực tiếp với môi trường, không cần dữ liệu gán nhãn trước. * **Thích nghi và tối ưu hóa trong môi trường động:** RL có khả tự động thích nghi với các thay dổi trong môi trường. * **Khả năng giải quyết bài toán phức tạp:** RL đã chứng minh hiệu quả trong việc giải quyết các bài toán mà không có chiến lược rõ ràng bàn đầu. * **Tốc ưu hóa mục tiêu dài hạn:** RL không chỉ tập trung vào các quyết định cục bộ, mà còn tối ưu hóa cho mục tiêu dài hạn qua khái niệm phần thưởng tính lũy. | * **Yêu cầu nhiều tài nguyên:** thường yêu cầu nhiều thời gian và tài nguyên tính toán do cần thực hiện nhiều lần tương tác với môi trường * **Khó khăn trong việc thiết kế phần thưởng:** Việc thiết kế hàm phần thưởng phù hợp là một thách thức lớn. Nếu phần thưởng không được thiết kế cẩn thận, tác tử có thể học các hành vi không mong muốn. * **Không đảm bảo hội tụ:** có thể không hội tụ đến tối ưu, đặc biệt là trong môi trường phức tập có tính ngẫu nhiên cao * **Phụ thuộc vào chất lượng môi trường mô phỏng:** Kết quả của RL phụ thuộc vào môi trường mô phỏng. Nếu môi trường không phản ánh chính xác thế giới thực -> mô hình RL sẽ không hoạt động tốt khi truyển khai thực tế |

**Khi nào nên sử dụng:**

Khi có môi trường mô phỏng hoặc thực tế để tác tử có thể tương tác, phù hợp khi các hành động hiện tại ảnh hưởng đến kết quả trong tương lai xa, như trong việc điều khiển xe tự lái hoặc quản lý tài nguyên,… trong các tình huống như giao dịch tài chính hoặc tối ưu hóa hệ thống năng lượng, RL có thể học trực tiếp từ tương tác hiện tại.

**Khi nào không nên sử dụng:**

Khi bài toán có sẵn dữ liệu nhãn lớn và rõ ràng, khi môi trường không thể mô phỏng hoặc chi phí tương tác cao, khi vấn đề quá đơn giản và không cần tối ưu hóa dài hạn,…