Công thức ở trên mô tả quy tắc chọn hành động tối ưu dựa trên tham số lợi ích dự kiến được tối ưu hóa. Công thức được xây dựng trên lý thuyết quyết định hợp lý. Thuyết quyết định hợp lý (Rational Decision Theory) dựa trên giả định rằng người ra quyết định là “hợp lý” nghĩa là họ sẽ chọn hành động sao cho hành động đó mang lại lợi ích cao nhất.

Trong bối cảnh AI, như với xe tự lái, việc áp dụng lý thuyết này có phần bị hạn chế bởi các yếu tố như dữ liệu không hoàn chỉnh, hoặc môi trường khi chạy trên đường có sự phức tạp trong việc đưa ra các hành động cụ thể.

Trong công thức trên ta có:

* **a**: Đại diện cho hành động (hay còn gọi là action) được chọn.
* **argmax**: Điểm đối số khi đạt giá trị lớn nhất (argument of the maximun). Mục đích là để tìm giá trị đối số cho một hành động cụ thể làm cho hàm bên trong đạt giá trị lớn nhất. Ví dụ như ta có nhiều phương án chạy trên đường đi, argmax sẽ chọn ra phương án mang lại hiệu quả cao nhất.
* **A**: Tập hợp các hành động hay các phương án (action space). Một tập hợp các hành động mà có thể thực hiện trong một tình huống nhất định (đối với xe thì sẽ là trái, phải, tiến, lùi, dừng lại).
* **E(U | a)**: Lợi ích được dự kiến sẽ được chọn (Expected Utility). Đây là giá trị số học đại diện cho mức độ đánh giá cho hành động đó (bao gồm tính xác xuất xảy ra các kết quả khác nhau và xem xét các yếu tố như an toàn, hiệu quả, chi phí,…).

Cách thức mà nó được tính toán như sau:

* Đầu tiên là liệt kê tất cả các hành động khả thi trong 1 tình huống cụ thể (tập hợp A).
* Tính toán E(U | a) bằng cách ước lượng các kết quả từ hành động đó mang lại, gán xác xuất cho từng kết quả.
* Gán giá trị mà các yếu tố mang lại (Utility).
* Áp dụng argmax, so sánh và chọn ra a có giá trị EU lớn nhất.

Ứng dụng:

* Lái xe tự động.
* Robot học và điều khiển.
* Trò chơi và Học máy.
* Kinh tế và Thương mại điện tử.
* Y tế.

**Câu hỏi: Nếu chúng ta có hai xe, và một xe cung cấp nhiều lợi ích dự kiến hơn. Xe nào là hợp lý?**

Xe cung cấp nhiều lợi ích dự kiến (EU) hơn là xe hợp lý Vì nếu giá trị EU lớn hơn thì xe sẽ mang lại lợi ích cao hơn hoặc xe kia có thể hợp lý nếu theo ngữ cảnh như xe tiết kiệm năng lượng hơn,..

**Câu hỏi: Một xe tự lái hợp lý có thể bị liên quan đến tai nạn không?**

Một xe tự lái hợp lý vẫn có thể bị liên quan đến tai nạn, ngay cả khi nó luôn chọn hành động tối đa hóa. Vì:

* EU dựa trên xác suất dự kiến, không phải chắc chắn 100%. Ví dụ, xe có thể chọn rẽ phải với EU cao (dựa trên dữ liệu đã từng đi qua), nhưng một sự cố ngẫu nhiên (như người đi bộ bất ngờ băng qua) vẫn xảy ra.
* Các yếu tố không lường trước (thời tiết xấu, lỗi cảm biến) có thể làm EU tính toán sai.

**Câu hỏi: Xe tự lái sẽ khám phá và học hỏi như thế nào?**

Xe tự lái khám phá và học hỏi thông qua học tăng cường (reinforcement learning - RL), kết hợp với công thức trên.

Thỉnh thoảng chọn hành động ngẫu nhiên để thử nghiệm. Ví dụ như sử dụng epsilon-greedy nhằm thu thập dữ liệu mới về môi trường.

Xe có thể thử đổi làn ở tốc độ cao để học EU thực tế, sau đó điều chỉnh để tránh tai nạn. Dữ liệu từ hàng triệu km lái xe (fleet learning) giúp cải thiện toàn bộ hệ thống.

**Câu hỏi: Tính hợp lý có giới hạn nghĩa là gì đối với xe tự lái?**

Tính hợp lý có giới hạn (bounded rationality) nghĩa là xe tự lái không thể đạt được tính hợp lý hoàn hảo (perfect rationality theo argmax EU) do các ràng buộc thực tế như thời gian tính toán hạn chế, dữ liệu không đầy đủ, hoặc không gian hành động quá lớn, bằng cách sử dụng xấp xỉ thay cho việc tính hết toàn bộ.

**Fully observable:**

Môi trường được coi là hoàn toàn có thể quan sát nếu các cảm biến của agent luôn cung cấp đầy đủ thông tin về trạng thái hiện tại của môi trường.

**Partially observable:**

Môi trường chỉ có thể quan sát một phần, nghĩa là đại lý chỉ nhận được một phần thông tin trạng thái và phải dựa vào bộ nhớ hoặc suy luận để hiểu phần còn lại.

**Deterministic:**

Môi trường xác định, tức là:

- Các cảm biến cung cấp thông tin 100% đáng tin cậy.

- Các thay đổi trong môi trường được xác định hoàn toàn bởi trạng thái hiện tại và hành động của đại lý, không có yếu tố ngẫu nhiên.

**Stochastic:**

Môi trường ngẫu nhiên, tức là:

- Các cảm biến không đáng tin cậy.

- Chuyển đổi trạng thái tuân theo mô hình ngẫu nhiên (stochastic model), dẫn đến các xác suất chuyển đổi (transition probabilities) và được mô tả bằng quá trình Markov.

**Known:**

Agent biết trước hàm chuyển đổi (transition function), nghĩa là nó có mô hình hoàn chỉnh về cách hành động ảnh hưởng đến trạng thái tiếp theo của môi trường.

**Unknown:**

Agent không biết hàm chuyển đổi và phải học nó thông qua thử nghiệm các hành động để khám phá mối quan hệ giữa hành động và kết quả.