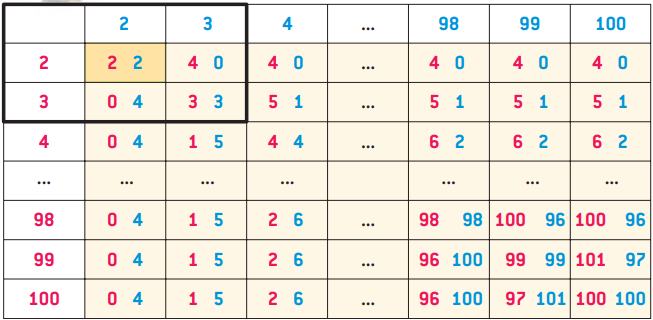
# I - Phát biểu bài toán: mô tả các bài toán cần giải quyết.

Bài toán “Travel’s dilemma” nói về việc 2 người bị mất hành lý ở 1 sân bay và 2 hành lý đó giống y hệt nhau. Người quản lý sân bay vì muốn tránh lừa đảo nên đã đưa ra phương pháp sau: Cho 2 người viết xuống 1 tờ giấy giá tiền của hành lý trong khoảng giữa 2$ và 100$ và họ không được trao đổi với nhau. Nếu 2 người viết giá tiền bằng nhau, thì quản lý sẽ cho là 2 người đều thành thật và 2 người đều nhận được số tiền đó. Nếu 2 người viết giá tiền khác nhau, thì người viết giá tiền cao hơn sẽ nhận được số tiền bằng giá tiền thấp hơn -2, còn người viết giá tiền thấp hơn sẽ được thưởng 2$ vào giá tiền thấp hơn đó.

Ví dụ với 2 người Tuấn và Lan, ban đầu Tuấn nghĩ rằng mình sẽ ghi số lớn nhất là 100$ và sẽ nhận được 100$ nếu Lan cũng tham lam như vậy. Nhưng rồi Tuấn nghĩ là nếu mình viết 99$, thì mình sẽ nhận được 101$ và Lan nhận được 98$, nhưng sau đó Tuấn chợt nhận ra rằng Lan cũng sẽ nghĩ giống mình là cũng sẽ viết 99$. Thế là Tuấn lại viết 98$, và lại nghĩ rằng Lan nghĩ giống mình, quy trình này lặp lại cho tới khi cả 2 người viết 2$ trên giấy và cùng đi về với 2$. Kết quả này là Nash Equilibrium của bài toán, và cách suy nghĩ này được gọi là quy nạp ngược (Backward induction).

Nhưng thực tế con người sẽ không có suy nghĩ như vậy và chỉ viết trong khoảng từ 95$ tới 97$. Bài toán này là một nghịch lý trong Game Theory, con người hành động bất hợp lý (irrational) lại mang về lợi ích nhiều hơn hành động hợp lý (rational). Như John Nash từng nói “Game theory predicts that the Nash equilibrium will occur when Traveler’s Dilemma is played rationally.”.

Việc thưởng phạt dựa trên lựa chọn của 2 người có thể được biểu diễn bằng 1 payoff matrix như sau:



Hình 1: The Traveler’s Dilemma, Kaushik Basu, Scientific American , Vol. 296, No. 6 (JUNE 2007), pp. 90-95

Payoff matrix này có thể tóm gọn hết ý tưởng của bài toán Travel’s Dilemma cho người xem, cột bên trái cùng là đại diện cho những lựa chọn của Lan, hàng trên cùng là đại diện cho những lựa chọn của Tuấn. Như ta thấy thì khi Lan chọn 100, Tuấn chọn 100 thì cả 2 người đều nhận được 100, và khi Tuấn chọn 100, Lan chọn 99 thì Tuấn nhận được 97 còn Lan lại nhận được 101, và ngược lại. Bảng cứ theo quy luật này đi lên cho tới khi gặp (2; 2) là Nash equilibrium, (2; 2) là Nash equilibrium vì nó là trạng thái cuối cùng mà người chơi sẽ nghĩ tới và sẽ không muốn chuyển qua trạng thái nào nữa.

# II - Thách thức: khó khăn, thách thức cụ thể đối với từng bài toán.

Thách thức lớn nhất về bài Traveler’s Dilemma đó chính là sự nghịch lý về hành vi hợp lí và hành vi thực tế của con người như đã có đề cập ở trên. Hành vi hợp lí theo Lý thuyết trò chơi đó chính là người chơi sẽ luôn cố gắng làm cho mình hơn đối thủ, nên sẽ cứ tiếp tục suy nghĩ tới số tiền nhỏ hơn vì nó sẽ làm cho mình nhận được tiền thưởng cao hơn đối thủ, hành vi này hội tụ về việc cả 2 người cùng chọn 2 – Nash equilibrim – vì sau đó không còn lựa chọn nào mà 2 người sẽ muốn chọn nữa. Còn hành vi thực tế của con người là sẽ không quan tâm nhiều tới người kia mà quan tâm tới số tiền mình nhận được, nên thường lựa chọn sẽ nằm dao động ở khoảng trên 95. Việc mâu thuẫn như vậy dẫn tới việc chọn policy để mô phỏng lại bài toán phải được cân nhắc kỹ cho cả 2 loại hành vi.

Còn khó khăn đối với em đó chính là việc hiểu và tìm được code tham khảo cho bài toán này, vì lý thuyết trò chơi và những khái niệm liên quan là rất mới đối với em nên em còn hơi lúng túng khi tra cứu trên mạng.

# III - Thực nghiệm:

1. **Mô hình hoá tính toán:**

struct SimpleGame

      γ  # discount factor

      ℐ  # agents

      𝒜  # joint action space

      R  # joint reward function

end

Trong simple game, những agent chọn một hành động ai để tối ưu hoá phần thưởng ri của họ.

Một joint action space bao gồm mọi hoán vị của những hành động mà mỗi agent có thể có. Những hành động được chọn cùng một lúc qua khắp các agent có thể được gộp lại thành 1 joint action từ joint action space trên.

Joint reward function mô tả phần thưởng nhận được từ joint action . Joint reward là .

Còn với γ (discount factor), đó chỉ là biến đại diện cho xác suất “đi tiếp” của lần lặp hiện tại để cho nó không lặp vô tận, giá trị này nằm giữa 0 và 1 nhưng không thể bằng 0 và 1.

Với bài toán traveler’s dilemma, discount factor sẽ bằng 0.9, agents sẽ gồm 1 và 2, joint action space sẽ là một list chứa các hành động chọn 2…100 của mỗi agent, và joint reward function sẽ gồm 2 hàm reward giống hệt nhau của 2 agent có dạng như sau:

1. **Phương pháp giải quyết, phương pháp, thuật toán được sử dụng cho việc giải quyết bài toán**

Đối với bài này thì em sẽ chọn 2 cấu hình policy Iterated Best Response và Hierarchical Softmax, tương ứng cho 2 hành vi là hành vi hợp lí và hành vi thực tế.

1. Iterated Best Response
2. Best response

Với việc người chơi biết những gì mà những người chơi khác đang làm, một chiến thuật là một best response nếu và chỉ nếu người chơi không thể có lợi hơn khi chuyển sang chiến thuật khác. Có thể có nhiều best response.

Định nghĩa Nash Equilibrium bằng Best response: một trò chơi đang ở trong trạng thái Nash equilibrium nếu và chỉ nếu tất cả người chơi đều đang có phản ứng tốt nhất (best response) cho việc mà những người chơi khác đang làm.

Cách tìm best response cho 1 trò chơi là cô lập 1 chiến thuật của 1 người chơi, và tìm xem phản ứng tốt nhất mà những người chơi còn lại nên có để đáp trả chiến thuật đó. Cụ thể với traveler’s dilemma:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 2 | 3 | 4 | 5 | … |
| 2 | 2; 2 | 4; 0 | 4; 0 | 4; 0 |  |
| 3 | 0; 4 | 3; 3 | 5; 1 | 5; 1 |  |
| 4 | 0; 4 | 1; 5 | 4; 4 | 6; 2\* |  |
| 5 | 0; 4 | 1; 5 | 2; 6 | 5; 5 |  |
| … |  |  |  |  |  |

Những lựa chọn ở cột trái cùng sẽ là của player1, còn những lựa chọn ở hàng trên cùng sẽ là của player2. Em chỉ xét những giá trị từ 2 tới 5 để dễ ví dụ.

Như bảng thì em đã cô lập chiến thuật của player2 là chọn giá tiền 5, thì với chiến thuật của player2 như vậy, thì chiến thuật tốt nhất của player1 sẽ là chọn 4 vì nó cho ra giá tiền cao nhất cho player đó, là 6. Cho nên best response của player1 khi player2 chọn 5 là chọn 4.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 2 | 3 | 4 | 5 | … |
| 2 | 2; 2 | 4; 0 | 4; 0 | 4; 0 |  |
| 3 | 0; 4 | 3; 3 | 5; 1\* | 5; 1 |  |
| 4 | 0; 4 | 1; 5 | 4; 4 | 6; 2\* |  |
| 5 | 0; 4 | 1; 5 | 2; 6 | 5; 5 |  |
| … |  |  |  |  |  |

Tương tự, khi cô lập chiến thuật chọn 4 của player2, thì chiến thuật tốt nhất của player1 là 3 vì nó cho ra giá tiền lớn nhất cho player1 là 5, nên best response của player1 khi player2 chọn 4 là chọn 3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 2 | 3 | 4 | 5 | … |
| 2 | 2\*; 2\* | 4; 0 | 4; 0 | 4; 0 |  |
| 3 | 0; 4 | 3; 3 | 5; 1\* | 5; 1 |  |
| 4 | 0; 4 | 1; 5 | 4; 4 | 6; 2\* |  |
| 5 | 0; 4 | 1; 5 | 2; 6 | 5; 5 |  |
| … |  |  |  |  |  |

Giờ em sẽ xét cùng lúc 2 chiến thuật của 2 người chơi là chọn 2. Em sẽ cô lập lựa chọn 2 của player2 trước (cột màu đỏ), thì khi player2 chọn 2, best response của player1 sẽ là 2 vì nó cho ra giá tiền cao nhất cho player1 là 2 thay vì 0 như những lựa chọn dưới. Tương tự, khi cô lập lựa chọn 2 của player1 (hàng màu xanh), thì player2 tốt nhất nên chọn 2 luôn vì nó cũng ra giá tiền cao nhất là 2 thay vì là 0 như những lựa chọn sau.

Vì ô chiến thuật đó, cả 2 player đều có best response cho lẫn nhau, nên (2; 2) chính là Nash equilibrium cho bài toán.

1. Interated best response

Với phương pháp tìm Nash Equilibrium bằng Best response em vừa mô tả ở trên, thì chi phí sẽ rất là cao đối với 1 bài toán có nhiều hành vi như vậy, cho nên, ta phải có 1 phương pháp tối ưu hơn để tính được Nash Equilibrium cho bài toán.

o Với mỗi thử nghiệm:

 Mô hình hóa tính toán: cấu trúc dữ liệu hoặc phương pháp thể hiện các bàin toán.

 Phương pháp giải quyết: phương pháp, thuật toán được sử dụng cho việc giải quyết bài toán, lý do lựa chọn phương pháp.

Với hành vi hợp lí, policy Iterated best response là phù hợp nhất vì nó vừa đảm bảo tính được Nash Equilibrium, vừa dễ cài đặt. Và policy Hierarchical Softmax thì lại phù hợp với hành vi thực tế và đưa ra kết quả đúng dự đoán.

 Code: hướng dẫn và mô tả về đoạn code mà bạn đã viết (hay sử dụng). Lưu ý: cần ghi rõ nguồn (nếu bạn sao chép đoạn code ở đâu đó).

 Phân tích: kết quả thử nghiệm và ý nghĩa.

# IV - Tóm tắt kết quả:

o Dựa trên các tiêu chí đánh giá kết quả của bạn.

o Điểm mạnh và điểm yếu trong đồ án này của bạn.